Solutions manual for *Linear Algebra*

Elizabeth Meckes Mark Meckes

May 3, 2019

Contents

Chapter 1

Linear systems and vector spaces

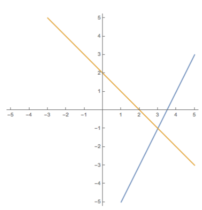
1.1 Linear systems of equations

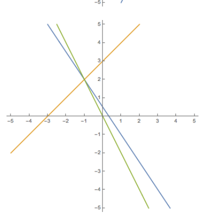
1.1.1. (a) If we use the available ingredients to make x loaves of bread, y pints of beer, and z rounds of cheese, then

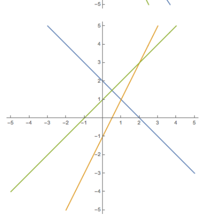
$$\begin{aligned} 2x + \frac{1}{4}y &= 35 \\ x + 8z &= 176 \\ 2x + \frac{1}{4}y + z &= 56. \end{aligned}$$

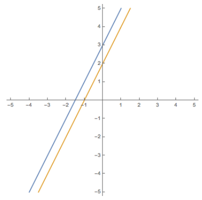
Subtracting the first equation from the third yields $z = 21$. Substituting this in the second equation yields $x = 8$, and then the first equation yields $y = 76$. So you can make 8 loaves of bread, 76 pints of beer, and 21 rounds of cheese.

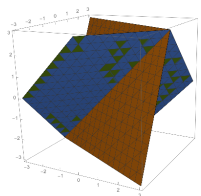
(b) You'll use $x + \frac{1}{4}y = 27$ pounds of barley.

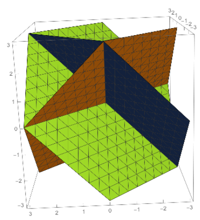
1.1.2. (a)  Consistent with a unique solution.

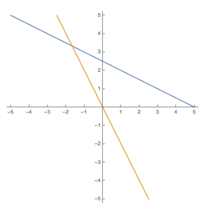
(b)  Consistent with a unique solution.

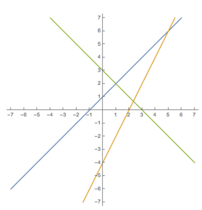
(c)  Inconsistent.

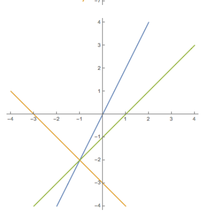
(d)  Inconsistent.

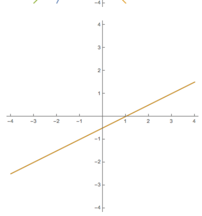
(e)  Consistent; solution is not unique.

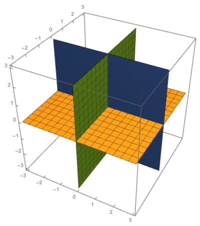
(f)  Consistent with a unique solution.

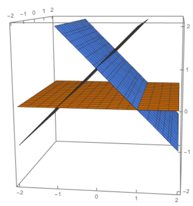
1.1.3. (a)  Consistent with a unique solution.

(b)  Inconsistent.

(c)  Consistent with a unique solution.

(d)  Consistent; solution is not unique.

(e)  Consistent with a unique solution.

(f)  Inconsistent.

1.1.4. The system consisting of the first three equations in (1.3) has a unique solution, so changing the fourth equation so that $(16, 16, 20)$ is not a solution will produce an inconsistent system, for example

$$\begin{aligned} x + \frac{1}{4}y &= 20 \\ 2x + \frac{1}{4}y &= 36 \\ x + 8z &= 176 \\ 2x + \frac{1}{4}y + z &= 50. \end{aligned}$$

The inconsistency of this system means that it is not possible to exactly use up 20 pounds of barley, 36 tsps of yeast, 176 cups of milk, and 50 tsps of rosemary using our recipes.

1.1.5. (a) The set of $(x, y, z) \in \mathbb{R}^3$ with $z = 0$ is exactly the x-y plane.

(b) Here, $z = y = 0$, so this is the x-axis.

(c) Since $x = y = z = 0$, this is just the origin.

(d) Again, we have $x = y = z = 0$, and the fourth equation is redundant, so this is the origin.

(e) Only $x = y = z = 0$ satisfies the first three equations, but this is not a solution to the fourth equation, so the system is inconsistent.

1.1.6. (a) The set of $(x, y, z) \in \mathbb{R}^3$ with $y = 1$ is a plane parallel to the x-z plane, shifted one unit along the positive y-axis.

(b) Here, $x = y = 0$, so this is the z-axis.

(c) Adding the first two equations and subtracting the third yields $z = 0$, from which it follows that $x = y = 0$, so this is the origin.

(d) Again, $x = y = z = 0$ from the first three equations. Since this satisfies the fourth equation, the set is again the origin.

(e) Only $x = y = z = 0$ satisfies the first three equations, but this is not a solution to the fourth equation, so the system is inconsistent.

1.1.7. We know that $f(-1) = 1$, $f(0) = 0$, and $f(1) = 2$. This is the same as the system of equations

$$\begin{aligned} a - b + c &= 1, \\ c &= 0, \\ a + b + c &= 2. \end{aligned}$$

Subsituting $c = 0$ into the first and third equations, we get

$$\begin{aligned} a - b &= 1, \\ a + b &= 2. \end{aligned}$$

Adding these equations together gives $2a = 3$, so $a = 3/2$, and then the last equation gives $b = 1/2$. Thus $(3/2, 1/2, 0)$ is the unique solution.

1.1.8. (a) Since $f(0) = a + b\cos(0) + c\sin(0) = a + b$, we have $a + b = -1$. Similarly, $2 = f(\frac{\pi}{2}) = a + c$ and $3 = f(\pi) = a - b$, so the linear system is

$$\begin{aligned} a + b &= -1 \\ a + c &= 2 \\ a - b &= 3. \end{aligned}$$

(b) Adding the first and third equations shows $2a = 2$ or $a = 1$. Using this in the first equation then gives $b = -2$, and using $a = 1$ in the second equation gives $c = 1$: the (unique) solution is $(1, -2, 1)$.

1.1.9. We just use the Rat Poison Principle:

$$a\frac{de - bf}{ad - bc} + b\frac{af - ec}{ad - bc} = \frac{ade - abf + baf - bec}{af - ec} = \frac{(ad - bc)e}{ad - bc} = e$$

and

$$c\frac{de - bf}{ad - bc} + d\frac{af - ec}{ad - bc} = \frac{cde - cbf + daf - dec}{af - ec} = \frac{(ad - bc)f}{ad - bc} = f.$$

1.1.10. We are given that

$$\begin{aligned} a_{11}c_1 + \cdots + a_{1n}c_n &= b_1, \\ &\vdots \\ a_{m1}c_1 + \cdots + a_{mn}c_n &= b_m \end{aligned}$$

and

$$\begin{aligned} a_{11}d_1 + \cdots + a_{1n}d_n &= b_1, \\ &\vdots \\ a_{m1}d_1 + \cdots + a_{mn}d_n &= b_m, \end{aligned}$$

so

$$\begin{aligned} a_{11}(c_1+d_1)+\cdots+a_{1n}(c_n+d_n) &= 2b_1, \\ &\vdots \\ a_{m1}(c_1+d_1)+\cdots+a_{mn}(c_n+d_n) &= 2b_m. \end{aligned}$$

So for $(c_1+d_1,\ldots,c_n+d_n)$ to be a solution, we'd need $2b_j = b_j$ for each j; that is, $b_j = 0$ for each j.

1.1.11. We are given that

$$\begin{aligned} a_{11}c_1+\cdots+a_{1n}c_n &= b_1, \\ &\vdots \\ a_{m1}c_1+\cdots+a_{mn}c_n &= b_m \end{aligned}$$

and

$$\begin{aligned} a_{11}d_1+\cdots+a_{1n}d_n &= b_1, \\ &\vdots \\ a_{m1}d_1+\cdots+a_{mn}d_n &= b_m, \end{aligned}$$

so

$$\begin{aligned} a_{11}(tc_1+(1-t)d_1)+\cdots+a_{1n}(tc_n+(1-t)d_n) &= tb_1+(1-t)b_1 = b_1, \\ &\vdots \\ a_{m1}(tc_1+(1-t)d_1)+\cdots+a_{mn}(tc_n+(1-t)d_n) &= tb_m+(1-t)b_m = b_m. \end{aligned}$$

So for any real number t, $(tc_1+(1-t)d_1,\ldots,tc_n+(1-t)d_n)$ is be a solution.

1.2 Gaussian elimination

1.2.1.
(a) **R1:** Add -1 times the fourth row to the third row.
(b) **R3:** Switch the first and second rows.
(c) **R1:** Add (1 times) the third row to the first row.
(d) **R1:** Add -10 times either the first or second row to the fourth row.
(e) **R2:** Multiply the fourth row by $-1/9$.
(f) **R2:** Multiply the third row by $-1/3$.
(g) **R1:** Add -1 times the third row to the first row.
(h) **R1:** Add -1 times the third row to the second row.

1.2.2.
(a) **R1:** Add -1 times the first row to the third row.

(b) **R1:** Add -4 times the first row to the second row.

(c) **R2:** Multiply the second row by -1.

(d) **R1:** Add -1 times the second row to the first row.

(e) **R1:** Add (1 times) the first row to the thrid row.

(f) **R1:** Add -1 times the third row to the fourth row.

(g) **R1:** Add -1 times the fourth row to the second row.

(h) **R3:** Switch the second and fourth rows.

1.2.3. (a) $\begin{bmatrix}1 & 0 & -\frac{1}{2}\\ 0 & 1 & -\frac{7}{6}\end{bmatrix}$ (b) $\left[\begin{array}{cccc|c}1 & 0 & -1 & 0 & 3\\ 0 & 1 & 2 & 0 & -1\\ 0 & 0 & 0 & 1 & 2\end{array}\right]$ (c) $\begin{bmatrix}1 & 0\\ 0 & 1\\ 0 & 0\end{bmatrix}$ (d) $\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}$ (e) $\begin{bmatrix}1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\end{bmatrix}$

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

1.2.4. (a) $\begin{bmatrix}1 & 0 & \frac{2}{3}\\ 0 & 1 & -\frac{1}{12}\end{bmatrix}$ (b) $\left[\begin{array}{cccc|c}1 & 0 & 2 & 0 & 1\\ 0 & 1 & -1 & 0 & -2\\ 0 & 0 & 0 & 1 & 0\end{array}\right]$ (c) $\begin{bmatrix}1 & 0\\ 0 & 1\\ 0 & 0\end{bmatrix}$ (d) $\begin{bmatrix}1 & 0 & -1\\ 0 & 1 & 2\\ 0 & 0 & 0\end{bmatrix}$ (e) $\begin{bmatrix}1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\end{bmatrix}$

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

1.2.5. Perform Gaussian elimination on the augmented matrix:

$$\left[\begin{array}{cccc|c}1 & \frac{1}{4} & 0 & \frac{1}{2} & 20\\ 2 & \frac{1}{4} & 0 & 0 & 36\\ 1 & 0 & 8 & 1 & 176\end{array}\right] \xrightarrow{\mathbf{R1}\text{ (twice)}} \left[\begin{array}{cccc|c}1 & \frac{1}{4} & 0 & \frac{1}{2} & 20\\ 0 & -\frac{1}{4} & 0 & -1 & -4\\ 0 & -\frac{1}{4} & 8 & \frac{1}{2} & 156\end{array}\right] \xrightarrow{\mathbf{R2}} \left[\begin{array}{cccc|c}1 & \frac{1}{4} & 0 & \frac{1}{2} & 20\\ 0 & 1 & 0 & 4 & 16\\ 0 & -\frac{1}{4} & 8 & \frac{1}{2} & 156\end{array}\right]$$

$$\xrightarrow{\mathbf{R1}} \left[\begin{array}{cccc|c}1 & \frac{1}{4} & 0 & \frac{1}{2} & 20\\ 0 & 1 & 0 & 4 & 16\\ 0 & 0 & 8 & \frac{3}{2} & 160\end{array}\right] \xrightarrow{\mathbf{R2}} \left[\begin{array}{cccc|c}1 & \frac{1}{4} & 0 & \frac{1}{2} & 20\\ 0 & 1 & 0 & 4 & 16\\ 0 & 0 & 1 & \frac{3}{16} & 20\end{array}\right]$$

1.2.6. (a) The RREF is $\left[\begin{array}{ccc|c}1 & 0 & -1 & -2\\ 0 & 1 & 2 & 3\end{array}\right]$. Solutions are $x = -2 + z$, $y = 3 - 2z$ for any $z \in \mathbb{R}$.

(b) The RREF is $\left[\begin{array}{cccc|c}1 & 0 & 5 & -1 & -1\\ 0 & 1 & 2 & -2 & -2\end{array}\right]$. Solutions are $x = -1-5z+w$, $y = -2-2z+2w$ for any $z, w \in \mathbb{R}$.

(c) The RREF is $\left[\begin{array}{ccc|c}1 & 0 & 1 & 2\\ 0 & 1 & -1 & 3\\ 0 & 0 & 0 & 0\end{array}\right]$. Solutions are $x = 2 - z$, $y = 3 + z$ for any $z \in \mathbb{R}$.

(d) The RREF is $\left[\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right]$. The system is inconsistent.

(e) The RREF is $\left[\begin{array}{ccc|c} 1 & 0 & 0 & -2 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \end{array}\right]$. The unique solution is $x = -2$, $y = 1$, $z = -1$.

1.2.7. (a) The RREF is $\left[\begin{array}{ccc|c} 1 & 0 & -1 & -\frac{7}{2} \\ 0 & 1 & 2 & \frac{15}{4} \end{array}\right]$. Solutions are $x = -\frac{7}{2} + z$, $y = \frac{15}{4} - 2z$ for any $z \in \mathbb{R}$.

(b) The RREF is $\left[\begin{array}{cccc|c} 1 & 0 & -\frac{1}{7} & 2 & 1 \\ 0 & 1 & -\frac{5}{7} & 1 & 0 \end{array}\right]$. Solutions are $x = 1 + \frac{1}{7}z - 2w$, $y = \frac{5}{7}z - w$ for any $z, w \in \mathbb{R}$.

(c) The RREF is $\left[\begin{array}{ccc|c} 1 & 0 & 1 & 2 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right]$. Solutions are $x = 2 - z$, $y = 1 - z$ for any $z \in \mathbb{R}$.

(d) The RREF is $\left[\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right]$. The unique solution is $x = 1$, $y = 0$, $z = 1$.

(e) The RREF is $\left[\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right]$. The system is inconsistent.

1.2.8. We know that $f(-1) = 2$, $f(0) = 3$, $f(1) = 4$, and $f(2) = 15$. Using the formula for $f(x)$, this yields the linear system

$$\begin{aligned} -a + b - c + d &= 2 \\ d &= 3 \\ a + b + c + d &= 4 \\ 8a + 4b + 2c + d &= 15 \end{aligned}$$

with augmented matrix

$$\left[\begin{array}{cccc|c} -1 & 1 & -1 & 1 & 2 \\ 0 & 0 & 0 & 1 & 3 \\ 1 & 1 & 1 & 1 & 4 \\ 8 & 4 & 2 & 1 & 15 \end{array}\right] \rightsquigarrow \left[\begin{array}{cccc|c} 1 & 0 & 0 & 0 & \frac{5}{3} \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & -\frac{2}{3} \\ 0 & 0 & 0 & 1 & 3 \end{array}\right].$$

Therefore $a = 5/3$, $b = 0$, $c = -2/3$, $d = 3$.

1.2.9. We know that $f(-2) = 2$, $f(-1) = -1$, $f(0) = 2$, and $f(1) = 5$. Using the formula for $f(x)$, this yields the linear system

$$\begin{aligned} -8a + 4b - 2c + d &= 2 \\ -a + b - c + d &= -1 \\ d &= 2 \\ a + b + c + d &= 5 \end{aligned}$$

with augmented matrix

$$\left[\begin{array}{cccc|c} -8 & 4 & -2 & 1 & 2 \\ -1 & 1 & -1 & 1 & -1 \\ 0 & 0 & 0 & 1 & 2 \\ 1 & 1 & 1 & 1 & 5 \end{array}\right] \rightsquigarrow \left[\begin{array}{cccc|c} 1 & 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 4 \\ 0 & 0 & 0 & 1 & 2 \end{array}\right].$$

Therefore $a = -1$, $b = 0$, $c = 4$, $d = 2$.

1.2.10. First suppose that $a \neq 0$. Rewriting the system as an augmented matrix and performing Gaussian elimination yields

$$\left[\begin{array}{cc|c} a & b & e \\ c & d & f \end{array}\right] \rightsquigarrow \left[\begin{array}{cc|c} 1 & \frac{b}{a} & \frac{e}{a} \\ c & d & f \end{array}\right] \rightsquigarrow \left[\begin{array}{cc|c} 1 & \frac{b}{a} & \frac{e}{a} \\ 0 & d - \frac{bc}{a} & f - \frac{ec}{a} \end{array}\right].$$

If $ad - bc = a(d - \frac{bc}{a}) \neq 0$, then $d - \frac{bc}{a} \neq 0$, and so the REF of the augmented matrix of the system has a pivot in each column to the left of the bar and no pivot in the final column, which means the system has a single unique solution.

Now suppose that $a = 0$. Then saying $ad - bc \neq 0$ is the same thing as saying $bc \neq 0$, which is equivalent to the statment that neither b nor c is 0. So again, we can perform Gaussian elimination:

$$\left[\begin{array}{cc|c} 0 & b & e \\ c & d & f \end{array}\right] \rightsquigarrow \left[\begin{array}{cc|c} c & d & f \\ 0 & b & e \end{array}\right] \rightsquigarrow \left[\begin{array}{cc|c} 1 & \frac{d}{c} & \frac{f}{c} \\ 0 & 1 & \frac{e}{b} \end{array}\right].$$

Again, each column to the left of the bar contains a pivot, and there is no pivot in the last column, so the system has a unique solution.

1.2.11. We wish to solve the linear system

$$\begin{aligned} x + \frac{1}{4}y &= 1 \\ 2x + \frac{1}{4}y &= 2 \\ x + 8z &= 5 \\ 2x + \frac{1}{4}y + z &= 2 \end{aligned}$$

which has RREF

$$\left[\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right].$$

Therefore the system is inconsistent; there is no solution.

1.2.12. (a) A linear system has solutions at all exactly when its RREF does not include $0 = 1$. In that case, there is a unique solution exactly when there are no free variables. Otherwise, the free variables can take on infinitely many values, leading to infinitely many solutions.

(b) The system given here isn't linear, so things we know about linear systems don't apply.

1.2.13. Except for part (b) the justifications are obvious.

(a) $$\begin{aligned} x + y + z &= 1 \\ x + y + z &= 2 \end{aligned}$$

(b) Impossible by Corollary 1.4.

(c) $x + y = 0$

(d) $$\begin{aligned} x &= 1 \\ x &= 2 \end{aligned}$$

(e) $$\begin{aligned} x + y &= 1 \\ 2x + 2y &= 2 \\ 3x + 3y &= 3 \end{aligned}$$

(f) $$\begin{aligned} x &= 0 \\ y &= 0 \\ x + y &= 0 \end{aligned}$$

(g) $$\begin{aligned} x + y &= 0 \\ x + y &= 1 \end{aligned}$$

(h) $$\begin{aligned} x &= 0 \\ y &= 0 \end{aligned}$$

(i) $$\begin{aligned} x + y &= 0 \\ 2x + 2y &= 0 \end{aligned}$$

1.2.14. Performing Gaussian elimination on the system:

$$\left[\begin{array}{ccc|c} 2 & 1 & 0 & b_1 \\ 1 & -1 & 3 & b_2 \\ 1 & -2 & 5 & b_3 \end{array}\right] \rightsquigarrow \left[\begin{array}{ccc|c} 1 & \frac{1}{2} & 0 & \frac{b_1}{2} \\ 0 & -\frac{3}{2} & 3 & b_2 - \frac{b_1}{2} \\ 0 & -\frac{5}{2} & 5 & b_3 - \frac{b_1}{2} \end{array}\right] \rightsquigarrow \left[\begin{array}{ccc|c} 1 & \frac{1}{2} & 0 & \frac{b_1}{2} \\ 0 & 1 & -2 & -\frac{2}{3}b_2 + \frac{b_1}{3} \\ 0 & 0 & 0 & -\frac{2}{5}b_3 + \frac{2}{3}b_2 - \frac{2}{15}b_1 \end{array}\right].$$

The system has a solution if and only if there is no pivot in the final column; that is, if and only if $-\frac{2}{5}b_3 + \frac{2}{3}b_2 - \frac{2}{15}b_1 = 0$.

1.2.15. Performing Gaussian elimination on the system:

$$
\left[\begin{array}{cccc|c} 1 & 2 & -1 & 7 & b_1 \\ 0 & -1 & 1 & -3 & b_2 \\ 2 & 0 & -2 & 2 & b_3 \\ -3 & 1 & 4 & 0 & b_4 \end{array}\right] \rightsquigarrow \left[\begin{array}{cccc|c} 1 & 2 & -1 & 7 & b_1 \\ 0 & -1 & 1 & -3 & b_2 \\ 0 & -4 & 0 & -12 & -2b_1 + b_3 \\ 0 & 7 & 1 & 21 & 3b_1 + b_4 \end{array}\right]
$$

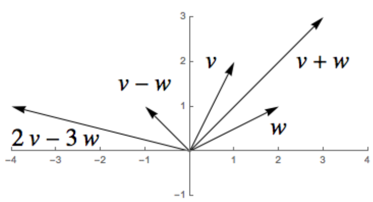
$$
\rightsquigarrow \left[\begin{array}{cccc|c} 1 & 2 & -1 & 7 & b_1 \\ 0 & -1 & 1 & -3 & b_2 \\ 0 & 0 & -4 & 0 & -2b_1 - 4b_2 + b_3 \\ 0 & 0 & 8 & 0 & 3b_1 + 7b_2 + b_4 \end{array}\right]
$$

$$
\rightsquigarrow \left[\begin{array}{cccc|c} 1 & 2 & -1 & 7 & b_1 \\ 0 & -1 & 1 & -3 & b_2 \\ 0 & 0 & -4 & 0 & -2b_1 - 4b_2 + b_3 \\ 0 & 0 & 0 & 0 & -b_1 - b_2 + 2b_3 + b_4 \end{array}\right].
$$

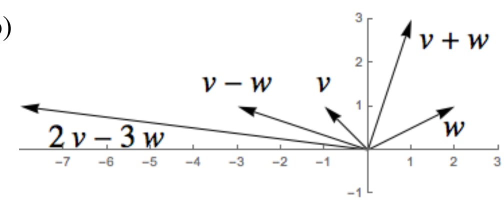
The system has a solution if and only if there is no pivot in the final column; that is, if and only if $b_1 + b_2 - 2b_3 - b_4 = 0$.

1.3 Vectors and the geometry of linear systems

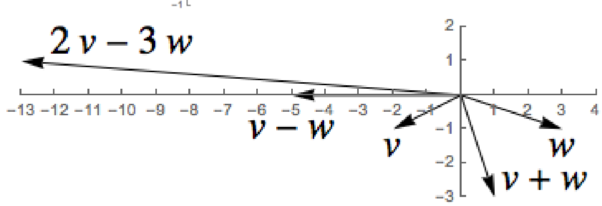
1.3.1. (a)



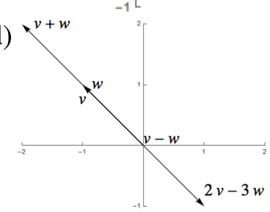
(b)



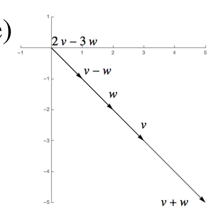
(c)



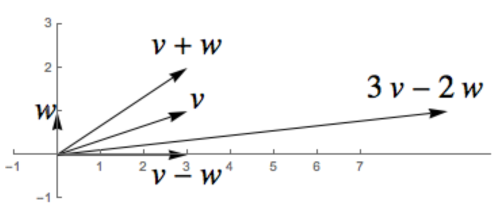
(d)



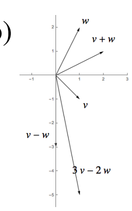
(e)



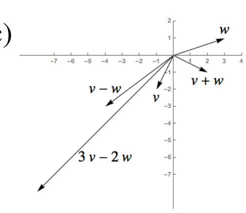
1.3.2. (a)



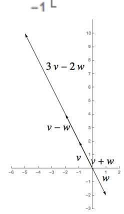
(b)



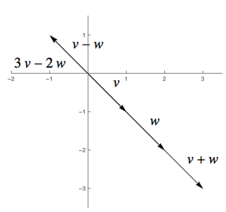
(c)



(d)



(e)



1.3.3. (a) The system has RREF $\left[\begin{array}{ccc|c} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & -1 \\ 0 & 0 & 0 & 0 \end{array}\right]$, hence solutions $x = 1 - z$ and $y = -1 - z$ for any $z \in \mathbb{R}$, which can be written in vector form as $\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + z \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix}$ for $z \in \mathbb{R}$.

(b) The system has RREF $\left[\begin{array}{ccc|c} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{array}\right]$, hence solutions $x = 1 - z$ and $y = 2 - z$ for any $z \in \mathbb{R}$, which can be written in vector form as $\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} + z \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix}$ for $z \in \mathbb{R}$.

(c) The system has RREF $\left[\begin{array}{ccc|c} 1 & 0 & \frac{5}{3} & 2 \\ 0 & 1 & -\frac{2}{3} & -1 \end{array}\right]$, hence solutions $x = 2 - \frac{5}{3}z$ and $y = -1 + \frac{2}{3}z$ for $z \in \mathbb{R}$, which can be written in vector form as $\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix} + z\frac{1}{3} \begin{bmatrix} -5 \\ 2 \\ 3 \end{bmatrix}$ for $z \in \mathbb{R}$.

(d) The system has RREF $\left[\begin{array}{cccc|c} 1 & 0 & -1 & -2 & -3 \\ 0 & 1 & 2 & 3 & 4 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$, hence solutions $x_1 = -3 + x_3 + 2x_4$ and $x_2 = 4 - 2x_3 - 3x_4$ for any $x_3, x_4 \in \mathbb{R}$, which can be written in vector form as $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} -3 \\ 4 \\ 0 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 1 \\ -2 \\ 1 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 2 \\ -3 \\ 0 \\ 1 \end{bmatrix}$ for $x_3, x_4 \in \mathbb{R}$.

(e) The system has RREF $\left[\begin{array}{cc|c} 1 & -2 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{array}\right]$, hence solutions $x = 2y$ for any $y \in \mathbb{R}$, which can be written in vector form as $\begin{bmatrix} x \\ y \end{bmatrix} = y \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ for $y \in \mathbb{R}$.

(f) The system has RREF $\left[\begin{array}{cc|c} 1 & -2 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{array}\right]$, hence solutions $x = 1 + 2y$ for any $y \in \mathbb{R}$, which can be written in vector form as $\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} + y \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ for $y \in \mathbb{R}$.

1.3.4. (a) The system has RREF $\left[\begin{array}{ccc|c} 1 & 0 & 1 & 2 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$, hence solutions $x = 2 - z$ and $y = -z$ for any $z \in \mathbb{R}$, which can be written in vector form as $\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix} + z \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix}$ for $z \in \mathbb{R}$.

(b) The system has RREF $\left[\begin{array}{ccc|c} 1 & 0 & 1 & 2 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right]$, hence solutions $x = 2 - z$ and $y = 1 - z$ for any

$z \in \mathbb{R}$, which can be written in vector form as $\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} + z \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix}$ for $z \in \mathbb{R}$.

(c) The system has RREF $\left[\begin{array}{ccc|c} 1 & 0 & \frac{3}{5} & \frac{1}{5} \\ 0 & 1 & -\frac{9}{5} & \frac{2}{5} \end{array}\right]$, hence solutions $x = \frac{1}{5} - \frac{3}{5}z$ and $y = \frac{2}{5} + \frac{9}{5}z$ for $z \in \mathbb{R}$, which can be written in vector form as $\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \frac{1}{5} \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} + z\frac{1}{5} \begin{bmatrix} -3 \\ 9 \\ 5 \end{bmatrix}$ for $z \in \mathbb{R}$.

(d) The system has RREF $\left[\begin{array}{cccc|c} 1 & 0 & -1 & 2 & -3 \\ 0 & 1 & -2 & 3 & -4 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$, hence solutions $x_1 = -3 + x_3 - 2x_4$ and $x_2 = -4 + 2x_3 - 3x_4$ for any $x_3, x_4 \in \mathbb{R}$, which can be written in vector form as $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} -3 \\ -4 \\ 0 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 1 \\ 2 \\ 1 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -2 \\ -3 \\ 0 \\ 1 \end{bmatrix}$ for $x_3, x_4 \in \mathbb{R}$.

(e) The system has RREF $\left[\begin{array}{cc|c} 1 & -1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{array}\right]$, hence solutions $x = y$ for any $y \in \mathbb{R}$, which can be written in vector form as $\begin{bmatrix} x \\ y \end{bmatrix} = y \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ for $y \in \mathbb{R}$.

(f) The system has RREF $\left[\begin{array}{cc|c} 1 & -1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{array}\right]$, and is therefore inconsistent.

1.3.5. (a) We need to know under what conditions the system with augmented matrix

$$\left[\begin{array}{ccc|c} 1 & 2 & -1 & b_1 \\ 0 & -1 & 1 & b_2 \\ 2 & 0 & -2 & b_3 \\ -3 & 1 & 4 & b_4 \end{array}\right]$$

is consistent. We can reduce this matrix to

$$\left[\begin{array}{ccc|c} 1 & 2 & -1 & b_1 \\ 0 & -1 & 1 & b_2 \\ 0 & 0 & -4 & -2b_1 - 4b_2 + b_3 \\ 0 & 0 & 0 & -b_1 + -b_2 + 2b_3 + b_4 \end{array}\right]$$

so that the system is consistent iff $-b_1 - b_2 + 2b_3 + b_4 = 0$.

(b) Using the condition in part (a): (i) No (ii) No (iii) No (iv) Yes

1.3.6. (a) We need to know under what conditions the system with augmented matrix

$$\left[\begin{array}{ccc|c} 1 & 0 & 2 & b_1 \\ 0 & 2 & -1 & b_2 \\ -1 & 1 & 0 & b_3 \\ 0 & 1 & 0 & b_4 \end{array}\right]$$

is consistent. We can reduce this matrix to

$$\left[\begin{array}{ccc|c} 1 & 0 & 2 & b_1 \\ 0 & 1 & 0 & b_4 \\ 0 & 0 & -1 & b_2 - 2b_4 \\ 0 & 0 & 0 & b_1 + 2b_2 + b_3 - 5b_4 \end{array}\right]$$

so that the system is consistent iff $b_1 + 2b_2 + b_3 - 5b_4 = 0$.

(b) Using the condition in part (a): (i) Yes (ii) No (iii) No (iv) Yes

1.3.7. We want all solutions of the system

$$x\begin{bmatrix}1\\2\end{bmatrix} + y\begin{bmatrix}3\\4\end{bmatrix} + z\begin{bmatrix}5\\6\end{bmatrix} = \begin{bmatrix}-1\\1\end{bmatrix}.$$

The augmented matrix $\left[\begin{array}{ccc|c} 1 & 3 & 5 & -1 \\ 2 & 4 & 6 & 1 \end{array}\right]$ has RREF $\left[\begin{array}{ccc|c} 1 & 0 & -1 & \frac{7}{2} \\ 0 & 1 & 2 & -\frac{3}{2} \end{array}\right]$, so z is a free variable, $x - z = \frac{7}{2}$, and $y + 2z = -\frac{3}{2}$. So for any $z \in \mathbb{R}$,

$$\begin{bmatrix}-1\\1\end{bmatrix} = \left(z + \frac{7}{2}\right)\begin{bmatrix}1\\2\end{bmatrix} + \left(-2z - \frac{3}{2}\right)\begin{bmatrix}3\\4\end{bmatrix} + z\begin{bmatrix}5\\6\end{bmatrix}.$$

1.3.8. We want all solutions of the system

$$x\begin{bmatrix}3\\1\end{bmatrix} + y\begin{bmatrix}4\\1\end{bmatrix} + z\begin{bmatrix}5\\1\end{bmatrix} = \begin{bmatrix}2\\1\end{bmatrix}.$$

The augmented matrix $\left[\begin{array}{ccc|c} 3 & 4 & 5 & 2 \\ 1 & 1 & 1 & 1 \end{array}\right]$ has RREF $\left[\begin{array}{ccc|c} 1 & 0 & -1 & 2 \\ 0 & 1 & 2 & -1 \end{array}\right]$, so z is a free variable, $x - z = 2$, and $y + 2z = -1$. So for any $z \in \mathbb{R}$,

$$\begin{bmatrix}2\\1\end{bmatrix} = (z + 2)\begin{bmatrix}3\\1\end{bmatrix} + (-2z - 1)\begin{bmatrix}4\\1\end{bmatrix} + z\begin{bmatrix}5\\1\end{bmatrix}.$$

1.3.9. For any $a, b \in \mathbb{R}$,

$$\begin{bmatrix}1\\2\\3\\4\end{bmatrix} = (a-2)\begin{bmatrix}1\\1\\0\\0\end{bmatrix} + (b-1)\begin{bmatrix}1\\0\\1\\0\end{bmatrix} + (4-a-b)\begin{bmatrix}1\\0\\0\\1\end{bmatrix} + (4-a-b)\begin{bmatrix}0\\1\\1\\0\end{bmatrix} + b\begin{bmatrix}0\\1\\0\\1\end{bmatrix} + a\begin{bmatrix}0\\0\\1\\1\end{bmatrix}.$$

The goal is to find all solutions to

$$c_1\begin{bmatrix}1\\1\\0\\0\end{bmatrix} + c_2\begin{bmatrix}1\\0\\1\\0\end{bmatrix} + c_3\begin{bmatrix}1\\0\\0\\1\end{bmatrix} + c_4\begin{bmatrix}0\\1\\1\\0\end{bmatrix} + c_5\begin{bmatrix}0\\1\\0\\1\end{bmatrix} + c_6\begin{bmatrix}0\\0\\1\\1\end{bmatrix} = \begin{bmatrix}1\\2\\3\\4\end{bmatrix}.$$

Perform Gaussian elimination on

$$\left[\begin{array}{cccccc|c} 1 & 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 & 2 \\ 0 & 1 & 0 & 1 & 0 & 1 & 3 \\ 0 & 0 & 1 & 0 & 1 & 1 & 4 \end{array}\right]$$

to get

$$\left[\begin{array}{cccccc|c} 1 & 0 & 0 & 0 & 0 & -1 & -2 \\ 0 & 1 & 0 & 0 & -1 & 0 & -1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 4 \\ 0 & 0 & 0 & 1 & 1 & 1 & 4 \end{array}\right].$$

Setting $c_5 = a$ and $c_6 = b$ and solving for the pivot variables in terms of a and b shows that $c_1 = b - 2$, $c_2 = a - 1$, $c_3 = -a - b + 4$ and $c_4 = -a - b + 4$. In other words, the following represents all possible ways to write $\begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}$ as a linear combination of the given vectors:

$$\begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} = (a-2)\begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} + (b-1)\begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix} + (4-a-b)\begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} + (4-a-b)\begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} + b\begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix} + a\begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}.$$

1.3.10. The consistency of the system and the statement that $\mathbf{b} \in \langle \mathbf{v}_1, \ldots, \mathbf{v}_n \rangle$ both mean precisely that there exist $x_1, \ldots, x_n \in \mathbb{R}$ such that $\mathbf{b} = x_1\mathbf{v}_1 + \cdots + x_n\mathbf{v}_n$.

1.3.11. Write $(x, y, z) = (x, y, 0) + (0, 0, z)$. Apply the Pythagorean theorem to get that $\|(x, y, 0)\| = \sqrt{x^2 + y^2}$, and then again to get $\|(x, y, z)\| = \sqrt{\|(x, y, 0)\|^2 + \|(0, 0, z)\|^2} = \sqrt{x^2 + y^2 + z^2}$.

1.3.12. If the system has infinitely many solutions, then there is a solution in which at least one of x, y, z are non-zero; without loss of generality, say that there is a solution with $x \neq 0$. Then we can solve for $\mathbf{v}_1$:

$$\mathbf{v}_1 = -\frac{y}{x}\mathbf{v}_2 - \frac{z}{x}\mathbf{v}_3.$$

This means that $\mathbf{v}_1 \in \langle \mathbf{v}_2, \mathbf{v}_3 \rangle$, which is a plane containing the origin (and $\mathbf{v}_2$ and $\mathbf{v}_3$).

1.3.13. If vectors $\mathbf{v} = (v_1, v_2)$ and $\mathbf{w} = (w_1, w_2)$ do not span $\mathbb{R}^2$ then there exist some x and y such that the linear system

$$\left[\begin{array}{cc|c} v_1 & w_1 & x \\ v_2 & w_2 & y \end{array}\right]$$

is inconsistent. If $v_1 \neq 0$, then the system is equivalent to

$$\left[\begin{array}{cc|c} 1 & \frac{w_1}{v_1} & \frac{x}{v_1} \\ 0 & w_2 - \frac{w_1 v_2}{v_1} & y - \frac{x v_2}{v_1} \end{array}\right],$$

which can only be inconsistent if $w_2 - \frac{w_1 v_2}{v_1} = 0$. In that case, $\mathbf{w} = \frac{w_1}{v_1}\mathbf{v}$. If $v_2 \neq 0$, then $\mathbf{w} = \frac{w_2}{v_2}\mathbf{v}$ similarly. If $v_1 = v_2 = 0$, then $\mathbf{v} = 0\mathbf{w}$.

1.3.14. $\mathbf{u} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \mathbf{v} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \mathbf{w} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}.$

1.3.15. If $k = 0$, there is a single solution: one point. If $k = 1$, the set of solutions is a line. If $k = 2$, the set of solutions is a plane. If $k = 3$, then every point in $\mathbb{R}^3$ is a solution.

1.4 Fields

1.4.1. We have $0 = 0 + 0\sqrt{5} \in \mathbb{F}$ and $1 = 1 + 0\sqrt{5} \in \mathbb{F}$. As noted in the hint, associativity, commutativity, and the distributive law are all known because we are using the usual addition and multiplication in $\mathbb{R}$.

If $a, b, c, d \in \mathbb{Q}$ then

$$(a + b\sqrt{5}) + (c + d\sqrt{5}) = (a + c) + (b + d)\sqrt{5} \in \mathbb{F}$$

and

$$(a + b\sqrt{5})(c + d\sqrt{5}) = (ac + 5bd) + (ad + bc)\sqrt{5} \in \mathbb{F}.$$

Thus addition and multiplication actually give operations on $\mathbb{F}$.

Finally, if $a, b \in \mathbb{Q}$, then

$$-(a + b\sqrt{5}) = (-a) + (-b)\sqrt{5} \in \mathbb{F},$$

and if also a and b are not both 0 then

$$(a + b\sqrt{5})^{-1} = \frac{a - b\sqrt{5}}{(a + b\sqrt{5})(a - b\sqrt{5})} = \frac{a}{a^2 - 5b^2} - \frac{b}{a^2 - 5b^2}\sqrt{5} \in \mathbb{F}.$$

This only works because $a^2 - 5b^2 \neq 0$ for $a, b \in \mathbb{Q}$. Fortunately, that is indeed always true: if $a^2 = 5b^2$, then $a = \pm\sqrt{5}b$, and since $\sqrt{5}$ is irrational, we couldn't have both $a, b \in \mathbb{Q}$. Thus $\mathbb{F}$ contains additive and multiplicative inverses, and so $\mathbb{F}$ is a field.

1.4.2. We have $0 = 0 + 0i \in \mathbb{F}$ and $1 = 1 + 0i \in \mathbb{F}$. As in the hint for Exercise 1.4.1, associativity, commutativity, and the distributive law are all known because we are using the usual addition and multiplication in $\mathbb{C}$.

If $a, b, c, d \in \mathbb{Q}$ then

$$(a + bi) + (c + di) = (a + c) + (b + d)i \in \mathbb{F}$$

and

$$(a + bi)(c + di) = (ac - bd) + (ad + bc)i \in \mathbb{F}.$$

Thus addition and multiplication actually give operations on $\mathbb{F}$.

Finally, if $a, b \in \mathbb{Q}$, then

$$-(a + bi) = (-a) + (-b)i \in \mathbb{F},$$

and if also a and b are not both 0 then

$$(a + bi)^{-1} = \frac{a - bi}{(a + bi)(a - bi} = \frac{a}{a^2 + b^2} - \frac{b}{a^2 + b^2}i \in \mathbb{F}.$$

Thus $\mathbb{F}$ contains additive and multiplicative inverses, and so $\mathbb{F}$ is a field.

1.4.3. (a) Doesn't contain 0. (b) Doesn't have all additive inverses. (c) Not closed under multiplication (multiply by i, e.g.). (d) Doesn't have all multiplicative inverses.

1.4.4. Since $d \neq 0$, d has a multiplicative inverse $d^{-1} \in \mathbb{F}$. Then

$$x = 1x = (d^{-1}d)x = d^{-1}(dx) = d^{-1}f$$

where the third equality follows from associativity of multiplication in $\mathbb{F}$. Substituting this into the first equation, we obtain

$$ad^{-1}f + by = c.$$

The associativity of multiplication means we can omit the parentheses in $a(d^{-1}f)$ without ambiguity. Then

$$by = 0 + by = (-ad^{-1}f + ad^{-1}f) + by = -ad^{-1}f + (ad^{-1}f + by) = -ad^{-1}f + c,$$

where the third equality follows from associativity of addition in $\mathbb{F}$. Since $b \neq 0$, c has a multiplicative inverse $b^{-1} \in \mathbb{F}$. Then

$$y = 1y = (b^{-1}b)y = b^{-1}(by) = b^{-1}(-ad^{-1}f + c),$$

where the third equality follows from associativity of multiplication in $\mathbb{F}$.

1.4.5. (a) Use Gaussian elimination:

$$\left[\begin{array}{cc|c} 1 & 2-i & 3 \\ -i & 1 & 1+i \end{array}\right] \rightsquigarrow \left[\begin{array}{cc|c} 1 & 2-i & 3 \\ 0 & 2+2i & 1+4i \end{array}\right] \rightsquigarrow \left[\begin{array}{cc|c} 1 & 2-i & 3 \\ 0 & 1 & \frac{5+3i}{4} \end{array}\right] \rightsquigarrow \left[\begin{array}{cc|c} 1 & 0 & \frac{-1-i}{4} \\ 0 & 1 & \frac{5+3i}{4} \end{array}\right]$$

Thus the unique solution is $\frac{1}{4}\begin{bmatrix} -1-i \\ 5+3i \end{bmatrix}$.

(b) The system has RREF $\left[\begin{array}{ccc|c} 1 & 0 & i & 1 \\ 0 & 1 & 1 & i \\ 0 & 0 & 0 & 0 \end{array}\right]$ and hence solutions $x = 1 - iz$, $y = i - z$ for $z \in \mathbb{C}$, or in vector form, $\begin{bmatrix} 1 \\ i \\ 0 \end{bmatrix} + z \begin{bmatrix} -i \\ -1 \\ 1 \end{bmatrix}$ for $z \in \mathbb{C}$.

(c) The system has RREF $\left[\begin{array}{cc|c} 1 & 0 & \frac{27}{13} \\ 0 & 1 & -\frac{6i}{13} \end{array}\right]$, so the unique solution is $\frac{1}{13}\begin{bmatrix} 27 \\ -6i \end{bmatrix}$.

1.4.6. Use Gaussian elimination over $\mathbb{F}_2$:

(a)

$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \end{array}\right] \rightsquigarrow \left[\begin{array}{ccc|c} 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{array}\right] \rightsquigarrow \left[\begin{array}{ccc|c} 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{array}\right]$$

Solving for the pivot variables (x and y) in terms of the free variable (z) gives $x = z + 1$ and $y = 1$.

(b)

$$\left[\begin{array}{ccc|c} 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \end{array}\right] \rightsquigarrow \left[\begin{array}{ccc|c} 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 \end{array}\right] \rightsquigarrow \left[\begin{array}{ccc|c} 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{array}\right] \rightsquigarrow \left[\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{array}\right]$$

The system has the unique solution $x = y = 1, z = 0$.

1.4.7. If $a \neq 0$, then the system can be put into the REF $\left[\begin{array}{cc|c} 1 & a^{-1}b & a^{-1}e \\ 0 & 1 & (ad-bc)^{-1}(af-ce) \end{array}\right]$, and therefore has a unique solution. If $a = 0$, then since $0 \neq ad - bc = bc$, we must have $b, c \neq 0$. Then the system can be put into the REF $\left[\begin{array}{cc|c} 1 & c^{-1}d & c^{-1}f \\ 0 & 1 & b^{-1}e \end{array}\right]$, and therefore has a unique solution.

1.4.8. (a) If the system is consistent (thought of as a system over $\mathbb{C}$), then its RREF does not have a pivot in the last column. But the RREF is the same when the system is thought of as a system over $\mathbb{R}$ (Gaussian elimination doesn't introduce numbers with non-zero imaginary parts if there weren't any to start with), so the system is also consistent when thought of as a system over $\mathbb{R}$.

(b) Again, if the system has a unique solution in $\mathbb{R}^n$, then when Gaussian elimination is performed, there is a pivot in each of the first n columns of the RREF of the augmented matrix. This is then also true if we think of doing the row operations over $\mathbb{C}$, so there is still a unique solution (and we already know that it's in $\mathbb{R}^n$).

(c) The equation $x^2 = -1$ has solutions ($\pm i$) in $\mathbb{C}$, but no solutions in $\mathbb{R}$.

1.4.9. (a) If the system is consistent (thought of as a system over $\mathbb{R}$), then its RREF does not have a pivot in the last column. But the RREF is the same when the system is thought of as a system over $\mathbb{Q}$, so the system is also consistent when thought of as a system over $\mathbb{Q}$.

(b) Again, if the system has a unique solution in $\mathbb{Q}^n$, then when Gaussian elimination is performed, there is a pivot in each of the first n columns of the RREF of the augmented matrix. This is then also true if we think of doing the row operations over $\mathbb{R}$, so there is still a unique solution (and we already know that it's in $\mathbb{Q}^n$).

(c) $2x = 1$.

1.4.10. Multiplying the i^{th} equation by a_{ii} puts the system into REF with a pivot in each of the first n columns and no pivot in the last column, so the system is consistent with a uniques solution.

1.4.11. (a) $\begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} \xrightarrow{\mathbf{R3}} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix} \xrightarrow{\mathbf{R1}} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & -1 \end{bmatrix} \xrightarrow{\mathbf{R1}} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & -2 \end{bmatrix} \xrightarrow{\mathbf{R2}} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{\mathbf{R1}} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

(b) $\begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} \xrightarrow{\mathbf{R3}} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix} \xrightarrow{\mathbf{R1}} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix} \xrightarrow{\mathbf{R1}} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$

1.4.12. (a) The equation $x + y = 0$ over $\mathbb{F}_2$ has the two solutions $x = y = 0$ and $x = y = 1$.

(b) Not possible: see Exercise 1.2.12.

1.4.13. The solution is the same as in Exercise 1.4.8, with $\mathbb{F}$ in place of $\mathbb{R}$ and $\mathbb{K}$ in place of $\mathbb{C}$.

1.4.14. The closure properties hold by definition. The commutative, associative, and distributive properties operations follow immediately from the properties of the usual operations in $\mathbb{Z}$. The numbers 0 and 1 also clearly act as identities.

If $a \in \mathbb{F}_p$, then $p - a$ is an additive identity for a in $\mathbb{F}_p$.

Now suppose that $a, b, c \in \mathbb{F}_p$ are all nonzero and that $ab = ac$ (in $\mathbb{F}_p$). That means that $ab - ac = a(b - c)$ is a multiple of p. Since p is prime this means that either a or $b - c$ is a multiple of p, but since $a, b, c \in \{1, \dots, p-1\}$, this implies that $b - c = 0$, so $b = c$. It follows that if $a \in \mathbb{F}_p$ is nonzero, then the $p - 1$ products

$$a1, a2, \dots, a(p-1)$$

in $\mathbb{F}_p$ are all distint elements. Furthermore, none of them can be a multiple of p, so they are all distinct elements of $\mathbb{F}_p$. Therefore they represent all $p - 1$ nonzero elements of $\mathbb{F}_p$. It follows that one of these products is equal to 1 in $\mathbb{F}_p$. That is, there exists a $b \in \mathbb{F}_p$ such that $ab = 1$ (in $\mathbb{F}_p$).

1.4.15. Mod 4, $2 \cdot 2 = 0$, but $2 \neq 0$. This cannot happen in a field (see part 9 of Theorem 1.5).

1.4.16. Suppose that $a \in \mathbb{F}$ is such that $ab = b$ for all $b \in \mathbb{F}$. Apply this to $b = 1$: $a \cdot 1 = 1$. But 1 is known to be a multiplicative identity, and so $a \cdot 1 = a$; putting these two together, we have $a = 1$.

1.4.17. For $a \in \mathbb{F}$, suppose that $a + b = a + c = 0$. Then

$$c = c + 0 = c + (a + b) = (c + a) + b = 0 + b = b.$$

1.4.18. Let $a \in \mathbb{F}$ with $a \neq 0$. Then $(a^{-1})^{-1}$ is the unique element $b \in \mathbb{F}$ such that $a^{-1}b = 1$. Since $b = a$ is such an element ($a^{-1}a = 1$ by definition of a^{-1}), we have $a = (a^{-1})^{-1}$.

1.4.19. $a + (-1)a = (1 - 1)a = 0a = 0$, so $(-1)a$ is an additive inverse of a.

1.5 Vector spaces

1.5.1. (a) The x axis in $\mathbb{R}^3$ is the set $\left\{ \begin{bmatrix} x \\ 0 \\ 0 \end{bmatrix} \middle| x \in \mathbb{R}^3 \right\}$. This is closed under scalar multiplication and vector addition, and includes the 0 vector $\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$. Therefore it is a subspace.

(b) The first quadrant is not a subspace, because it is not closed under scalar multiplication: -1 times any nonzero vector in the first quadrant is not in the first quadrant.

(c) This is not a subspace, because it is not closed under vector addition: $\begin{bmatrix}1\\2\end{bmatrix}$ and $\begin{bmatrix}-2\\-1\end{bmatrix}$ are both in the set, but

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

is not.

(d) This is not a subspace because it does not contain the 0 vector: $(0,0,0)$ is not a solution of the given linear system. (In fact, the other properties of a subspace also fail, but this is the quickest one to check.)

1.5.2. Suppose that

$$\begin{aligned} a_{11}x_1+\cdots+a_{1n}x_n&=b_1,\\ &\vdots\\ a_{m1}x_1+\cdots+a_{mn}x_n&=b_m \end{aligned}$$

is a non-homogeneous system; that is, there is a $j\in\{1,\dots,m\}$ such that $b_j\neq 0$. Then $(0,\dots,0)$ is not a solution, because if $x_1=\cdots=x_n=0$, then the left-hand side of the jth equation is 0, and the right-hand side is $b_j\neq 0$. Since the set of solutions does not contain the zero vector, it is not a subspace.

1.5.3. If $\mathbf{0}$ is the zero matrix then $\operatorname{tr}\mathbf{0}=0$, so $\mathbf{0}\in S$. If $\mathbf{A},\mathbf{B}\in S$ and $c\in\mathbb{F}$, then $\operatorname{tr}(c\mathbf{A}+\mathbf{B})=\sum_i(ca_{ii}+b_{ii})=c\operatorname{tr}(\mathbf{A})+\operatorname{tr}(\mathbf{B})=0$, and so $c\mathbf{A}+\mathbf{B}\in S$.

1.5.4. (a) The function $f(x)=0$ acts as an additive identity in $C[a,b]$ and lies in V, so V contains 0. If $f,g\in V$, then for any $\alpha,\beta\in\mathbb{R}$, $\alpha f+\beta g$ is a continuous function on $[a,b]$ and $(\alpha f+\beta g)(c)=\alpha f(c)+\beta g(c)=0$, so $\alpha f+\beta g\in V$, and thus V is a subspace of $C[a,b]$.

(b) This set V is not closed under addition: if $f(c)=g(c)=1$, then $(f+g)(c)=2$. So V is not a subspace.

(c) This V is a subspace: it contains 0, and if $f,g\in V$, then for any $\alpha,\beta\in\mathbb{R}$, $\alpha f+\beta g$ is a differentiable function on $[a,b]$ and $(\alpha f+\beta g)'(c)=\alpha f'(c)+\beta g'(c)=0$, so $\alpha f+\beta g\in V$.

(d) This V is a subspace: if $f(x)=0$ on $[a,b]$, then $f'(x)=0$, hence constant. If $f,g\in V$, then for any $\alpha,\beta\in\mathbb{R}$, $\alpha f+\beta g$ is a differentiable function on $[a,b]$ and $(\alpha f+\beta g)'=\alpha f'+\beta g'$, which is constant on $[a,b]$ because f' and g' are.

(e) This set V is not closed under addition: if $f'(c)=g'(c)=1$, then $(f+g)'(c)=2$. So V is not a subspace.

1.5.5. (a) The properties in the definition of a vector space all follow from the field properties of $\mathbb{C}$..

(b) If you multiply a "vector" (element of $\mathbb{Q}$) by a "scalar" (element of $\mathbb{R}$), you don't necessarily get another "vector" (element of $\mathbb{Q}$): e.g., take $v=1$ and $c=\sqrt{2}$.

1.5.6. If V is a complex vector space, then since $\mathbb{R} \subseteq \mathbb{C}$ we can multiply vectors in V by scalars in $\mathbb{R}$. The properties in the definition of a vector space which refer to scalar multiplication are assumed to hold for all scalars in $\mathbb{C}$, so they hold in particular for all scalars in $\mathbb{R}$. Therefore V is also a vector space over $\mathbb{R}$.

1.5.7. The solution is the same as for Exercise 1.5.6, with $\mathbb{F}$ in place of $\mathbb{R}$ and $\mathbb{K}$ in place of $\mathbb{C}$.

1.5.8. Define "vector addition" by (field) addition in $\mathbb{K}$, "scalar multiplication" multiplication in $\mathbb{K}$ by an element of the subfield $\mathbb{F}$, and $0 \in \mathbb{K}$ as a vector additive identity. Then because $\mathbb{K}$ is closed under (field) addition, $\mathbb{K}$ is closed under vector addition. Given $c \in \mathbb{F} \subseteq \mathbb{K}$ and $k \in \mathbb{K}$, $ck \in \mathbb{K}$ since $\mathbb{K}$ is closed under multiplication, so $\mathbb{K}$, as a vector space over $\mathbb{F}$, is closed under scalar multiplication. Since (field) addition in $\mathbb{K}$ is commutative and associative, so are vector space addition. Similarly, 0 is an additive identity for vector addition because it is an additive identity for field addition, and every element of $\mathbb{K}$ has an additive inverse in the field, hence also when thought of as a vector. By definition of $1 \in \mathbb{F} \subseteq \mathbb{K}$, $1k = k$ for any $k \in \mathbb{K}$. If $a, b \in \mathbb{F}$ and $k \in \mathbb{K}$, then $a(bk) = (ab)k$ by associativity of multiplication in $\mathbb{K}$. Finally, if $a, b \in \mathbb{F}$ and $k_1, k_2 \in \mathbb{K}$, then $a(k_1 + k_2) = ak_1 + ak_2$ and $(a + b)k_1 = ak_1 + bk_1$ by the (field) distributive law.

1.5.9. By definition $(f + g)(a) = f(a) + g(a)$ (with the second $+$ in $\mathbb{F}$), so $f + g$ is a funtion from A to $\mathbb{F}$. Commutativity and associativity of vector addition follows from commutativity and associativity of addition in $\mathbb{F}$. The zero vector is the function $z : A \to \mathbb{F}$ with $z(a) = 0$ for each a. Given $c \in \mathbb{F}$, by definition $(cf)(a) = cf(a)$ (with multiplication here in $\mathbb{F}$). The properties of scalar multiplication follow from the properties of multiplication in $\mathbb{F}$.

1.5.10. Since $0 \in U$ and $0 \in W$, $0 \in U \cap W$.

Suppose that $v_1, v_2 \in U \cap W$. Then $v_1, v_2 \in U$, so $v_1 + v_2 \in U$ since U is a subspace. Also $v_1, v_2 \in W$, so $v_1 + v_2 \in W$ since W is a subspace. Thus $v_1 + v_2 \in U \cap W$.

Finally, suppose that $v \in U \cap W$ and $a \in \mathbb{F}$. Then $v \in U$, so $av \in U$ since U is a subspace. Also $v \in W$, so $av \in W$ since U is a subspace. Thus $av \in U \cap W$.

1.5.11. The subspaces U_1 and U_2 both contain 0, so we can take $u_1 = u_2 = 0$ to see $0 = 0 + 0 \in U_1 + U_2$. If $v \in U_1 + U_2$ and $c \in \mathbb{F}$, then $v = u_1 + u_2$ for some $u_1 \in U_1$ and $u_2 \in U_2$, so $cv = c(u_1 + u_2) = cu_1 + cu_2$. Now $cu_1 \in U_1$ and $cu_2 \in U_2$ since U_1 and U_2 are subspaces of V, and it follows that $cv \in U_1 + U_2$. Finally, if $u, v \in U_1 + U_2$, then $u = u_1 + u_2$ and $v = v_1 + v_2$ for some $u_1, v_1 \in U_1$ and $u_2, v_2 \in U_2$, and so $u + v = (u_1 + u_2) + (v_1 + v_2) = (u_1 + v_1) + (u_2 + v_2)$. Now $u_1 + v_1 \in U_1$ and $u_2 + v_2 \in U_2$ since U_1 and U_2 are subspaces of V, and it follows that $u + v \in U_1 + U_2$.

1.5.12. In $\mathbb{F}_2$, $0^2 + 0 = 0 + 0 = 0$ and also $1^2 + 1 = 1 + 1 = 0$.

1.5.13. Addition and scalar multiplication are given by $\begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} + \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix} = \begin{bmatrix} v_1 w_1 \\ \vdots \\ v_n w_n \end{bmatrix}$ and $\lambda \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} = \begin{bmatrix} v_1^\lambda \\ \vdots \\ v_n^\lambda \end{bmatrix}$, which have strictly positive entries. Commutativity and associativity of vector addition follows from commutativity and associativity of (ordinary) multiplication. If the "zero vector" is $\begin{bmatrix}1\\ \vdots\\ 1\end{bmatrix}$, then $\begin{bmatrix}1\\ \vdots\\ 1\end{bmatrix} + \begin{bmatrix}v_1\\ \vdots\\ v_n\end{bmatrix} = \begin{bmatrix}1\cdot v_1\\ \vdots\\ 1\cdot v_n\end{bmatrix} = \begin{bmatrix}v_1\\ \vdots\\ v_n\end{bmatrix}$. The additive inverse of $\begin{bmatrix}v_1\\ \vdots\\ v_n\end{bmatrix}$ is $\begin{bmatrix}\frac{1}{v_1}\\ \vdots\\ \frac{1}{v_n}\end{bmatrix}$. Multiplication by 1 and associativity of multiplication: $1\begin{bmatrix}v_1\\ \vdots\\ v_n\end{bmatrix} = \begin{bmatrix}v_1^1\\ \vdots\\ v_n^1\end{bmatrix} = \begin{bmatrix}v_1\\ \vdots\\ v_n\end{bmatrix}$, and $a\left(b\begin{bmatrix}v_1\\ \vdots\\ v_n\end{bmatrix}\right) = a\begin{bmatrix}v_1^b\\ \vdots\\ v_n^b\end{bmatrix} = \begin{bmatrix}v_1^{ab}\\ \vdots\\ v_n^{ab}\end{bmatrix} = (ab)\begin{bmatrix}v_1\\ \vdots\\ v_n\end{bmatrix}$. Distributive laws: $a\left(\begin{bmatrix}v_1\\ \vdots\\ v_n\end{bmatrix} + \begin{bmatrix}w_1\\ \vdots\\ w_n\end{bmatrix}\right) = a\begin{bmatrix}v_1w_1\\ \vdots\\ v_nw_n\end{bmatrix} = \begin{bmatrix}(v_1w_1)^a\\ \vdots\\ (v_nw_n)^a\end{bmatrix} = \begin{bmatrix}v_1^a\\ \vdots\\ v_n^a\end{bmatrix} + \begin{bmatrix}w_1^a\\ \vdots\\ w_n^a\end{bmatrix} = a\begin{bmatrix}v_1\\ \vdots\\ v_n\end{bmatrix} + a\begin{bmatrix}w_1\\ \vdots\\ w_n\end{bmatrix}$, and $(a+b)\begin{bmatrix}v_1\\ \vdots\\ v_n\end{bmatrix} = \begin{bmatrix}v_1^{a+b}\\ \vdots\\ v_n^{a+b}\end{bmatrix} = \begin{bmatrix}v_1^av_1^b\\ \vdots\\ v_n^av_n^b\end{bmatrix} = \begin{bmatrix}v_1^a\\ \vdots\\ v_n^a\end{bmatrix} + \begin{bmatrix}v_1^b\\ \vdots\\ v_n^b\end{bmatrix} = a\begin{bmatrix}v_1\\ \vdots\\ v_n\end{bmatrix} + b\begin{bmatrix}v_1\\ \vdots\\ v_n\end{bmatrix}$.

1.5.14. It simplifies the calculations below to notice that for $(p_1, \dots, p_n), (q_1, \dots, q_n) \in V$ and $c \in \mathbb{R}$,

$$\begin{aligned}(cp_1, \dots, cp_n) + (q_1, \dots, q_n) &= \frac{(cp_1q_1, \dots, cp_nq_n)}{(cp_1q_1 + \cdots + cp_nq_n)}\\ &= \frac{(p_1q_1, \dots, p_nq_n)}{(p_1q_1 + \cdots + p_nq_n)} = (p_1, \dots, p_n) + (q_1, \dots, q_n),\end{aligned}$$

and

$$\lambda(cp_1, \dots, cp_n) = \frac{(c^\lambda p_1^\lambda, \dots, c^\lambda p_n^\lambda)}{c^\lambda p_1^\lambda + \cdots + c^\lambda p_n^\lambda} = \frac{(p_1^\lambda, \dots, p_n^\lambda)}{p_1^\lambda + \cdots + p_n^\lambda} = \lambda(p_1, \dots, p_n).$$

(i) Given $(p_1, \dots, p_n), (q_1, \dots, q_n) \in V$, for any j, $\frac{p_jq_j}{p_1q_1+\cdots+p_nq_n} > 0$ and $\sum_{j=1}^n \frac{p_jq_j}{p_1q_1+\cdots+p_nq_n} = 1$, thus the sum of two elements of V is an element of V.

(ii) Given $(p_1, \dots, p_n) \in V$ and $\lambda \in \mathbb{R}$, for any j, $\frac{p_j^\lambda}{p_1^\lambda+\cdots+p_n^\lambda} > 0$ and $\sum_{j=1}^n \frac{p_j^\lambda}{p_1^\lambda+\cdots+p_n^\lambda} = 1$, so a scalar multiple of an element of V is an element of V.

(iii) Commutativity of addition is easy to see from the definition.

(iv) Let $(p_1, \dots, p_n), (q_1, \dots, q_n), (r_1, \dots, r_n) \in V$. Then if $c := p_1q_1 + \cdots + p_nq_n$,

$$\begin{aligned}&\Big[(p_1, \dots, p_n) + (q_1, \dots, q_n)\Big] + (r_1, \dots, r_n)\\ &\qquad = \left[\frac{1}{c}(p_1q_1, \dots, p_nq_n)\right] + (r_1, \dots, r_n)\\ &\qquad = \frac{(p_1q_1r_1, \dots, p_nq_nr_n)}{p_1q_1r_1 + \cdots + p_nq_nr_n},\end{aligned}$$

where we used our observation at the beginning of the solution to drop the $\frac{1}{c}$. Similarly, if $d = q_1 r_1 + \cdots + q_n r_n$,

$$\begin{aligned}
(p_1, \ldots, p_n) &+ \Big[(q_1, \ldots, q_n) + (r_1, \ldots, r_n)\Big] \\
&= (p_1, \ldots, p_n) + \left[\frac{1}{d}(q_1 r_1, \ldots, q_n r_n)\right] \\
&= \frac{(p_1 q_1 r_1, \ldots, p_n q_n r_n)}{p_1 q_1 r_1 + \cdots + p_n q_n r_n},
\end{aligned}$$

and associativity follows.

(v) The element $\left(\frac{1}{n}, \ldots, \frac{1}{n}\right)$ functions as an additive identity:

$$\left(\frac{1}{n}, \ldots, \frac{1}{n}\right) + (p_1, \ldots, p_n) = \frac{\left(\frac{p_1}{n}, \ldots, \frac{p_n}{n}\right)}{\frac{p_1}{n} + \cdots + \frac{p_n}{n}} = (p_1, \ldots, p_n),$$

since $p_1 + \cdots + p_n = 1$ by assumption.

(vi) Given $(p_1, \ldots, p_n) \in V$, the element $\frac{(p_1^{-1}, \ldots, p_n^{-1})}{p_1^{-1} + \cdots + p_n^{-1}}$ is an additive inverse: if $c = p_1^{-1} + \cdots + p_n^{-1}$, then

$$(p_1, \ldots, p_n) + \frac{(p_1^{-1}, \ldots, p_n^{-1})}{p_1^{-1} + \cdots + p_n^{-1}} = \frac{\left(\frac{1}{c}, \ldots, \frac{1}{c}\right)}{\frac{n}{c}} = \left(\frac{1}{n}, \ldots, \frac{1}{n}\right).$$

(vii) Given $(p_1, \ldots, p_n) \in V$,

$$1 \cdot (p_1, \ldots, p_n) = \frac{(p_1, \ldots, p_n)}{p_1 + \cdots + p_n},$$

again since $p_1 + \cdots + p_n = 1$ by assumption.

(viii) Given $\lambda, \mu \in \mathbb{R}$ and $(p_1, \ldots, p_n) \in V$, if $c = p_1^{\mu} + \cdots + p_n^{\mu}$,

$$\begin{aligned}
\lambda(\mu(p_1, \ldots, p_n)) &= \lambda\left[\frac{1}{c}\left(p_1^{\mu}, \ldots, p_n^{\mu}\right)\right] \\
&= \frac{\left(p_1^{\lambda\mu}, \ldots, p_n^{\lambda\mu}\right)}{p_1^{\lambda\mu} + \cdots + p_n^{\lambda\mu}} = (\lambda\mu)(p_1, \ldots, p_n).
\end{aligned}$$

(ix) For $\lambda \in \mathbb{R}$ and $(p_1, \ldots, p_n), (q_1, \ldots, q_n) \in V$,

$$\begin{aligned}
\lambda\left[(p_1, \ldots, p_n) + (q_1, \ldots, q_n)\right] &= \lambda\left[\frac{(p_1 q_1, \ldots, p_n q_n)}{p_1 q_1 + \cdots +, p_n q_n}\right] \\
&= \frac{(p_1 q_1)^{\lambda}, \ldots, (p_n q_n)^{\lambda}}{(p_1 q_1)^{\lambda} + \cdots + (p_n q_n)^{\lambda}} \\
&= \left[\frac{(p_1^{\lambda}, \ldots, p_n^{\lambda})}{p_1^{\lambda} + \cdots + p_n^{\lambda}}\right] + \left[\frac{(q_1^{\lambda}, \ldots, q_n^{\lambda})}{q_1^{\lambda} + \cdots + q_n^{\lambda}}\right].
\end{aligned}$$

(x) For $\lambda, \mu \in \mathbb{R}$ and $(p_1, \ldots, p_n) \in V$,

$$
\begin{aligned}
(\lambda+\mu)(p_1, \ldots, p_n) &= \frac{(p_1^{\lambda+\mu}, \ldots, p_n^{\lambda+\mu})}{p_1^{\lambda+\mu} + \cdots + p_n^{\lambda+\mu}} \\
&= \frac{(p_1^{\lambda}, \ldots, p_n^{\lambda})}{p_1^{\lambda} + \cdots + p_n^{\lambda}} + \frac{(p_1^{\mu}, \ldots, p_n^{\mu})}{p_1^{\mu} + \cdots + p_n^{\mu}} = \lambda(p_1, \ldots, p_n) + \mu(p_1 \ldots, p_n).
\end{aligned}
$$

1.5.15. We write $+$ in place of Δ below so as to avoid writing out additional steps in which the only change would be to replace one symbol with the other. It is clear from the definition that the sum of two vectors and the product of a scalar and a vector are vectors.

Vector addition is commutative: By definition $A+B = B+A$, so addition is commutative.

Vector addition is associative: Suppose $A, B, C \subseteq S$. We need to show that

$$(A+B)+C = A+(B+C).$$

An element $x \in (A+B)+C$ if and only if $x \in A+B$ or $x \in C$, but not both. Further, $x \in A+B$ are in either A or B but not both. Thus $x \in (A+B)+C$ if and only if x is in one and only one of A, B, and C. In exactly the same way we check that $x \in A+(B+C)$ if and only if x is in one and only one of A, B, and C. Therefore the sets $(A+B)+C$ and $A+(B+C)$ have the same elements, so they are equal.

Additive identity: For any $A \subseteq S$, $A+\emptyset$ consists of all x which are in A or in $\emptyset$ but not both. Since x is never in $\emptyset$, $A+\emptyset = A$. Thus the empty set is the zero vector in this space.

Additive inverses: For any $A \subseteq S$, $A+A$ consists of all x which are in A or in A, but are not in A and in A. Since (obviously) every x in A is in A, there are no such x. Therefore $A+A = \emptyset$, and so $-A = A$.

Multiplication by 1: This follows immediately from the definition of scalar multiplication here.

Scalar multiplication is "associative": Let $A \subseteq S$. Then

$$
\begin{aligned}
(0 \cdot 0)A &= 0A = \emptyset = 0\emptyset = 0(0A), \\
&= 0A + 0A, \\
(0 \cdot 1)A &= 0A = 0(1A), \\
(1 \cdot 1)A &= 1A = 1(1A).
\end{aligned}
$$

Distributive law #1: Suppose $A, B \subseteq S$. Then by the definition of scalar multiplication here,

$$1(A+B) = A+B = 1A+1B$$

and (using A2)

$$0(A+B) = \emptyset = \emptyset + \emptyset = 0A + 0B.$$

Distributive law #2: Let $A \subseteq S$. Then, using A2 and A3,

$$
\begin{aligned}
(0+0)A &= 0A = \emptyset = \emptyset + \emptyset = 0A + 0A,\\
(0+1)A &= 1A = A = A + \emptyset = 1A + 0A,\\
(1+1)A &= 0A = \emptyset = A + A = 1A + 1A.
\end{aligned}
$$

1.5.16. Suppose that 0 and $0'$ are both additive identities in V. Then $0 + 0' = 0'$ since 0 is an additive identity, and also $0 + 0' = 0$ since $0'$ is an additive identity. Therefore $0 = 0'$.

1.5.17. Suppose that u and w are both additive inverses of v. Then $w + v = u + v = 0$, and so

$$w = 0 + w = (u + v) + w = u + (v + w) = u + 0 = u.$$

1.5.18. Since $v + (-v) = 0$ by definition of $-v$, v acts as an additive inverse to $-v$: $v = -(-v)$.

1.5.19. By the distributive law for scalar multiplication, $0v = (0 + 0)v = 0v + 0v$. Therefore

$$0 = 0v - 0v = 0v + 0v - 0v = 0v.$$

1.5.20. Let $v \in V$. Then

$$(-1)v + v = (-1 + 1)v = 0v = 0,$$

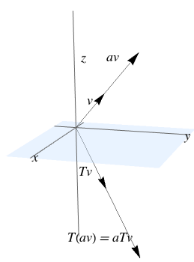
where the first step follows from distributive law #2 in the definition of a vector space, and the last step is by part 4. of Theorem 1.11. Since $(-1)v + v = 0$, (that is, $(-1)v$ acts as an additive inverse to v), $(-1)v = -v$.

Chapter 2

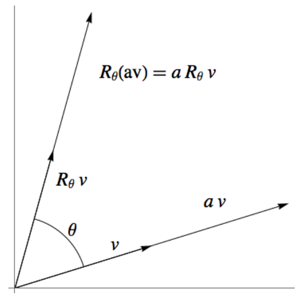
Linear maps and matrices

2.1 Linear maps

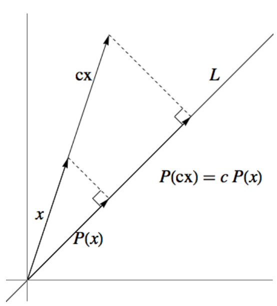
2.1.1. (a)

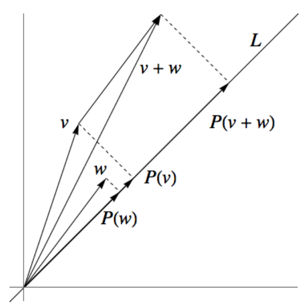


(b)



2.1.2.





2.1.3. (a) $\begin{bmatrix} 9 \\ 37 \end{bmatrix}$ (b) not possible (c) $\begin{bmatrix} 2 + \frac{2\pi}{3} + 2\sqrt{2} \\ \frac{138}{7} \end{bmatrix}$ (d) 70 (e) $\begin{bmatrix} 8.03 \\ 7.94 \\ -1.11 \end{bmatrix}$ (f) $\begin{bmatrix} -7 + 3i \\ -5 + 11i \end{bmatrix}$

2.1.4. (a) not possible (b) $\begin{bmatrix} 12 \\ -22 \\ -52 \end{bmatrix}$ (c) $[32]$ (d) $\begin{bmatrix} 0 \\ 3 \\ -2 \end{bmatrix}$ (e) $\begin{bmatrix} 3.44 \\ 7.28 \end{bmatrix}$ (f) $\begin{bmatrix} 7 \\ 4 - 2i \end{bmatrix}$

2.1.5. For continuous functions f and g, and $a \in \mathbb{R}$, $[\boldsymbol{T}(af + g)](x) = (af + g)(x)\cos(x) = [af(x) + g(x)]\cos(x) = af(x)\cos(x) + g(x)\cos(x) = a[\boldsymbol{T}f](x) + [\boldsymbol{T}g](x)$.

2.1.6. Let $f, g \in C[0,1]$ and let $\alpha, \beta \in \mathbb{R}$. Then

$$\boldsymbol{T}(\alpha f + \beta g) = \int_0^1 (\alpha f(x) + \beta g(x))\,dx = \alpha \int_0^1 f(x)\,dx + \beta \int_0^1 g(x)\,dx = \alpha \boldsymbol{T}(f) + \beta \boldsymbol{T}(g),$$

by properties of the integral (normally referred to as "linearity of the integral"!).

2.1.7. $\begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 1 \\ 3 & 4 & 1 & 2 \\ 4 & 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \end{bmatrix} = \begin{bmatrix} -2 \\ 2 \\ -2 \\ 2 \end{bmatrix} = -2 \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \end{bmatrix}$, so -2 is the eigenvalue.

2.1.8. $\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 2 & 2 \\ 1 & 2 & 3 & 3 \\ 1 & 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 0 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 0 \\ -1 \end{bmatrix} = 1 \begin{bmatrix} 1 \\ 1 \\ 0 \\ -1 \end{bmatrix}$, so 1 is the eigenvalue.

2.1.9. (a) We wish to know for which λ the system $\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \lambda \begin{bmatrix} x \\ y \end{bmatrix}$ has nonzero solutions, and what they are. We subtract $\lambda \begin{bmatrix} x \\ y \end{bmatrix}$ from both sides and perform Gaussian elimination:

$$\left[\begin{array}{cc|c} 1-\lambda & 1 & 0 \\ 1 & 1-\lambda & 0 \end{array}\right] \rightsquigarrow \left[\begin{array}{cc|c} 1 & 1-\lambda & 0 \\ 0 & 1-(1-\lambda)^2 & 0 \end{array}\right].$$

So for the system to have any nonzero solutions, we must have

$$1-(1-\lambda)^2 = 0.$$

Solving for λ yields that the eigenvalues are $\lambda = 0, 2$.
The system above also shows that if λ is an eigenvalue, then any eigenvector must satisfy $x = -(1-\lambda)y$.
So if $\lambda = 0$, then any solution to the system is of the form $\begin{bmatrix} -1 \\ 1 \end{bmatrix} y$ with $y \neq 0$, and so the set of all eigenvectors corresponding to this eigenvalue is $\left\langle \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right\rangle \setminus \{\mathbf{0}\}$. Similarly, if $\lambda = 2$, then any solution to the system is of the form $\begin{bmatrix} 1 \\ 1 \end{bmatrix} y$ with $y \neq 0$, and so the set of all eigenvectors corresponding to this eigenvalue is $\left\langle \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\rangle \setminus \{\mathbf{0}\}$.

(b) The remaining parts can be solved by the same procedure.
-2 is an eigenvalue with corresponding eigenvectors $\left\langle \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right\rangle \setminus \{\mathbf{0}\}$.
2 is an eigenvalue with corresponding eigenvectors $\left\langle \begin{bmatrix} \frac{3}{5} + \frac{4}{5}i \\ 1 \end{bmatrix} \right\rangle \setminus \{\mathbf{0}\}$.

(c) 2 is the only eigenvalue, with eigenvectors $\left\langle \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\rangle \setminus \{\mathbf{0}\}$.

(d) In this case the system is equivalent to

$$\left[\begin{array}{ccc|c} 1 & 0 & -\lambda & 0 \\ 0 & 2+\lambda & 0 & 0 \\ 0 & 0 & 1-\lambda^2 & 0 \end{array}\right].$$

This system has nonzero solutions if either $2+\lambda=0$ or $1-\lambda^2=0$, so the eigenvalues are -2, -1, and 1.
Eigenvectors corresponding to $\lambda=-2$ are the nonzero solutions to the system

$$\left[\begin{array}{ccc|c} 1 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -3 & 0 \end{array}\right] \rightsquigarrow \left[\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right],$$

which are $\left\langle \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\rangle \setminus \{\mathbf{0}\}$.

Eigenvectors corresponding to $\lambda=-1$ are the nonzero solutions to the system $\left[\begin{array}{ccc|c} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$,

which are $\left\langle \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \right\rangle \setminus \{\mathbf{0}\}$.

Eigenvectors corresponding to $\lambda=1$ are the nonzero solutions to the system $\left[\begin{array}{ccc|c} 1 & 0 & -1 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$,

which are $\left\langle \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \right\rangle \setminus \{\mathbf{0}\}$.

2.1.10. (a) We wish to know for which λ the system $\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \lambda \begin{bmatrix} x \\ y \end{bmatrix}$ has nonzero solutions, and what they are. We subtract $\lambda \begin{bmatrix} x \\ y \end{bmatrix}$ from both sides and perform Gaussian elimination:

$$\left[\begin{array}{cc|c} 1-\lambda & 2 & 0 \\ 3 & 4-\lambda & 0 \end{array}\right] \rightsquigarrow \left[\begin{array}{cc|c} 1 & \frac{1}{3}(4-\lambda) & 0 \\ 0 & 2-\frac{1}{3}(1-\lambda)(4-\lambda) & 0 \end{array}\right].$$

So for the system to have any nonzero solutions, we must have

$$2-\frac{1}{3}(1-\lambda)(4-\lambda)=0.$$

Solving for λ yields that the eigenvalues are $\lambda=\frac{5\pm\sqrt{33}}{2}$.
The system above also shows that if λ is an eigenvalue, then any eigenvector must satisfy $x=-\frac{1}{3}(4-\lambda)y$.
So if $\lambda=\frac{5+\sqrt{33}}{2}$, then any solution to the system is of the form $\begin{bmatrix} \frac{-3+\sqrt{33}}{6} \\ 1 \end{bmatrix} y$ with $y \neq 0$, and so the set of all eigenvectors corresponding to this eigenvalue is $\left\langle \begin{bmatrix} \frac{-3+\sqrt{33}}{6} \\ 1 \end{bmatrix} \right\rangle \setminus \{\mathbf{0}\}$. Similarly, if $\lambda=\frac{5-\sqrt{33}}{2}$, then any solution to the system is of the form $\begin{bmatrix} \frac{-3-\sqrt{33}}{6} \\ 1 \end{bmatrix} y$

with $y \neq 0$, and so the set of all eigenvectors corresponding to this eigenvalue is $\left\langle \begin{bmatrix} \frac{-3-\sqrt{33}}{6} \\ 1 \end{bmatrix} \right\rangle \setminus \{\mathbf{0}\}$.

(b) The remaining parts can be solved by the same procedure.

i is an eigenvalue with corresponding eigenvectors $\left\langle \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\rangle \setminus \{\mathbf{0}\}$.

1 is an eigenvalue with corresponding eigenvectors $\left\langle \begin{bmatrix} 2+i \\ 1-i \end{bmatrix} \right\rangle \setminus \{\mathbf{0}\}$.

(c) $\frac{3}{2} + \frac{\sqrt{7}}{2}i$ is an eigenvalue with corresponding eigenvectors $\left\langle \begin{bmatrix} \frac{1}{2} - \frac{\sqrt{7}}{2}i \\ 1 \end{bmatrix} \right\rangle \setminus \{\mathbf{0}\}$.

$\frac{3}{2} - \frac{\sqrt{7}}{2}i$ is an eigenvalue with corresponding eigenvectors $\left\langle \begin{bmatrix} \frac{1}{2} + \frac{\sqrt{7}}{2}i \\ 1 \end{bmatrix} \right\rangle \setminus \{\mathbf{0}\}$.

(d) In this case the system is equivalent to

$$\left[\begin{array}{ccc|c} 1 & 1 & -\lambda & 0 \\ 0 & 1+\lambda & 1-\lambda^2 & 0 \\ 0 & 0 & 2+\lambda-\lambda^2 & 0 \end{array}\right].$$

This system has nonzero solutions if either $\lambda^2 - \lambda - 2 = 0$ or $\lambda + 1 = 0$, so the eigenvalues are 2 and -1.

Eigenvectors corresponding to $\lambda = 2$ are the nonzero solutions to the system

$$\left[\begin{array}{ccc|c} 1 & 1 & -2 & 0 \\ 0 & 3 & -3 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] \rightsquigarrow \left[\begin{array}{ccc|c} 1 & 0 & -1 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right],$$

which are $\left\langle \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\rangle \setminus \{\mathbf{0}\}$.

Eigenvectors corresponding to $\lambda = -1$ are the nonzero solutions to the system $\left[\begin{array}{ccc|c} 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$,

which are $\left\langle \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \right\rangle \setminus \{\mathbf{0}\}$.

2.1.11. If $\mathbf{x}$ lies on the line L, then $\boldsymbol{P}(\mathbf{x}) = \mathbf{x}$, and so $\mathbf{x}$ is an eigenvector of $\boldsymbol{P}$ with eigenvalue 1.

If $\mathbf{x}$ lies on the line perpendicular to L, then $\boldsymbol{P}(\mathbf{x}) = \mathbf{0}$, and so $\mathbf{x}$ is an eigenvector of $\boldsymbol{P}$ with eigenvalue 0.

If $\mathbf{x}$ is any other vector, then $\mathbf{x}$ and $\boldsymbol{P}(\mathbf{x})$ do not point in the same direction, so $\mathbf{x}$ is not an eigenvector. Therefore the eigenvalues and eigenvectors described above are the only ones.

2.1.12. Suppose $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$. Each entry of $\mathbf{A}\mathbf{x}$ is equal to $x_1 + \cdots + x_n$, so $\lambda x_i = x_1 + \cdots + x_n$ for each i. Since $\lambda x_1 = \lambda x_2 = \cdots = \lambda x_n$, either $\lambda = 0$ or else $x_1 = x_2 = \cdots = x_n$.

If $\lambda = 0$, then $x_1 + \cdots + x_n = 0$. On the other hand, if $x_1 + \cdots + x_n = 0$, then $\mathbf{A}\mathbf{x} = \mathbf{0} = 0\mathbf{x}$. Therefore 0 is indeed an eigenvalue of $\mathbf{A}$, and the corresponding eigenvectors are precisely those nonzero vectors $\mathbf{x}$ with $x_1 + \cdots + x_n$.

If $\lambda \neq 0$, then $x_1 = \cdots = x_n$, so $x_1 + \cdots + x_n = nx_i$ for each i. Therefore $\lambda x_i = nx_i$ for each i, and if $\mathbf{x} \neq \mathbf{0}$ this implies that $\lambda = n$. Therefore the only nonzero eigenvalue of $\mathbf{A}$ is n, and the corresponding eigenvectors are precisely those nonzero $\mathbf{x}$ whose entries are all equal; that is, the scalar multiples of $\begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$.

2.1.13. By assumption, $\mathbf{A}\mathbf{x} = \mathbf{b}$ and $\mathbf{A}\mathbf{y} = \mathbf{b}$, so

$$\mathbf{A}(\mathbf{x} + \mathbf{y}) = \mathbf{A}\mathbf{x} + \mathbf{A}\mathbf{y} = \mathbf{b} + \mathbf{b}.$$

So $\mathbf{A}(\mathbf{x} + \mathbf{y}) = \mathbf{b}$ if and only if

$$\mathbf{b} = \mathbf{b} + \mathbf{b},$$

which is true if and only if $\mathbf{b} = \mathbf{0}$.

2.1.14. By assumption, for each j there is a scalar $d_j \in \mathbb{F}$ such that $\mathbf{A}\mathbf{e}_j = d_j \mathbf{e}_j$. This means that the j^{th} column of $\mathbf{A}$ is $d_j \mathbf{e}_j$. Therefore, $a_{jj} = d_j$ and every other entry in the j^{th} column is 0. Since this is true for each j, $\mathbf{A}$ is diagonal.

2.1.15. As in Exercise 2.1.14, $\mathbf{A}$ must be diagonal, say $\mathbf{A} = \mathbf{diag}(d_1, \ldots, d_n)$. Now since $\mathbf{u} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$ is an eigenvector of $\mathbf{A}$, there exists a $\lambda \in \mathbb{F}$ such that $\mathbf{A}\mathbf{u} = \lambda \mathbf{u}$, which implies that $d_i = \lambda$ for each i.

2.1.16. Since λ is an eigenvalue of $\boldsymbol{ST}$, there is a vector $v \in V$ with $v \neq 0$ and $\boldsymbol{ST}v = \lambda v$. Taking $\boldsymbol{T}$ of both sides gives

$$\boldsymbol{T}(\boldsymbol{ST}v) = \boldsymbol{TS}(\boldsymbol{T}v) = \boldsymbol{T}(\lambda v) = \lambda \boldsymbol{T}(v),$$

where the first equality is by associativity of function composition, and the last is because $\boldsymbol{T}$ is linear. This looks like $\boldsymbol{T}v$ is an eigenvector of $\boldsymbol{TS}$ with eigenvalue λ, so we'd be done. But we need to check that $\boldsymbol{T}v \neq 0$, because 0 is never an eigenvector. Going back to the equation $\boldsymbol{ST}v = \lambda v$, if $\boldsymbol{T}v = 0$ then $\boldsymbol{S}(\boldsymbol{T}v) = 0$, which would make $\lambda = 0$ (since we know $v \neq 0$). But we're assuming $\lambda \neq 0$, so this is impossible. That means $\boldsymbol{T}v$ does count as an eigenvector of $\boldsymbol{TS}$, and it has eigenvalue λ.

2.1.17. (a) Let $a = (a_1, a_2, \ldots)$ and $b = (b_1, b_2, \ldots)$ be sequences in $\mathbb{F}^\infty$ and $c \in \mathbb{F}$. Then

$$\begin{aligned} \boldsymbol{S}(a + b) &= \boldsymbol{S}((a_1 + b_1, a_2 + b_2, \ldots)) = (0, a_1 + b_1, a_2 + b_2, \ldots) \\ &= (0, a_1, a_2, \ldots) + (0, b_1, b_2, \ldots) = \boldsymbol{S}(a) + \boldsymbol{S}(b) \end{aligned}$$

and

$$\boldsymbol{S}(ca) = \boldsymbol{S}((ca_1, ca_2, \ldots)) = (0, ca_1, ca_2, \ldots) = c(0, a_1, a_2, \ldots) = c\boldsymbol{S}(a).$$

(b) Suppose λ is an eigenvalue with corresponding eigenvector $a = (a_1, a_2, \dots)$. Then $\boldsymbol{S}(a) = \lambda a$ implies that $\lambda a_1 = 0$ and $\lambda a_i = a_{i-1}$ for every $i \geq 2$. If $\lambda = 0$ this immediately implies that $a_i = 0$ for every i. On the other hand, if $\lambda \neq 0$, then it follows by induction that $a_i = 0$ for every i. In either case this contradicts the fact that the eigenvector a must be nonzero.

2.2 More on linear maps

2.2.1. (a) Yes, $\boldsymbol{T}$ is linear. The verification is straightforward.

(b) No, $\boldsymbol{T}$ is not injective. For example,

$$\boldsymbol{T}\left(\begin{bmatrix}1\\0\\0\end{bmatrix}\right) = \boldsymbol{T}\left(\begin{bmatrix}0\\0\\0\end{bmatrix}\right).$$

(c) No, $\boldsymbol{T}$ is not surjective. There is no $\begin{bmatrix}x\\y\\z\end{bmatrix} \in \mathbb{R}^3$ such that

$$\boldsymbol{T}\left(\begin{bmatrix}x\\y\\z\end{bmatrix}\right) = \begin{bmatrix}0\\0\\1\end{bmatrix}.$$

2.2.2. The augmented matrix of the system has RREF

$$\left[\begin{array}{cccc|c}1 & -1 & 0 & 3 & 0\\0 & 0 & 1 & 2 & 0\\0 & 0 & 0 & 0 & 0\end{array}\right],$$

which leads to the general expression of a solution as

$$\begin{bmatrix}w\\x\\y\\z\end{bmatrix} = \begin{bmatrix}x-3z\\x\\-2z\\z\end{bmatrix} = x\begin{bmatrix}1\\1\\0\\0\end{bmatrix} + z\begin{bmatrix}-3\\0\\-2\\1\end{bmatrix};$$

that is, the space of solutions is

$$U = \left\langle \begin{bmatrix}1\\1\\0\\0\end{bmatrix}, \begin{bmatrix}-3\\0\\-2\\1\end{bmatrix} \right\rangle.$$

Define $\boldsymbol{T} : \mathbb{R}^2 \to U$ by

$$\boldsymbol{T}\left(\begin{bmatrix}x\\y\end{bmatrix}\right) = x\begin{bmatrix}1\\1\\0\\0\end{bmatrix} + y\begin{bmatrix}-3\\0\\-2\\1\end{bmatrix}.$$

It is straightforward to verify that $\boldsymbol{T}$ is linear. It is surjective by the definition of U. It is injective because if

$$\begin{bmatrix} x_1 - 3y_1 \\ x_1 \\ -2y_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} x_2 - 3y_2 \\ x_2 \\ -2y_2 \\ y_2 \end{bmatrix},$$

then $x_1 = x_2$ (by comparing the second components) and $y_1 = y_2$ (by comparing the fourth components).

2.2.3. Begin by solving the given sytem via Gaussian elimination:

$$\left[\begin{array}{cccc|c} 1 & i & -2 & 0 & 1+i \\ -1 & 1 & 0 & 3i & 2 \end{array}\right] \rightsquigarrow \left[\begin{array}{cccc|c} 1 & i & -2 & 0 & 1+i \\ 0 & 1+i & -2 & 3i & 3+i \end{array}\right] \rightsquigarrow \left[\begin{array}{cccc|c} 1 & 0 & -1+i & \frac{3}{2} - \frac{3}{2}i & -i \\ 0 & 1 & -1+i & \frac{3}{2} + \frac{3}{2}i & 2-i \end{array}\right].$$

Solving for the pivot variables x and y in terms of the free variables z and w gives that the set of all solutions of the system is $\left\{ \begin{bmatrix} -i \\ 2-i \\ 0 \\ 0 \end{bmatrix} + a \begin{bmatrix} 1-i \\ 1-i \\ 1 \\ 0 \end{bmatrix} + b \begin{bmatrix} -3+3i \\ -3-3i \\ 0 \\ 2 \end{bmatrix} : a, b \in \mathbb{R} \right\}$. Define the map $\boldsymbol{T}$ by

$$\boldsymbol{T}((z_1, z_2)) = \begin{bmatrix} -i \\ 2-i \\ 0 \\ 0 \end{bmatrix} + z_1 \begin{bmatrix} 1-i \\ 1-i \\ 1 \\ 0 \end{bmatrix} + z_2 \begin{bmatrix} -3+3i \\ -3-3i \\ 0 \\ 2 \end{bmatrix}.$$

It is easy to see that $\boldsymbol{T}$ is linear, and by the expression above for the set of all solutions to the system, $\boldsymbol{T}$ is surjective. To see that $\boldsymbol{T}$ is injective, suppose that $\boldsymbol{T}((z_1, z_2)) = \boldsymbol{T}((w_1, w_2))$. Then $(z_1 - w_1, z_2 - w_2)$ is a solution to the linear system

$$\left[\begin{array}{cc|c} 1-i & -3+3i & 0 \\ 1-i & -3-3i & 0 \\ 1 & 0 & 0 \\ 0 & 2 & 0 \end{array}\right] \rightsquigarrow \left[\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{array}\right],$$

and since $(0, 0)$ is the only solution to this system, $z_1 = w_1$ and $z_2 = w_2$; that is, $\boldsymbol{T}$ is injective.

2.2.4. Let $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ be the standard basis vectors in $\mathbb{R}^3$. If $\boldsymbol{R}$ denotes reflection across the x-y plane in $\mathbb{R}^3$, then $\boldsymbol{R}\mathbf{e}_1 = \mathbf{e}_1$ and $\boldsymbol{R}\mathbf{e}_2 = \mathbf{e}_2$, but $\boldsymbol{R}\mathbf{e}_3 = -\mathbf{e}_1$. The matrix of $\boldsymbol{R}$ is thus

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

2.2.5. From the formula in Exercise 2.2.1,

$$\boldsymbol{T}\mathbf{e}_1 = \mathbf{0} \qquad \boldsymbol{T}\mathbf{e}_2 = \mathbf{e}_1 \qquad \boldsymbol{T}\mathbf{e}_3 = \mathbf{e}_2,$$

and so the matrix of $\boldsymbol{T}$ is $\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$.

2.2.6. If $\boldsymbol{P}$ denotes orthogonal projection onto $y = x$, then $\boldsymbol{P}\mathbf{e}_1 = \boldsymbol{P}\mathbf{e}_2 = \frac{1}{2}\begin{bmatrix}1\\1\end{bmatrix}$, so the matrix of $\boldsymbol{P}$ is $\frac{1}{2}\begin{bmatrix}1 & 1\\1 & 1\end{bmatrix}$.

2.2.7. The operator $\boldsymbol{T}$ acts on $\mathbf{e}_1$ as:

$$\mathbf{e}_1 \mapsto -\mathbf{e}_1 \mapsto -\frac{1}{\sqrt{2}}\begin{bmatrix}1\\1\end{bmatrix} \mapsto -\frac{1}{\sqrt{2}}\begin{bmatrix}1\\2\end{bmatrix}$$

and on $\mathbf{e}_2$ as

$$\mathbf{e}_2 \mapsto \mathbf{e}_2 \mapsto \frac{1}{\sqrt{2}}\begin{bmatrix}-1\\1\end{bmatrix} \mapsto \frac{1}{\sqrt{2}}\begin{bmatrix}-1\\2\end{bmatrix},$$

and so the matrix of $\boldsymbol{T}$ is $\begin{bmatrix}-\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}}\\ -\sqrt{2} & \sqrt{2}\end{bmatrix}$.

2.2.8. If $\boldsymbol{T}$ has 0 as an eigenvalue, then there is some $v \neq 0$ such that $\boldsymbol{T}v = 0$. Since $\boldsymbol{T}(0) = 0$ automatically, we have that $\boldsymbol{T}$ cannot be injective.

2.2.9. If v is an eigenvector of $\boldsymbol{T}$ with eigenvalue λ, then $\boldsymbol{T}v = \lambda v$; applying $\boldsymbol{T}^{-1}$ to both sides (and using that $\boldsymbol{T}^{-1}$ is necessarily linear) gives that $v = \lambda\boldsymbol{T}^{-1}v$. By Exercise 2.2.8, $\lambda \neq 0$ since $\boldsymbol{T}$ is invertible, so we may divide both sides by λ to get that $\boldsymbol{T}^{-1}v = \lambda^{-1}v$. That is, v is an eigenvector of $\boldsymbol{T}^{-1}$ with eigenvalue λ^{-1}.

2.2.10. For each $t \in \mathbb{R}$,

$$\boldsymbol{T}\big((1-t)\mathbf{x} + t\mathbf{y}\big) = (1-t)\boldsymbol{T}(\mathbf{x}) + t\boldsymbol{T}(\mathbf{y})$$

since $\boldsymbol{T}$ is linear. Therefore

$$\boldsymbol{T}(L) := \{(1-t)\boldsymbol{T}(\mathbf{x}) + t\boldsymbol{T}(\mathbf{y}) \mid 0 \le t \le 1\},$$

which is the line segment between $\boldsymbol{T}(\mathbf{x})$ and $\boldsymbol{T}(\mathbf{y})$.

Remark: Notice that according to our definition, a single point $\mathbf{x} \in \mathbb{R}^n$ is a line segment: it is the line segment between $\mathbf{x}$ and itself. So even if it happens that $\boldsymbol{T}(\mathbf{x}) = \boldsymbol{T}(\mathbf{y})$, $\boldsymbol{T}(L)$ still counts as a line segment.

2.2.11. The line segments making up the sides get mapped to line segments by Exercise 2.2.10, so the image of the unit square is a quadrilateral (with one corner at the origin). Call the right-hand side of the square R and the left-hand side L; then $L = \left\{t\begin{bmatrix}0\\1\end{bmatrix} : t \in [0,1]\right\}$ and $R = \left\{\begin{bmatrix}1\\0\end{bmatrix} + t\begin{bmatrix}0\\1\end{bmatrix} : t \in [0,1]\right\}$. This means $\boldsymbol{T}(R) = \left\{\boldsymbol{T}\left(\begin{bmatrix}1\\0\end{bmatrix}\right) + t\boldsymbol{T}\begin{bmatrix}0\\1\end{bmatrix} : t \in [0,1]\right\} = \boldsymbol{T}\left(\begin{bmatrix}1\\0\end{bmatrix}\right) + \boldsymbol{T}(L)$, and both $\boldsymbol{T}(R)$ and $\boldsymbol{T}(L)$ have direction $\boldsymbol{T}\begin{bmatrix}0\\1\end{bmatrix}$. The same argument shows that the images of the top and bottom are parallel.

2.2.12. Let $u \in U$. Then by linearity of $\boldsymbol{S}$,

$$\boldsymbol{S} \circ (c\boldsymbol{T})(u) = \boldsymbol{S}(c\boldsymbol{T}u) = c\boldsymbol{S}(\boldsymbol{T}u) = c(\boldsymbol{S}\boldsymbol{T})u.$$

2.2.13. (a) Computing directly,

$$\begin{bmatrix} -1 & \frac{3}{2} \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \end{bmatrix} = (-1) \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

Thus $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ is an eigenvector of $\boldsymbol{T}$ with eigenvalue -1. Similarly,

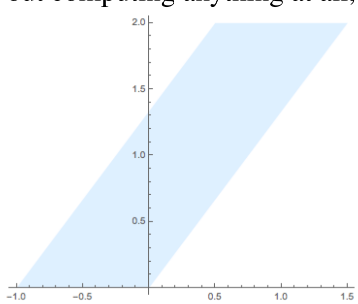
$$\begin{bmatrix} -1 & \frac{3}{2} \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} -1 \cdot 1 + \frac{3}{2} \cdot 2 \\ 1 \cdot 0 + 2 \cdot 2 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \end{bmatrix} = 2 \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

Thus $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$ is an eigenvector of $\boldsymbol{T}$ with eigenvalue 2.

(b) We compute (or just read off from the matrix)

$$\boldsymbol{T}\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} -1 \\ 0 \end{bmatrix}, \quad \boldsymbol{T}\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} \frac{3}{2} \\ 2 \end{bmatrix}, \quad \boldsymbol{T}\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} -1 \\ 0 \end{bmatrix} + \begin{bmatrix} \frac{3}{2} \\ 2 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} \\ 2 \end{bmatrix}.$$

Thus the desired image is the parallelogram with vertices $(0,0)$ (we know this one without computing anything at all, just by linearity), $(-1,0)$, $(3/2,2)$, and $(1/2,2)$:

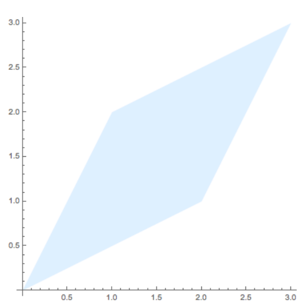


2.2.14. (a)

$$\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \end{bmatrix} \qquad \text{and} \qquad \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix},$$

so $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ is an eigenvector with eigenvalue 3 and $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$ is an eigenvector with eigenvalue 1.

(b)



2.2.15. For any $\lambda \in \mathbb{R}$, if $f(x) = e^{\lambda x}$, then $Df(x) = \lambda e^{\lambda x} = \lambda f(x)$, so f is an eigenvector with eigenvalue λ.

2.2.16. (a) Let $f, g \in C[0,\infty)$ and let $c \in \mathbb{R}$. Then

$$\boldsymbol{T}(cf+g)(x) = \int_0^x (cf+g)(y)\,dy = c\int_0^x f(y)\,dy + \int_0^x g(y)\,dy = c[\boldsymbol{T}f](x) + [\boldsymbol{T}g](x).$$

(b) Let

$$k(x,y) := \begin{cases} 1, & y \le x; \\ 0, & y > x. \end{cases}$$

Then

$$\boldsymbol{T}f(x) = \int_0^x f(y)\,dy = \int_0^\infty k(x,y)f(y)\,dy,$$

so we can think of k as the (discontinuous) kernel for $\boldsymbol{T}$.

2.2.17. (a) It's easiest to approach this by proving the contrapositive: if $\boldsymbol{T}$ is not injective, there are $u_1 \neq u_2$ such that $\boldsymbol{T}u_1 = \boldsymbol{T}u_2$, and so

$$\boldsymbol{ST}u_1 = \boldsymbol{ST}u_2,$$

and hence $\boldsymbol{ST}$ is not injective.

(b) If $\boldsymbol{ST}$ is surjective, then every $w \in W$ can be written $w = \boldsymbol{ST}u = \boldsymbol{S}(\boldsymbol{T}u)$ for some u. This shows that $w = \boldsymbol{S}v$ for $v = \boldsymbol{T}u$.

2.2.18. (a) Let $\mathbf{A}, \mathbf{B} \in \mathrm{M}_{m,n}(\mathbb{F})$ and let $c \in \mathbb{F}$. Then for any $\mathbf{v} \in \mathbb{F}^n$,

$$D(c\mathbf{A}+\mathbf{B})\mathbf{v} = (c\mathbf{A}+\mathbf{B})\mathbf{v} = c\mathbf{A}\mathbf{v}+\mathbf{B}\mathbf{v} = cD(\mathbf{A})\mathbf{v}+D(\mathbf{B})\mathbf{v} = (cD(\mathbf{A})+D(\mathbf{B}))\mathbf{v},$$

so $D(c\mathbf{A}+\mathbf{B}) = cD(\mathbf{A}) + D(\mathbf{B})$.

(b) By definition, the matrix of $\boldsymbol{T}$ is the unique matrix $\mathbf{B}$ such that $\boldsymbol{T}(\mathbf{v}) = \mathbf{B}\mathbf{v}$ for all $\mathbf{v}$. Since $\mathbf{A}$ is such a matrix, the matrix of $\boldsymbol{T}$ is $\mathbf{A}$; i.e., $\boldsymbol{CD}(\mathbf{A}) = \mathbf{A}$.

(c) Again by definition of the matrix of $\boldsymbol{T}$, if $\boldsymbol{T}$ has matrix $\mathbf{A}$, then $\boldsymbol{T}(v) = \mathbf{A}\mathbf{v}$ for all v, and so the linear map given by multiplication by $\mathbf{A}$ is exactly $\boldsymbol{T}$: $\boldsymbol{DC}(\boldsymbol{T}) = \boldsymbol{T}$.

2.2.19. Let $u_1, u_2 \in U$ and $c \in \mathbb{F}$. Then

$$\boldsymbol{S}\left[\boldsymbol{T}(cu_1+u_2)\right] = \boldsymbol{S}(c\boldsymbol{T}u_1 + \boldsymbol{T}u_2) = c\boldsymbol{ST}u_1 + \boldsymbol{ST}u_2,$$

and so $\boldsymbol{ST}$ is linear.

2.2.20. Let $v_1, v_2 \in V$ and $a \in \mathbb{F}$. Then

$$\begin{aligned}(\boldsymbol{S}+\boldsymbol{T})(av_1+v_2) &= \boldsymbol{S}(av_1+v_2) + \boldsymbol{T}(av_1+v_2) \\ &= a\boldsymbol{S}v_1 + \boldsymbol{S}v_2 + a\boldsymbol{T}v_1 + \boldsymbol{T}v_2 = a(\boldsymbol{S}+\boldsymbol{T})v_1 + (\boldsymbol{S}+\boldsymbol{T})v_2,\end{aligned}$$

so $\boldsymbol{S}+\boldsymbol{T}$ is linear.

Similarly,

$$\begin{aligned}(c\boldsymbol{T})(av_1+v_2) &= c\left[\boldsymbol{T}(av_1+v_2)\right] \\ &= c\left[a\boldsymbol{T}v_1 + \boldsymbol{T}v_2\right] = a(c\boldsymbol{T})v_1 + (c\boldsymbol{T})v_2,\end{aligned}$$

so $c\boldsymbol{T}$ is linear.

The operations of pointwise addition and scalar multiplication are trivially seen to have the required commutativity and associativity properties; these follow from the corresponding properties in W. The zero map acts as an additive identity in $\mathcal{L}(V,W)$, and for $\boldsymbol{T} \in \mathcal{L}(V,W)$, $(-1)\boldsymbol{T} \in \mathcal{L}(V,W)$ is an additive inverse. We thus have that $\mathcal{L}(V,W)$ is a vector space with pointwise addition and scalar multiplication.

2.3 Matrix multiplication

2.3.1. (a) $\begin{bmatrix} 8 & 5 \\ 20 & 13 \end{bmatrix}$ (b) $\begin{bmatrix} -3+6i & 2-2i \\ -2-11i & 8-5i \end{bmatrix}$ (c) not possible (d) -12 (e) not possible

2.3.2. (a) $\begin{bmatrix} -3+6i & 1+3i & 0 \\ 4 & 8 & -10i \\ 11 & 7-3i & -5i \end{bmatrix}$ (b) $\begin{bmatrix} 14 & 0 & -6 \\ 39 & 15 & -24 \end{bmatrix}$ (c) not defined (d) $\begin{bmatrix} -10 & 4 & -8 \\ 15 & -6 & 12 \\ 5 & -2 & 4 \end{bmatrix}$
(e) $\begin{bmatrix} 1 & 22 & -35 \\ -18 & 4 & -10 \end{bmatrix}$

2.3.3. (a) Computing column-by-column,

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

so

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

(b) Row-by-row,

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

so again

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

(c) Entry-by-entry.

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

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

and again,

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

2.3.4. The matrix of the rotation counterclockwise by θ radians is $\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$.

The matrix of the reflection across the line $y = x$ is $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$.

Therefore, the matrix of $\boldsymbol{T}$ is

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} = \begin{bmatrix} \sin\theta & \cos\theta \\ \cos\theta & -\sin\theta \end{bmatrix}.$$

2.3.5. The matrix of the reflection across the y-axis is $\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$.

The matrix of the map which stretches by a factor of 2 in the y-direction is $\begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$.

The matrix of the rotation counterclockwise by $\frac{\pi}{4}$ radians is $\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$.

Therefore, the matrix of $\boldsymbol{T}$ is

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

2.3.6. For $1 \le i \le m$ and $1 \le j \le p$,

$$[\mathbf{A}\mathbf{Z}_{n\times p}]_{ij} = \sum_{k=1}^{n} a_{ik} 0 = 0$$

and similarly, if $1 \le i \le p$ and $1 \le j \le n$, then

$$[\mathbf{Z}_{p\times m}\mathbf{A}]_{ij} = \sum_{k=1}^{n} 0 a_{kj} = 0.$$

For $1 \le i \le m$ and $1 \le j \le n$,

$$[\mathbf{A}\mathbf{I}_n]_{ij} = \sum_{k=1}^{n} a_{ik}\delta_{kj} = a_{ij}$$

and similarly, if $1 \le i \le m$ and $1 \le j \le n$, then

$$[\mathbf{I}_m\mathbf{A}]_{ij} = \sum_{k=1}^{n} \delta_{ik} a_{kj} = a_{ij}.$$

2.3.7. The i-j entry of $\mathbf{AB}$ is

$$\sum_{k=1}^{n} a_{ik} b_{kj} = \begin{cases} a_{ii} b_{ii}, & i = j; \\ 0, & i \ne j, \end{cases},$$

and so $\mathbf{AB} = \mathbf{diag}(a_1 b_1, \ldots, a_n b_n)$.

2.3.8. Let $\mathbf{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}$. Then

$$\mathbf{v}^T\mathbf{v} = \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix} \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} = v_1^2 + \cdots + v_n^2 = \|\mathbf{v}\|^2.$$

2.3.9. The matrix $[a]$ is invertible if and only if $a \neq 0$. If $a \neq 0$, $[a]^{-1} = [a^{-1}]$.

2.3.10. This is another application of the rat poison principle: observe that

$$(\mathbf{A}^{-1})^T\mathbf{A}^T = \left[\mathbf{A}\mathbf{A}^{-1}\right]^T = \mathbf{I}^T = \mathbf{I}$$

and

$$\mathbf{A}^T(\mathbf{A}^{-1})^T = \left[\mathbf{A}^{-1}\mathbf{A}\right]^T = \mathbf{I}^T = \mathbf{I}.$$

Since $(\mathbf{A}^{-1})^T$ functions as an inverse to $\mathbf{A}^T$,

$$(\mathbf{A}^{-1})^T = (\mathbf{A}^T)^{-1}.$$

2.3.11. Write $\mathbf{b}$ as $\mathbf{A}\mathbf{B}\mathbf{b}$; this says $\mathbf{x} = \mathbf{B}\mathbf{b}$ is a solution to $\mathbf{A}\mathbf{x} = \mathbf{b}$.

Suppose that $\mathbf{A}\mathbf{B} = \mathbf{I}_m$. Then for any $\mathbf{b} \in \mathbb{F}^m$,

$$\mathbf{A}(\mathbf{B}\mathbf{b}) = (\mathbf{A}\mathbf{B})\mathbf{b} = \mathbf{I}_m\mathbf{b} = \mathbf{b},$$

so $\mathbf{x} = \mathbf{B}\mathbf{b}$ is a solution to the linear system $\mathbf{A}\mathbf{x} = \mathbf{b}$. Hence the system $\mathbf{A}\mathbf{x} = \mathbf{b}$ is consistent for every $\mathbf{b} \in \mathbb{F}^m$.

By Theorem 1.2, this means that the RREF of the augmented matrix $[\mathbf{A}|\mathbf{b}]$ never has a pivot in the last column, which means that the RREF of $\mathbf{A}$ alone must have a pivot in each of its m rows (otherwise, we could reverse-engineer a $\mathbf{b}$ for which the RREF of $[\mathbf{A}|\mathbf{b}]$ did have a pivot in the last column, by reversing the Gaussian elimination from $[RREF(\mathbf{A})|\mathbf{e}_m]$). Since there is at most one pivot per column, this implies that $m \leq n$.

2.3.12. (a) The i, j entry of $\mathbf{A}\mathbf{B}$ is

$$\sum_{k=1}^{n} a_{ik}b_{kj} = \sum_{k=i}^{n} a_{ik}b_{kj}$$

since $a_{ik} = 0$ for $i > k$. Now if $i > j$, then for each term in the last sum, $k \geq i > j$, and so $b_{kj} = 0$, and so $\left[\mathbf{A}\mathbf{B}\right]_{ij} = 0$. Thus $\mathbf{A}\mathbf{B}$ is upper triangular.

Now for each i,

$$\left[\mathbf{A}\mathbf{B}\right]_{ii} = \sum_{k=1}^{n} a_{ik}b_{ki} = a_{ii}b_{ii},$$

since $a_{ik} = 0$ whenever $k < i$ and $b_{ki} = 0$ whenever $k > i$.

(b) This is essentially the same as the previous part: $\mathbf{AB}$ is lower triangular, with diagonal entries given by the products of the diagonal entries of $\mathbf{A}$ and $\mathbf{B}$.

2.3.13. Let $\mathbf{A}, \mathbf{A_1}, \mathbf{A_2} \in M_{m,n}(\mathbb{F})$, $\mathbf{B}, \mathbf{B}_1, \mathbf{B}_2 \in M_{n,p}(\mathbb{F})$, and $\mathbf{C} \in M_{p,q}(\mathbb{F})$.

1. The i-j entry of $\mathbf{A}(\mathbf{BC})$ is

$$\sum_{k=1}^{n} a_{ik}(\mathbf{BC})_{kj} = \sum_{k=1}^{n} a_{ik}\left(\sum_{\ell=1}^{p} b_{k\ell}c_{\ell j}\right) = \sum_{\ell=1}^{p}\left(\sum_{k=1}^{n} a_{ik}b_{k\ell}\right)c_{\ell j} = \sum_{\ell=1}^{p}(\mathbf{AB})_{i\ell}c_{\ell j},$$

which is the i-j entry of $(\mathbf{AB})\mathbf{C}$.

2. Let b^1_{ij} denote the entries of $\mathbf{B}_1$ and b^2_{ij} the entries of $\mathbf{B}_2$. The i-j entry of $\mathbf{A}(\mathbf{B}_1 + \mathbf{B}_2)$ is

$$\sum_{k=1}^{n} a_{ik}(b^1_{kj} + b^2_{kj}) = \sum_{k=1}^{n} a_{ik}b^1_{kj} + \sum_{k=1}^{n} a_{ik}b^2_{kj},$$

which is the i-j entry of $\mathbf{AB}_1 + \mathbf{AB}_2$.

3. Let a^1_{ij} denote the entries of $\mathbf{A}_1$ and a^2_{ij} the entries of $\mathbf{A}_2$. The i-j entry of $(\mathbf{A}_1 + \mathbf{A}_2)\mathbf{B}$ is

$$\sum_{k=1}^{n}(a^1_{ik} + a^2_{ik})b_{kj} = \sum_{k=1}^{n} a^1_{ik}b_{kj} + \sum_{k=1}^{n} a^2_{ik}b_{kj},$$

which is the i-j entry of $\mathbf{A}_1\mathbf{B} + \mathbf{A}_2\mathbf{B}$.

2.3.14. The i-jth entry of $\begin{bmatrix} —\mathbf{a}_1\mathbf{B}— \\ \vdots \\ —\mathbf{a}_m\mathbf{B}— \end{bmatrix}$ is

$$[\mathbf{a}_i\mathbf{B}]_j = \sum_{k=1}^{n} a_{ik}b_{kj} = (\mathbf{AB})_{ij}.$$

2.3.15. Recall that $(\mathbf{A}^{-1})^{-1}$ is defined to be any matrix $\mathbf{B}$ such that $\mathbf{BA}^{-1} = \mathbf{A}^{-1}\mathbf{B} = \mathbf{I}$. The matrix $\mathbf{B} = \mathbf{A}$ does the trick.

2.4 Row operations and the LU decomposition

2.4.1. (a) $\mathbf{Q}_{2,1}\mathbf{Q}_{3,2}\mathbf{P}_{-\frac{1}{2},1,2}\mathbf{I}_2$ (b) $\mathbf{P}_{3,2,1}\mathbf{P}_{-1,1,2}\mathbf{Q}_{-2,2}\mathbf{I}_2$ (c) $\mathbf{P}_{-2,2,1}\begin{bmatrix} 1 & -2 & -1 \\ 0 & 0 & 0 \end{bmatrix}$

(d) $\mathbf{P}_{2,2,1}\mathbf{P}_{-1,3,1}\mathbf{P}_{1,3,2}\begin{bmatrix} 1 & -2 & 0 & 2 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

2.4.2. (a) $\mathbf{Q}_{3,2}\begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$

(b) $\mathbf{P}_{2,2,1}\begin{bmatrix}1 & 1\\ 0 & 0\end{bmatrix}$

(c) $\mathbf{Q}_{2,1}\mathbf{P}_{1,2,1}\mathbf{Q}_{-1,2}\begin{bmatrix}1 & 0 & \frac{1}{2}\\ 0 & 1 & -\frac{5}{2}\end{bmatrix}$

(d) $\mathbf{Q}_{2,1}\mathbf{P}_{1,2,1}\mathbf{P}_{1,3,1}\mathbf{Q}_{\frac{1}{2},2}\mathbf{Q}_{-\frac{3}{2},3}\mathbf{P}_{1,3,2}\mathbf{Q}_{\frac{4}{3},3}\begin{bmatrix}1 & \frac{1}{2} & \frac{3}{2} & 0\\ 0 & 1 & -1 & -2\\ 0 & 0 & 0 & 1\end{bmatrix}$

2.4.3. (a) $\left[\begin{array}{cc|cc}1 & 2 & 1 & 0\\ 3 & 4 & 0 & 1\end{array}\right] \rightsquigarrow \left[\begin{array}{cc|cc}1 & 2 & 1 & 0\\ 0 & -2 & -3 & 1\end{array}\right] \rightsquigarrow \left[\begin{array}{cc|cc}1 & 0 & -2 & 1\\ 0 & 1 & \frac{3}{2} & -\frac{1}{2}\end{array}\right]$,

so $\begin{bmatrix}1 & 2\\ 3 & 4\end{bmatrix}^{-1} = \begin{bmatrix}-2 & 1\\ \frac{3}{2} & -\frac{1}{2}\end{bmatrix}$.

(b) $\begin{bmatrix}0+0i & 2/5-1/5i\\ -1/3+0i & 1/15+2/15i\end{bmatrix}$ (c) $\begin{bmatrix}1 & 1 & -1\\ 1 & -1 & 1\\ -1 & 1 & 1\end{bmatrix}$ (d) singular

(e) singular (f) $\begin{bmatrix}0 & 1 & 1 & 0\\ -1 & 2 & 1 & -2\\ 1 & -1 & 0 & 1\\ -2 & 3 & 2 & -2\end{bmatrix}$

2.4.4. (a) $\left[\begin{array}{cc|cc}1 & -2 & 1 & 0\\ -2 & 4 & 0 & 1\end{array}\right] \rightsquigarrow \left[\begin{array}{cc|cc}1 & -2 & 1 & 0\\ 0 & 0 & -2 & 1\end{array}\right]$, so the matrix is singular.

(b) singular (c) $\begin{bmatrix}2 & -1 & 0\\ -1 & 2 & -1\\ 0 & -1 & 1\end{bmatrix}$ (d) singular (e) $\begin{bmatrix}-\frac{1}{4} & \frac{1}{2} & \frac{1}{4}\\ \frac{7}{4} & -\frac{1}{2} & -\frac{3}{4}\\ -\frac{5}{4} & \frac{1}{2} & \frac{1}{4}\end{bmatrix}$ (f) $\begin{bmatrix}2 & 1 & -\frac{1}{2} & \frac{1}{2}\\ 1 & 0 & -\frac{1}{2} & \frac{1}{2}\\ -4 & -2 & 2 & -1\\ -7 & -3 & \frac{7}{2} & -\frac{3}{2}\end{bmatrix}$

2.4.5. Using Algorithm 2.25,

$$\begin{aligned}
\left[\begin{array}{ccc|ccc}1 & 0 & 3 & 1 & 0 & 0\\ 0 & 1 & 1 & 0 & 1 & 0\\ -1 & 0 & 2 & 0 & 0 & 1\end{array}\right] &\rightsquigarrow \left[\begin{array}{ccc|ccc}1 & 0 & 3 & 1 & 0 & 0\\ 0 & 1 & 1 & 0 & 1 & 0\\ 0 & 0 & 5 & 1 & 0 & 1\end{array}\right]\\
&\rightsquigarrow \left[\begin{array}{ccc|ccc}1 & 0 & 3 & 1 & 0 & 0\\ 0 & 1 & 1 & 0 & 1 & 0\\ 0 & 0 & 1 & 1/5 & 0 & 1/5\end{array}\right]\\
&\rightsquigarrow \left[\begin{array}{ccc|ccc}1 & 0 & 0 & 2/5 & 0 & -3/5\\ 0 & 1 & 1 & 0 & 1 & 0\\ 0 & 0 & 1 & 1/5 & 0 & 1/5\end{array}\right]\\
&\rightsquigarrow \left[\begin{array}{ccc|ccc}1 & 0 & 0 & 2/5 & 0 & -3/5\\ 0 & 1 & 0 & -1/5 & 1 & -1/5\\ 0 & 0 & 1 & 1/5 & 0 & 1/5\end{array}\right],
\end{aligned}$$

giving us the final answer

$$\begin{bmatrix}1 & 0 & 3\\ 0 & 1 & 1\\ -1 & 0 & 2\end{bmatrix}^{-1} = \frac{1}{5}\begin{bmatrix}2 & 0 & -3\\ -1 & 5 & -1\\ 1 & 0 & 1\end{bmatrix}.$$

(a) $(x, y, z) = \frac{1}{5}(-7, 6, 4)$
(b) $(x, y, z) = (-\frac{15}{2}, -\frac{35}{6}, \frac{5}{2})$
(c) $(x, y, z) = \frac{1}{5}(-1, 8, -3)$
(d) $(x, y, z) = \frac{1}{5}(2, -1, 1)$

2.4.6. Using Algorithm 2.25,

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 1 & 1 & 0 & 0 \\ 3 & 2 & 0 & 0 & 1 & 0 \\ 0 & 1 & -2 & 0 & 0 & 1 \end{array}\right] \rightsquigarrow \left[\begin{array}{ccc|ccc} 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 2 & -3 & -3 & 1 & 0 \\ 0 & 1 & -2 & 0 & 0 & 1 \end{array}\right]$$
$$\rightsquigarrow \left[\begin{array}{ccc|ccc} 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & -2 & 0 & 0 & 1 \\ 0 & 0 & 1 & -3 & 1 & -2 \end{array}\right]$$
$$\rightsquigarrow \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 4 & -1 & 2 \\ 0 & 1 & 0 & -6 & 2 & -3 \\ 0 & 0 & 1 & -3 & 1 & -2 \end{array}\right],$$

giving us the final answer

$$\begin{bmatrix} 1 & 0 & 1 \\ 3 & 2 & 0 \\ 0 & 1 & -2 \end{bmatrix}^{-1} = \begin{bmatrix} 4 & -1 & 2 \\ -6 & 2 & -3 \\ -3 & 1 & -2 \end{bmatrix}.$$

(a) $(x, y, z) = (8, -11, -7)$
(b) $(x, y, z) = \left(\frac{17}{3}, -\frac{53}{6}, -\frac{17}{3}\right)$
(c) $(x, y, z) = (-11, 17, 9)$
(d) $(x, y, z) = (4, -6, -3)$

2.4.7. (a) $\begin{bmatrix} 2 & 3 \\ -2 & 1 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 2 & 3 \\ 0 & 4 \end{bmatrix}$, so by Alorithm 2.28, $\begin{bmatrix} 2 & 3 \\ -2 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 3 \\ 0 & 4 \end{bmatrix}$.

(b) $\begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 2 & -1 \\ 0 & -1 \end{bmatrix}$ (c) $\begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -2 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 & -1 \\ 0 & -1 & 2 \\ 0 & 0 & 1 \end{bmatrix}$

(d) $\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} -1 & 1 & 2 \\ 0 & 2 & -1 \end{bmatrix}$

2.4.8. (a) $\begin{bmatrix} -1 & 3 \\ 2 & -4 \end{bmatrix} \rightsquigarrow \begin{bmatrix} -1 & 3 \\ 0 & 2 \end{bmatrix}$, so by Algorithm 2.28, $\begin{bmatrix} -1 & 3 \\ 2 & -4 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} -1 & 3 \\ 0 & 2 \end{bmatrix}$.

(b) $\begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ -1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 & 2 \\ 0 & 2 & -1 \\ 0 & 0 & -1 \end{bmatrix}$

(c) $\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ -2 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 & -3 \\ 0 & -2 & 3 \\ 0 & 0 & -3 \end{bmatrix}$

(d) $\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ -1 & 1 & 1 \end{bmatrix} \begin{bmatrix} -1 & 1 \\ 0 & -2 \\ 0 & 0 \end{bmatrix}$

2.4.9. (a) Using the LU factorization of the coefficient matrix found in Exercise 2.4.7(a), we solve

$$\begin{bmatrix} 2 & 3 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 3 \\ 0 & 4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

First let $\begin{bmatrix} z \\ w \end{bmatrix} = \begin{bmatrix} 2 & 3 \\ 0 & 4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$ and solve $\begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} z \\ w \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ to get $z = 1$ and $w = 3$. Then solve $\begin{bmatrix} 2 & 3 \\ 0 & 4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$ to get $x = -\frac{5}{8}$ and $y = \frac{3}{4}$

(b) Using the LU factorization of the coefficient matrix found in Exercise 2.4.7(b), we solve

$$\begin{bmatrix} 2 & -1 \\ 4 & -3 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 2 & -1 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ -2 \end{bmatrix}.$$

First let $\begin{bmatrix} z \\ w \end{bmatrix} = \begin{bmatrix} 2 & -1 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$ and solve $\begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} z \\ w \end{bmatrix} = \begin{bmatrix} 0 \\ -2 \end{bmatrix}$ to get $z = 0$ and $w = -2$. Then solve $\begin{bmatrix} 2 & -1 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ -2 \end{bmatrix}$ to get $(x, y) = (1, 2)$.

(c) Using the same method, $(x, y, z) = (0, 3, 2)$.

(d) Using the same method, $(x, y, z) = \frac{1}{5}(0, -9, 17) + c(5, 1, 2)$ for any $c \in \mathbb{R}$.

2.4.10. (a) Using the LU factorization of the coefficient matrix found in Exercise 2.4.8(a), we solve

$$\begin{bmatrix} -1 & 3 \\ 2 & -4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} -1 & 3 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -5 \\ 6 \end{bmatrix}.$$

First let $\begin{bmatrix} z \\ w \end{bmatrix} = \begin{bmatrix} -1 & 3 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$ and solve $\begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} z \\ w \end{bmatrix} = \begin{bmatrix} -5 \\ 6 \end{bmatrix}$ to get $z = -5$ and $w = -4$. Then solve $\begin{bmatrix} -1 & 3 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -5 \\ -4 \end{bmatrix}$ to get $x = -1$ and $y = -2$.

(b) Starting from the LU factorization from Exercise 2.4.8(b):

$$\begin{bmatrix} 1 & -1 & 2 \\ 3 & -1 & 5 \\ -1 & 3 & -4 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ -1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 & 2 \\ 0 & 2 & -1 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 3 \\ 10 \\ -3 \end{bmatrix}$$

First solve

$$\begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ -1 & 1 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 3 \\ 10 \\ -3 \end{bmatrix}$$

to get $a = 3$, $b = 1$, $c = -1$. Then solve

$$\begin{bmatrix} 1 & -1 & 2 \\ 0 & 2 & -1 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \\ -1 \end{bmatrix}$$

to get $x = 2$, $y = 1$, $z = 1$.

(c) Using the same method, $x = 1$, $y = 7$, $z = \frac{8}{3}$.

(d) Using the same method, $x = -2$, $y = -1$.

2.4.11. (a) Just switching the rows produces an upper triangular matrix, and so if $\mathbf{P} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, then

$$\mathbf{PA} = \begin{bmatrix} 2 & 3 \\ 0 & 1 \end{bmatrix} = \mathbf{LU},$$

where $\mathbf{L} = \mathbf{I}_2$ and $\mathbf{U} = \begin{bmatrix} 2 & 3 \\ 0 & 1 \end{bmatrix}$.

(b) Observe that

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \mathbf{A} = \begin{bmatrix} 2 & -1 & 2 \\ 0 & 2 & -1 \\ -2 & 1 & -1 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 2 & -1 & 2 \\ 0 & 2 & -1 \\ 0 & 0 & 1 \end{bmatrix},$$

so we can take

$$\mathbf{L} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix}, \mathbf{U} = \begin{bmatrix} 2 & -1 & 2 \\ 0 & 2 & -1 \\ 0 & 0 & 1 \end{bmatrix}, \text{and } \mathbf{P} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}.$$

2.4.12. (a) Just switching the rows produces an upper triangular matrix, and so if $\mathbf{P} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, then

$$\mathbf{PA} = \begin{bmatrix} -1 & 4 \\ 0 & 2 \end{bmatrix} = \mathbf{LU},$$

where $\mathbf{L} = \mathbf{I}_2$ and $\mathbf{U} = \begin{bmatrix} 2 & 3 \\ 0 & 1 \end{bmatrix}$.

(b) Observe that

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \mathbf{A} = \begin{bmatrix} 1 & 2 & -2 \\ 1 & 5 & -5 \\ 3 & 6 & -4 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 2 & -2 \\ 0 & 3 & -3 \\ 0 & 0 & 2 \end{bmatrix},$$

so we can take $\mathbf{P} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$, $\mathbf{U} = \begin{bmatrix} 1 & 2 & -2 \\ 0 & 3 & -3 \\ 0 & 0 & 2 \end{bmatrix}$, and $\mathbf{L} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 3 & 0 & 1 \end{bmatrix}$.

2.4.13. Using the decomposition from Exercise 2.4.11(b), we solve

$$\begin{bmatrix} 2 & -1 & 2 \\ -2 & 1 & -1 \\ 0 & 2 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & -1 & 2 \\ 0 & 2 & -1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 5 \\ -2 \\ 2 \end{bmatrix},$$

or equivalently,

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & -1 & 2 \\ 0 & 2 & -1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 5 \\ 2 \\ -2 \end{bmatrix}.$$

Solving

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} 5 \\ 2 \\ -2 \end{bmatrix}$$

gives $(x', y', z') = (5, 2, 3)$, and then solving

$$\begin{bmatrix} 2 & -1 & 2 \\ 0 & 2 & -1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 5 \\ 2 \\ 3 \end{bmatrix}$$

gives $(x, y, z) = \left(\frac{3}{4}, \frac{5}{2}, 3\right)$.

2.4.14. Using the decomposition from Exercise 2.4.12(b), we solve

$$\begin{bmatrix} 1 & 2 & -2 \\ 3 & 6 & -4 \\ 1 & 5 & -5 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 3 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & -2 \\ 0 & 3 & -3 \\ 0 & 0 & 2 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \\ 2 \end{bmatrix},$$

or equivalently,

$$\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 3 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & -2 \\ 0 & 3 & -3 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ 4 \end{bmatrix}.$$

Solving

$$\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 3 & 0 & 1 \end{bmatrix} \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ 4 \end{bmatrix}$$

gives $(x', y', z') = (2, 0, -2)$, and then solving

$$\begin{bmatrix} 1 & 2 & -2 \\ 0 & 3 & -3 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \\ -2 \end{bmatrix}$$

gives $(x, y, z) = (2, -1, -1)$.

2.4.15. These are easiest to see by considering the effect of multiplying an arbitrary matrix $\mathbf{A}$ by these products. $\mathbf{R}_{i,j}\mathbf{P}_{c,k,\ell}$ performs **R1** with rows k and ℓ, and then switches rows i and j, so if $\{i, j\} \cap \{k, \ell\} = \emptyset$, it doesn't matter which order you do these operations in: the matrices commute. If $i = k$ but $j \neq \ell$, then $\mathbf{R}_{i,j}\mathbf{P}_{c,k,\ell}$ adds c times row ℓ to row k, then switches row k with row j. This is the same as first switching rown k with row j, then adding c times row ℓ to row j: $\mathbf{R}_{k,j}\mathbf{P}_{c,k,\ell} = \mathbf{P}_{c,j,\ell}\mathbf{R}_{k,j}$. The other two are similar.

Altenatively, you can just write everything out in components and confirm the formulae.

2.4.16. (a) Since $\mathbf{L}$ is lower-triangular, $\ell_{1k} = 0$ for $k > 1$, and so

$$a_{11} = \sum_{k=1}^{m} \ell_{1k} u_{k1} = \ell_{11} u_{11}.$$

If $a_{11} = 0$, then we must have either $\ell_{11} = 0$ or $u_{11} = 0$.

(b) If $u_{11} = 0$, then since $\mathbf{U}$ is upper triangular, it must be that the entire first column is zero. A matrix is invertible if and only if its RREF is the identity; if it has a column of all zeroes, this is impossible. So $\mathbf{U}$ is singular.

If $\ell_{11} = 0$, we can use the fact that $\mathbf{L}$ is invertible if and only if $\mathbf{L}^T$ is invertible. But $\mathbf{L}^T$ is upper triangular, and $[\mathbf{L}^T]_{1,1} = \ell_{11}$. So $\mathbf{L}^T$ (and hence $\mathbf{L}$) is singular, by the same argument as for $\mathbf{U}$.

(c) Suppose that $u_{11} = 0$. Then $\mathbf{U}\mathbf{e}_1 = \mathbf{0}$, and so $\mathbf{A}\mathbf{e}_1 = \mathbf{L}\mathbf{U}\mathbf{e}_1 = \mathbf{0}$. Since $\mathbf{e}_1 \neq \mathbf{0}$, and $\mathbf{A}\mathbf{0} = \mathbf{0}$, this means that multiplication by $\mathbf{A}$ is not injective, and so $\mathbf{A}$ is singular. If it is $\ell_{11} = 0$, then apply the same argument to $\mathbf{A}^T$.

(d) $\mathbf{A} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ is invertible.

2.4.17. The $\mathbf{L}$ from the LU decomposition of $\mathbf{A}$ already works. The $\mathbf{U}$ is upper triangular but may not have 1's on the diagonal. Using row operation **R2** on $\mathbf{U}$ to get a new upper triangular matrix $\tilde{\mathbf{U}}$ with 1s on the diagonal:

$$\tilde{\mathbf{U}} = \mathbf{E}_1 \cdots \mathbf{E}_n \mathbf{U},$$

where if $u_{jj} \neq 0$, then $\mathbf{E}_j = \mathbf{Q}_{\frac{1}{u_{jj}},j}$, and otherwise let $\mathbf{E}_j = \mathbf{I}$. Taking $\mathbf{D} = \mathbf{diag}(u_{11}, \ldots, u_{nn})$ gives $\mathbf{A} = \mathbf{L}\mathbf{U} = \mathbf{L}\mathbf{D}\tilde{\mathbf{U}}$, which has the required form.

2.4.18. (a) From Exercise 2.4.8(a), $\mathbf{A} = \begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} -1 & 3 \\ 0 & 2 \end{bmatrix}$. Factoring out the diagonal entries as discussed in the solution of Exercise 2.4.17 then gives

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

(b) $\begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ -1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} 1 & -1 & 2 \\ 0 & 1 & -\frac{1}{2} \\ 0 & 0 & 1 \end{bmatrix}$

(c) $\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ -2 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -3 \end{bmatrix} \begin{bmatrix} 1 & \frac{1}{2} & -\frac{3}{2} \\ 0 & 1 & -\frac{3}{2} \\ 0 & 0 & 1 \end{bmatrix}$

(d) $\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ -1 & 1 & 1 \end{bmatrix} \begin{bmatrix} -1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$

2.4.19. If $\mathbf{A}$ is upper triangular, then the Gaussian elimination algorithm on $[\mathbf{A}|\mathbf{I}_n]$ will only involve multiplying rows by constants and adding a multiple of a row to a higher row. These operations applied to $\mathbf{I}_n$ will result in an upper triangular matrix, and thus $\mathbf{A}^{-1}$ is upper triangular.

2.4.20. (a) Subtract the second row from the first, then the third from the second, and so on:

$$
\left[\begin{array}{cccccc|cccccc}
1 & 1 & 1 & \cdots & \cdots & 1 & 1 & 0 & 0 & \cdots & \cdots & 0 \\
0 & 1 & 1 & \cdots & \cdots & 1 & 0 & 1 & 0 & \cdots & \cdots & 0 \\
0 & 0 & 1 & \ddots & & 1 & 0 & 0 & 1 & \ddots & & 0 \\
\vdots & & \ddots & \ddots & \ddots & \vdots & \vdots & & \ddots & \ddots & \ddots & \vdots \\
0 & & & \ddots & 1 & 1 & 0 & & & \ddots & 1 & 0 \\
0 & \cdots & \cdots & \cdots & 0 & 1 & 0 & \cdots & \cdots & \cdots & 0 & 1
\end{array}\right]
$$

$$
\rightsquigarrow \left[\begin{array}{cccccc|cccccc}
1 & 0 & 0 & \cdots & \cdots & 0 & 1 & -1 & 0 & \cdots & \cdots & 0 \\
0 & 1 & 1 & \cdots & \cdots & 1 & 0 & 1 & 0 & \cdots & \cdots & 0 \\
0 & 0 & 1 & \ddots & & 1 & 0 & 0 & 1 & \ddots & & 0 \\
\vdots & & \ddots & \ddots & \ddots & \vdots & \vdots & & \ddots & \ddots & \ddots & \vdots \\
0 & & & \ddots & 1 & 1 & 0 & & & \ddots & 1 & 0 \\
0 & \cdots & \cdots & \cdots & 0 & 1 & 0 & \cdots & \cdots & \cdots & 0 & 1
\end{array}\right]
$$

$$
\rightsquigarrow \left[\begin{array}{cccccc|cccccc}
1 & 0 & 0 & \cdots & \cdots & 0 & 1 & -1 & 0 & \cdots & \cdots & 0 \\
0 & 1 & 0 & \cdots & \cdots & 0 & 0 & 1 & -1 & \cdots & \cdots & 0 \\
0 & 0 & 1 & \ddots & & 1 & 0 & 0 & 1 & \ddots & & 0 \\
\vdots & & \ddots & \ddots & \ddots & \vdots & \vdots & & \ddots & \ddots & \ddots & \vdots \\
0 & & & \ddots & 1 & 1 & 0 & & & \ddots & 1 & 0 \\
0 & \cdots & \cdots & \cdots & 0 & 1 & 0 & \cdots & \cdots & \cdots & 0 & 1
\end{array}\right] \rightsquigarrow \textit{etc.}
$$

(b) In this case, one first adds the second row to the first, then the third row to the first two, and so on.

2.4.21. Write $\mathbf{Q}_{c,i} = [q_{jk}]_{j,k=1}^m$ and $\mathbf{A} = [a_{jk}]_{\substack{1\le j\le m \\ 1\le k\le n}}$. Then $[\mathbf{Q}_{c,i}\mathbf{A}]_{jk} = \sum_{\ell=1}^m q_{j\ell}a_{\ell k}$. Since $\mathbf{Q}_{c,i}$ is zero except on the diagonal, $q_{j\ell} = 1$ if $j = \ell$ but $j \neq i$, and $q_{ii} = c$, the only surviving term from the sum is $q_{jj}a_{jk}$, which is just a_{jk}, except when $j = i$, when it is ca_{ik}. This exactly means that the ith row of $\mathbf{A}$ has been multiplied by c.

2.4.22. The entries of $\mathbf{R}_{ij}$ are given by

$$
[\mathbf{R}_{ij}]_{k\ell} = \begin{cases} 1, & k = \ell, k \notin \{i,j\} \text{ or } i = k, j = \ell, \text{ or } i = \ell, j = k; \\ 0, & \text{otherwise.} \end{cases}
$$

Given $\mathbf{A} \in \mathrm{M}_{m,n}(\mathbb{F})$, this means that

$$
[\mathbf{R}_{ij}\mathbf{A}]_{k\ell} = \sum_{p=1}^m [\mathbf{R}_{ij}]_{kp} a_{p\ell} = \begin{cases} a_{k\ell}, & k \notin \{i,j\}; \\ a_{j\ell}, & k = i; \\ a_{i\ell}, & k = j. \end{cases}
$$

This is exactly saying that $\mathbf{R}_{ij}\mathbf{A}$ is the result of swapping the ith and jth rows of $\mathbf{A}$.

2.5 Range, kernel, and eigenspaces

2.5.1. (a) We are looking for the solutions to

$$\left[\begin{array}{ccc|c} 1 & 2 & 3 & 0 \\ 4 & 5 & 6 & 0 \end{array}\right] \rightsquigarrow \left[\begin{array}{ccc|c} 1 & 2 & 3 & 0 \\ 0 & -3 & -6 & 0 \end{array}\right] \rightsquigarrow \left[\begin{array}{ccc|c} 1 & 2 & 3 & 0 \\ 0 & 1 & 2 & 0 \end{array}\right] \rightsquigarrow \left[\begin{array}{ccc|c} 1 & 0 & -1 & 0 \\ 0 & 1 & 2 & 0 \end{array}\right];$$

solving for the pivot variables x and y in terms of the free variable z gives that $x = z$ and $y = -2z$, and so the kernel of our matrix is $\left\langle \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} \right\rangle$.

(b) $\left\langle \begin{bmatrix} -6 \\ 3 \\ 5 \end{bmatrix} \right\rangle$ (c) $\left\langle \begin{bmatrix} 2 \\ -3 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ -2 \\ 0 \\ 1 \end{bmatrix} \right\rangle$ (d) $\langle \mathbf{0} \rangle$ (e) $\left\langle \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\rangle$

2.5.2. (a) We are looking for the solutions to $\left[\begin{array}{cc|c} 1 & -2 & 0 \\ 4 & -8 & 0 \end{array}\right] \rightsquigarrow \left[\begin{array}{cc|c} 1 & -2 & 0 \\ 0 & 0 & 0 \end{array}\right]$; solving for the pivot variable x in terms of the free variable y gives $x = 2y$, and so the kernel of our matrix is $\left\{ \begin{bmatrix} 2y \\ y \end{bmatrix} : y \in \mathbb{R} \right\} = \left\langle \begin{bmatrix} 2 \\ 1 \end{bmatrix} \right\rangle$.

(b) $\left\langle \begin{bmatrix} 2 \\ 5 \\ 4 \end{bmatrix} \right\rangle$ (c) $\left\langle \begin{bmatrix} 11 \\ 23 \\ -14 \end{bmatrix} \right\rangle$ (d) $\left\langle \begin{bmatrix} 2 \\ -5 \\ 4 \\ 0 \end{bmatrix}, \begin{bmatrix} -6 \\ 1 \\ 0 \\ 4 \end{bmatrix} \right\rangle$ (e) $\left\langle \begin{bmatrix} -1 \\ -1 \\ 1 \\ 0 \end{bmatrix} \right\rangle$

2.5.3. A list of vectors in $\mathbb{F}^m$ spans $\mathbb{F}^m$ if and only if the RREF of the matrix with those columns has a pivot in every row.

(a) $REF\left(\begin{bmatrix} 1 & -3 \\ 2 & 4 \end{bmatrix}\right) = \begin{bmatrix} 1 & -3 \\ 0 & 1 \end{bmatrix}$, so the given vectors do span $\mathbb{R}^2$.

(b) $REF\left(\begin{bmatrix} -2 & 1 & 3 \\ 1 & 1 & -1 \end{bmatrix}\right) = \begin{bmatrix} 1 & -\frac{1}{2} & -\frac{3}{2} \\ 0 & 1 & \frac{1}{3} \end{bmatrix}$, so the given vectors do span $\mathbb{R}^2$.

(c) The REF of a 3×2 matrix cannot have a pivot in every row, so these vectors do not span $\mathbb{R}^3$.

(d) $REF\left(\begin{bmatrix} 1 & 3 & 2 \\ 0 & -1 & -1 \\ 2 & 1 & -1 \end{bmatrix}\right) = \begin{bmatrix} 1 & 3 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$, so the given vectors do not span $\mathbb{R}^3$.

(e) $REF\left(\begin{bmatrix} 1 & 3 & 0 \\ 0 & -1 & 1 \\ 2 & 1 & -1 \end{bmatrix}\right) = \begin{bmatrix} 1 & 3 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$, so the given vectors do span $\mathbb{R}^3$.

2.5.4. A list of vectors in $\mathbb{F}^m$ spans $\mathbb{F}^m$ if and only if the RREF of the matrix with those columns has a pivot in every row.

(a) $REF\left(\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}\right) = \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix}$, so the given vectors do not span $\mathbb{F}^2$.

(b) $REF\left(\begin{bmatrix}1 & -1 & 2\\ -1 & 2 & -1\end{bmatrix}\right) = \begin{bmatrix}1 & -1 & 2\\ 0 & 0 & 1\end{bmatrix}$, so the given vectors do span $\mathbb{R}^2$.

(c) $REF\left(\begin{bmatrix}1 & 3 & 2 & 0\\ 0 & -1 & -1 & 1\\ 2 & 1 & -1 & -1\end{bmatrix}\right) = \begin{bmatrix}1 & 3 & 2 & 0\\ 0 & 1 & 1 & -1\\ 0 & 0 & 0 & 1\end{bmatrix}$, so the given vectors do span $\mathbb{R}^3$.

(d) Over $\mathbb{R}$, $REF\left(\begin{bmatrix}1 & 1 & 0\\ 1 & 0 & 1\\ 0 & 1 & 1\end{bmatrix}\right) = \begin{bmatrix}1 & 1 & 0\\ 0 & 1 & -1\\ 0 & 0 & 1\end{bmatrix}$, and so the given vectors do span $\mathbb{R}^3$.

(e) Over $\mathbb{F}_2$, $REF\left(\begin{bmatrix}1 & 1 & 0\\ 1 & 0 & 1\\ 0 & 1 & 1\end{bmatrix}\right) = \begin{bmatrix}1 & 1 & 0\\ 0 & 1 & 1\\ 0 & 0 & 0\end{bmatrix}$, and so the given vectors do not span $\mathbb{F}_2^3$.

2.5.5. (a) Yes; the 3-eigenspace is the null space of $\begin{bmatrix}-1 & 2\\ 2 & -4\end{bmatrix} \rightsquigarrow \begin{bmatrix}1 & -2\\ 0 & 0\end{bmatrix}$, so $\mathrm{Eig}_3(\mathbf{A}) = \left\langle \begin{bmatrix}2\\ 1\end{bmatrix} \right\rangle$.

(b) No; $RREF\,(\mathbf{A}+\mathbf{I}) = \mathbf{I}$.

(c) Yes:

$$\mathbf{A}+2\mathbf{I} = \begin{bmatrix}2 & 1 & -1\\ 1 & 2 & -2\\ 0 & 1 & -1\end{bmatrix} \rightsquigarrow \begin{bmatrix}1 & 0 & 0\\ 0 & 1 & -1\\ 0 & 0 & 0\end{bmatrix}$$

and so $\mathrm{Eig}_{-2}(\mathbf{A}) = \left\langle \begin{bmatrix}0\\ 1\\ 1\end{bmatrix} \right\rangle$.

(d) Yes:

$$\mathbf{A}-\mathbf{I} = \begin{bmatrix}6 & -2 & 4\\ 0 & 0 & 0\\ -6 & 2 & -4\end{bmatrix} \rightsquigarrow \begin{bmatrix}3 & -1 & 2\\ 0 & 0 & 0\\ 0 & 0 & 0\end{bmatrix}$$

and so $\mathrm{Eig}_1(\mathbf{A}) = \left\langle \begin{bmatrix}1\\ 3\\ 0\end{bmatrix}, \begin{bmatrix}-2\\ 0\\ 3\end{bmatrix} \right\rangle$.

(e) Yes:

$$\mathbf{A}+3\mathbf{I} = \begin{bmatrix}1 & 0 & 1 & 0\\ 0 & 1 & 0 & 1\\ 1 & 0 & 1 & 0\\ 0 & 1 & 0 & 1\end{bmatrix} \rightsquigarrow \begin{bmatrix}1 & 0 & 1 & 0\\ 0 & 1 & 0 & 1\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\end{bmatrix}$$

and so $\mathrm{Eig}_{-3}(\mathbf{A}) = \left\langle \begin{bmatrix}1\\ 0\\ -1\\ 0\end{bmatrix}, \begin{bmatrix}0\\ 1\\ 0\\ -1\end{bmatrix} \right\rangle$.

2.5.6. (a) No; $RREF\,(\mathbf{A}-\mathbf{I})=\mathbf{I}$.

(b) Yes; the 0-eigenspace is just the null space of

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

so $\mathrm{Eig}_0(\mathbf{A})=\left\langle\begin{bmatrix}2\\3\\1\end{bmatrix}\right\rangle$.

(c) No:

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

(d) Yes:

$$\mathbf{A}-3\mathbf{I}=\begin{bmatrix}4 & -2 & 4\\ 0 & -2 & 0\\ -6 & 2 & -6\end{bmatrix}\rightsquigarrow\begin{bmatrix}1 & 0 & 1\\ 0 & 1 & 0\\ 0 & 0 & 0\end{bmatrix},$$

and so $\mathrm{Eig}_3(\mathbf{A})=\left\langle\begin{bmatrix}-1\\0\\1\end{bmatrix}\right\rangle$.

(e) Yes:

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

and so

$$\mathrm{Eig}_{-1}(\mathbf{A})=\left\{\begin{bmatrix}-2z-2w\\ w\\ z\\ w\end{bmatrix}:z,w\in\mathbb{R}\right\}=\left\langle\begin{bmatrix}-2\\0\\1\\0\end{bmatrix},\begin{bmatrix}-2\\1\\0\\1\end{bmatrix}\right\rangle.$$

2.5.7. (a) Put the system into matrix form and perform Gaussian elimination:

$$\left[\begin{array}{ccc|c}1 & 3 & -1 & 2\\ 2 & 2 & 1 & -3\end{array}\right]\rightsquigarrow\left[\begin{array}{ccc|c}1 & 3 & -1 & 2\\ 0 & -4 & 3 & -7\end{array}\right]\rightsquigarrow\left[\begin{array}{ccc|c}1 & 3 & -1 & 2\\ 0 & 1 & -\frac{3}{4} & \frac{7}{4}\end{array}\right]\rightsquigarrow\left[\begin{array}{ccc|c}1 & 0 & \frac{5}{4} & -\frac{13}{4}\\ 0 & 1 & -\frac{3}{4} & -\frac{7}{4}\end{array}\right].$$

Solving for the pivot variables in terms of the free variable gives that the set of solutions is $\left\{\begin{bmatrix}-\frac{13}{4}\\ -\frac{7}{4}\\ 0\end{bmatrix}+c\begin{bmatrix}-5\\3\\4\end{bmatrix}:c\in\mathbb{R}\right\}$.

(b) The RREF of the augmented matrix of the system is $\left[\begin{array}{ccc|c} 1 & 0 & 3 & 1 \\ 0 & 1 & 2 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right]$. Solving for the pivot variables in terms of the free variable gives that the set of solutions is $\left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + c \begin{bmatrix} -3 \\ -2 \\ 1 \end{bmatrix} : c \in \mathbb{R} \right\}$.

(c) The RREF of the augmented matrix of the system is $\left[\begin{array}{cccc|c} 1 & 0 & 2 & 3 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{array}\right]$. The system is therefore inconsistent.

(d) The RREF of the augmented matrix of the system is $\left[\begin{array}{ccc|c} 1 & 0 & -2 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$. Solving for the pivot variables in terms of the free variable gives that the set of solutions is $\left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + c \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix} : c \in \mathbb{R} \right\}$

(e) The RREF of the augmented matrix of the system is $\left[\begin{array}{cccc|c} 1 & 0 & 0 & -1 & 1 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & -2 & -1 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$. Solving for the pivot variables in terms of the free variable gives that the set of solutions is $\left\{ \begin{bmatrix} 1 \\ 1 \\ -1 \\ 0 \end{bmatrix} + c \begin{bmatrix} 1 \\ -1 \\ 2 \\ 1 \end{bmatrix} : c \in \mathbb{R} \right\}$

2.5.8. (a) Put the system into matrix form and perform Gaussian elimination:

$$\left[\begin{array}{cc|c} 1 & -2 & -3 \\ -2 & 4 & 6 \end{array}\right] \rightsquigarrow \left[\begin{array}{cc|c} 1 & -2 & -3 \\ 0 & 0 & 0 \end{array}\right].$$

Solving for the pivot variables in terms of the free variable gives that the set of solutions is $\left\{ \begin{bmatrix} -3 \\ 0 \end{bmatrix} + c \begin{bmatrix} 2 \\ 1 \end{bmatrix} : c \in \mathbb{R} \right\}$.

(b) The RREF of the augmented matrix of the system is

$$\left[\begin{array}{ccc|c} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right].$$

Solving for the pivot variables in terms of the free variable gives that the set of solutions is $\left\{ \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + c \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix} \right\}$.

(c) The RREF of the augmented matrix of the system is

$$\left[\begin{array}{cccc|c} 1 & 0 & \frac{1}{3} & 0 & 0 \\ 0 & 1 & -\frac{5}{3} & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{array}\right],$$

so the system is inconsistent.
The RREF of the augmented matrix of the system is

$$\left[\begin{array}{cccc|c} 1 & 0 & 0 & \frac{5}{2} & \frac{3}{2} \\ 0 & 1 & 0 & -\frac{1}{2} & \frac{3}{2} \\ 0 & 0 & 1 & -5 & -5 \end{array}\right].$$

Solving for the pivot variables in terms of the free variable gives that the set of solutions is $\left\{ \begin{bmatrix} \frac{3}{2} \\ \frac{3}{2} \\ -5 \\ 0 \end{bmatrix} + c \begin{bmatrix} -5 \\ 1 \\ 10 \\ 2 \end{bmatrix} \right\}$.

(d) The RREF of the augmented matrix of the system is

$$\left[\begin{array}{ccc|c} 1 & 0 & \frac{1}{4} & -\frac{1}{4} \\ 0 & 1 & \frac{1}{4} & \frac{11}{4} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right].$$

Solving for the pivot variables in terms of the free variable gives that the set of solutions is $\left\{ \begin{bmatrix} -\frac{1}{4} \\ \frac{11}{4} \\ 0 \end{bmatrix} + c \begin{bmatrix} -1 \\ -1 \\ 4 \end{bmatrix} \right\}$.

2.5.9. (a) If $f_p(t) = 2e^{-t}$, then

$$\frac{d^2}{dt^2} f_p(t) + f_p(t) = \frac{d}{dt}(-2e^{-t}) + 2e^{-t} = 2e^{-t} + 2e^{-t} = 4e^{-t}.$$

(b) $\boldsymbol{T}f = f'' + f$ is a linear map on $C^2(\mathbb{R})$, so the set of solutions to the given differential equation is the set of $f \in C^2(\mathbb{R})$ with $\boldsymbol{T}f = g$, where $g(t) = 4e^{-t}$. The kernel of $\boldsymbol{T}$ is the set of solutions to the homogenous differential equation $\frac{d^2 f}{dt^2} + f(t) = 0$, which we are told is the set $\{k_1 \sin(t) + k_2 \cos(t) : k_1, k_2 \in \mathbb{R}\}$. Since f_p is a solution to $\boldsymbol{T}f = g$, Proposition 4.2 gives that the set of all solutions is $\{f_p(t) + k_1 \sin(t) + k_2 \cos(t) : k_1, k_2 \in \mathbb{R}\}$.

(c) If $f(t) = f_p(t) + k_1 \sin(t) + k_2 \cos(t)$, then $f(0) = f_p(0) + k_2 = 2 + k_2$, so to get $f(0) = 1$, we must have $k_2 = -1$; we thus have that $f(t)$ is a solution with $f(0) = 1$ if and only if $f(t) = 2e^{-t} + k_1 \sin(t) - \cos(t)$ for some $k_1 \in \mathbb{R}$.

(d) Such a solution must have the form $f(t) = 2e^{-t} + k_1 \sin(t) + k_2 \cos(t)$ with $k_1 = -2e^{-\frac{\pi}{2}} + b$ and $k_2 = -2 + a$.

2.5.10. Since U is a subspace, $0 \in U$, and so $0 = \boldsymbol{T}(0) \in \boldsymbol{T}(U)$. If $w_1, w_2 \in \boldsymbol{T}(U)$, then there are $u_1, u_2 \in U$ such that $w_1 = \boldsymbol{T}(u_1)$ and $w_2 = \boldsymbol{T}(u_2)$. Since U is a subspace, if $c \in \mathbb{F}$, $cu_1 + u_2 \in U$, and so

$$cw_1 + w_2 = c\boldsymbol{T}u_1 + \boldsymbol{T}u_2 = \boldsymbol{T}(cu_1 + u_2) \in \boldsymbol{T}(U).$$

2.5.11. (a) Let $w \in \operatorname{range}(\boldsymbol{S} + \boldsymbol{T})$. Then there exists $v \in V$ such that

$$w = (\boldsymbol{S} + \boldsymbol{T})(v) = \boldsymbol{S}v + \boldsymbol{T}v.$$

Since $\boldsymbol{S}v \in \operatorname{range}(\boldsymbol{S})$ and $\boldsymbol{T}v \in \operatorname{range}(\boldsymbol{T})$, $\boldsymbol{S}v + \boldsymbol{T}v \in \operatorname{range}(\boldsymbol{S}) + \operatorname{range}(\boldsymbol{T})$.

(b) $C(\mathbf{A}) = \operatorname{range}(\mathbf{A})$ as an operator on $\mathbb{F}^n$, so this is immediate from the previous part.

2.5.12. A function $f \in C^\infty(\mathbb{R})$ is in the kernel of $\boldsymbol{D}$ if and only if $f'(x) = 0$ for all x, which holds if and only if f is constant.

2.5.13. The matrix $\mathbf{AB} = \mathbf{0}$ if and only if $\mathbf{A}(\mathbf{Bv}) = \mathbf{0}$ for every $\mathbf{v}$, if and only if $\mathbf{Bv} \in \ker \mathbf{A}$ for every $\mathbf{v}$. Since $C(\mathbf{B}) = \{\mathbf{Bv} : \mathbf{v} \in \mathbb{F}^p\}$, this shows that $\mathbf{AB} = 0$ if and only if $C(\mathbf{B}) \subseteq \ker \mathbf{A}$.

2.5.14. Every linear map $\boldsymbol{T}$ has $\boldsymbol{T}(0) = 0$, so if $\boldsymbol{T}$ is injective, there can be no $v \neq 0$ with $\boldsymbol{T}(v) = 0$; i.e., 0 is not an eigenvalue of $\boldsymbol{T}$. Conversely, if $\boldsymbol{T}$ is not injective, then by Theorem 2.37, $\ker \boldsymbol{T} \neq \{0\}$; i.e., there is a nonzero v with $\boldsymbol{T}v = 0$. This means v is an eigenvector of $\boldsymbol{T}$ with eigenvalue 0.

2.5.15. Let v be an eigenvector with eigenvalue λ. Then

$$\boldsymbol{T}^k(v) = \boldsymbol{T}^{k-1}(\boldsymbol{T}v) = \boldsymbol{T}^{k-1}(\lambda v) = \lambda \boldsymbol{T}^{k-1}(v) = \lambda \boldsymbol{T}^{-k-2}(\boldsymbol{T}v) = \cdots = \lambda^k v.$$

This shows both that λ^k is an eigenvalue of $\boldsymbol{T}^k$ and that if $v \in \operatorname{Eig}_\lambda(\boldsymbol{T})$, then $v \in \operatorname{Eig}_{\lambda^k}\big(\boldsymbol{T}^k\big)$.

2.5.16. Consider the map $\boldsymbol{T} : \mathbb{R}^2 \to \mathbb{R}^2$ given by the matrix $\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$; this map has eigenvalues ± 1. Since $\boldsymbol{T}^2 = \boldsymbol{I}$, $\operatorname{Eig}_{-1}(\boldsymbol{T}) \neq \operatorname{Eig}_1\big(\boldsymbol{T}^2\big)$.

2.5.17. No. For example, the set of eigenvectors together with 0 of $\begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$ is the x- and y-axes (and nothing else), which is not a subspace of $\mathbb{R}^2$.

2.5.18. Since $\boldsymbol{T}$ is linear, $\boldsymbol{T}(0) = 0 \in U$, so $0 \in X$. Suppose that $x_1, x_2 \in X$ and $c \in \mathbb{F}$. Then $\boldsymbol{T}x_1, \boldsymbol{T}x_2 \in U$, and so

$$\boldsymbol{T}(cx_1 + x_2) = c\boldsymbol{T}x_1 + \boldsymbol{T}x_2 \in U,$$

and thus $cx_1 + x_2 \in X$.

2.5.19. By definition, $C(\mathbf{AB})$ is the set of linear combinations of the columns of $\mathbf{AB}$. Since each column of $\mathbf{AB}$ is itself a linear combination of columns of $\mathbf{A}$, every element of $C(\mathbf{AB})$ is a linear combination of columns of $\mathbf{A}$, and so $C(\mathbf{AB}) \subseteq C(\mathbf{A})$.

2.5.20. Suppose that $(\mathbf{v}_1, \dots, \mathbf{v}_n)$ spans $\mathbb{F}^m$. By Corollary 2.34, this means that the $m \times n$ matrix whose j^{th} column is $\mathbf{v}_j$ has a pivot in every row, hence m pivots. Since there is at most one pivot in a given column, this means that $m \leq n$.

2.5.21. Since $\boldsymbol{T}$ is linear, $\boldsymbol{T}(0) = 0$, so $0 \in \operatorname{Eig}_\lambda(\boldsymbol{T})$. If $v_1, v_2 \in \operatorname{Eig}_\lambda(\boldsymbol{T})$ and $c \in \mathbb{F}$, then

$$\boldsymbol{T}(cv_1 + v_2) = c\boldsymbol{T}v_1 + \boldsymbol{T}v_2 = c\lambda v_1 + \lambda v_2 = \lambda(cv_1 + v_2),$$

so $cv_1 + v_2 \in \operatorname{Eig}_\lambda(\boldsymbol{T})$.

2.5.22. Let $\boldsymbol{T} \in \mathcal{L}(V, W)$ and suppose that $\boldsymbol{T}(v_0) = w$. If $y \in \ker \boldsymbol{T}$, then

$$\boldsymbol{T}(v_0 + y) = \boldsymbol{T}(v_0) + \boldsymbol{T}(y) = w + 0 = w.$$

Conversely, if $v \in V$ is such that $\boldsymbol{T}v = w$, then write $v = v_0 + (v - v_0)$ and observe that

$$\boldsymbol{T}(v - v_0) = \boldsymbol{T}v - \boldsymbol{T}v_0 = w - w = 0,$$

so $v - v_0 \in \ker \boldsymbol{T}$.

2.6 Error-correcting linear codes

2.6.1. (a) The parity bit encoding is $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}$

(b) The triple repitition code repeats each bit three times, so the encoded version is

$(1, 1, 1, 0, 0, 0, 1, 1, 1, 1, 1, 1)$.

(c) The Hamming code encodes the vector as $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}$.

2.6.2. (a) An error occurred, since the $z_1 + z_2 = 1 \neq z_3$. The original vector $\mathbf{x}$ could have been $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$, $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$, (if the error was in the first or second of the original bits) or $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ (if the error was in the parity bit).

(b) No error occurred: $z_1 + z_2 = z_3$.

(c) No error occurred: $z_1 + z_2 + z_3 = z_4$.

(d) An error occurred. The original vector could have been any of $\begin{bmatrix}0\\0\\1\end{bmatrix}$, $\begin{bmatrix}1\\1\\1\end{bmatrix}$, $\begin{bmatrix}1\\0\\0\end{bmatrix}$, or $\begin{bmatrix}1\\0\\1\end{bmatrix}$.

2.6.3. (a) No error occurred: $z_1 + z_2 = z_3$.

(b) An error occurred, since the $z_1 + z_2 = 1 \neq z_3$. The original vector could have been $\begin{bmatrix}1\\0\end{bmatrix}$ or $\begin{bmatrix}0\\1\end{bmatrix}$ (if the error was in the first or second of the original bits) or $\begin{bmatrix}1\\1\end{bmatrix}$ (if the error was in the parity bit).

(c) No error occurred: $z_1 + z_2 + z_3 = z_4$.

(d) An error occurred. The original vector could have been any of $\begin{bmatrix}1\\1\\0\end{bmatrix}$, $\begin{bmatrix}1\\0\\1\end{bmatrix}$, $\begin{bmatrix}0\\1\\1\end{bmatrix}$ or $\begin{bmatrix}1\\1\\1\end{bmatrix}$.

2.6.4. (a) Letting $\mathbf{z}$ denote the received vector, $\mathbf{y}$ the transmitted vector, and $\mathbf{B}$ the parity check matrix for the Hamming code, computing $\mathbf{Bz}$ gives

$$\mathbf{Bz} = \begin{bmatrix}1&1&1&0&1&0&0\\1&1&0&1&0&1&0\\1&0&1&1&0&0&1\end{bmatrix}\begin{bmatrix}0\\1\\1\\0\\0\\0\\1\end{bmatrix} = \begin{bmatrix}0\\1\\0\end{bmatrix} = \mathbf{Be}_6 = \mathbf{B}(\mathbf{y}+\mathbf{e}_6),$$

which means that an error occurred in the transmission of the sixth bit. We thus have that the transmitted vector $\mathbf{y}$ was $(0,1,1,0,0,1,1)$, which in turn means that the original message $\mathbf{x}$ was $(0,1,1,0)$ (just the first four entries of $\mathbf{y}$).

(b) Using the same method, $\mathbf{Bz} = (1,1,1) = \mathbf{b}_1$, so an error occurred in the first bit, and the original message was $\mathbf{x} = (1,1,0,0)$.

(c) Using the same method, $\mathbf{Bz} = (0,0,0)$, so in this case, no errors occurred and the original message was $\mathbf{x} = (1,0,0,1)$.

(d) Using the same method, $\mathbf{Bz} = (0,0,0)$, so in this case, no errors occurred and the original message was $\mathbf{x} = (1,1,1,1)$.

(e) Using the same method, $\mathbf{Bz} = (1,1,1) = \mathbf{b}_1$, so an error occurred in the first bit, and the original message was $\mathbf{x} = (0,0,1,0)$.

2.6.5. (a) Letting $\mathbf{z}$ denote the received vector, $\mathbf{y}$ the transmitted vector, and $\mathbf{B}$ the parity check matrix for the Hamming code, computing $\mathbf{Bz}$ gives

$$\mathbf{Bz} = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \\ 0 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \mathbf{Be}_6 = \mathbf{B}(\mathbf{y} + \mathbf{e}_6),$$

which means that an error occurred in the transmission of the sixth bit. We thus have that the transmitted vector $\mathbf{y}$ was $(1, 0, 1, 1, 0, 0, 1)$, which in turn means that the original message $\mathbf{x}$ was $(1, 0, 1, 1)$ (just the first four entries of $\mathbf{y}$).

(b) Using the same method, $\mathbf{Bz} = (1, 1, 1) = \mathbf{b}_1$, so an error occurred in the first bit, and the original message was $\mathbf{x} = (0, 0, 1, 1)$.

(c) Using the same method, $\mathbf{Bz} = (0, 1, 1) = \mathbf{b}_4$, so an error occurred in the fourth bit, and the original message was $\mathbf{x} = (1, 1, 0, 0)$.

(d) Using the same method, $\mathbf{Bz} = (0, 1, 0) = \mathbf{b}_6$, so an error occurred in the sixth bit, and the original message was $\mathbf{x} = (0, 1, 0, 1)$.

(e) Using the same method, $\mathbf{Bz} = (1, 0, 1) = \mathbf{b}_3$, so an error occurred in the third bit, and the original message was $\mathbf{x} = (1, 0, 1, 0)$.

2.6.6. It was shown in the section that $C(\mathbf{A}) \subseteq \ker \mathbf{B}$. From the forms of $\mathbf{A}$ and $\mathbf{B}$, one can see immediately that $\operatorname{rank} \mathbf{A} = 4$ and $\operatorname{rank} B = 3$. By the rank-nullity theorem, this means that $\dim(\ker \mathbf{B}) = 4 = \dim(C(\mathbf{A}))$, and so they are equal.

2.6.7. Suppose that errors occur in the bits $i_1, \dots, i_m$ for $m \geq 2$. That means

$$\mathbf{z} = \mathbf{y} + \mathbf{e}_{i_1} + \cdots + \mathbf{e}_{i_m},$$

and so

$$\mathbf{Bz} = \mathbf{B}(\mathbf{y} + \mathbf{e}_{i_1} + \cdots + \mathbf{e}_{i_m}) = \mathbf{B}(\mathbf{e}_{i_1} + \cdots + \mathbf{e}_{i_m}),$$

since $\mathbf{By} = \mathbf{BAx} = 0$. Remember that the columns of $\mathbf{B}$ consist of all the nonzero vectors in $\mathbb{F}_2^3$. Thus a linear combination of the columns of $\mathbf{B}$ is either 0 or else one of the columns of $\mathbf{B}$.

If there are errors in only $m = 2$ bits, then since $\mathbf{Be}_{i_1} \neq \mathbf{Be}_{i_2}$ and we're working over $\mathbb{F}_2$,

$$\mathbf{Bz} = \mathbf{B}(\mathbf{e}_{i_1} + \mathbf{e}_{i_2}) = \mathbf{Be}_{i_1} + \mathbf{Be}_{i_2} \neq \mathbf{0}.$$

Thus we can tell that an error has occured, but we will not be able to correct it since the output of the parity check will be the same as if there were a single error in whichever column of $\mathbf{B}$ is the same as the sum of the i_1 and i_2 columns. For example (using the version of the parity check matrix discussed in the section), if errors occur in both the first and second bits, the output of the parity check will be the same as if an error occured in the seventh bit.

If three or more errors occur, then $\mathbf{Bz}$ may, as above, be the same as $\mathbf{Be}_i$ for some i, and so the parity check would give a misleading indication of the location of the error. More seriously, it may occur that $\mathbf{B}(\mathbf{e}_{i_1} + \cdots + \mathbf{e}_{i_m}) = \mathbf{0}$, for example if errors occur in the first, second, and seventh bits. In that case the output of the parity check will be the same as if no error had occured at all.

This shows that the Hamming code is not very useful if there is a significant probability of having two or more errors when transmitting a 7-bit message.

2.6.8. If the original message is $\mathbf{x} = \mathbf{0}$, then $\mathbf{y} = \mathbf{A}\mathbf{x} = \mathbf{0}$, and so if an error occurs in the ith bit only then $\mathbf{z} = \mathbf{e}_i$. So we would need

$$\mathbf{C}\mathbf{e}_i = \mathbf{x} = \mathbf{0}$$

for each i; in other words, $\mathbf{C}$ must be the 4×7 zero matrix. But in that case we would also have $\mathbf{C}\mathbf{z} = \mathbf{0}$ no matter what the original message $\mathbf{x}$ was, so $\mathbf{C}$ would fail to recover any nonzero $\mathbf{x}$.

2.6.9. There are 2^4 vectors in $\mathbb{F}_2^4$. Each can lead to $8 = 2^3$ received messages (no errors or an error in one of the seven transmitted bits). Since we have shown that the original message can be recovered from the received message when at most one error occurs, these vectors must all be distinct, so there are 2^7 possible received messages, which is all of $\mathbb{F}_2^7$.

2.6.10. As in the case of the Hamming code, start by making an encoding matrix which contains all of the nonzero vectors in $\mathbb{F}_2^4$, and whose final 4×4 block is $\mathbf{I}_4$: $\mathbf{B} = \begin{bmatrix} \mathbf{M} & \mathbf{I}_4 \end{bmatrix}$, with

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

Then take $\mathbf{A} = \begin{bmatrix} \mathbf{I}_{11} \\ \mathbf{M} \end{bmatrix}$ as the encoding matrix, so that as in the case of the Hamming code, $\mathbf{BA} = \mathbf{M} + \mathbf{M} = \mathbf{0}$; it follows that $C(\mathbf{A}) \subseteq \ker \mathbf{B}$. Suppose that there is at most one error in the recieved message $\mathbf{z}$. Let $\mathbf{y}$ denote the transmitted (encoded) message, so that $\mathbf{y} = \mathbf{A}\mathbf{x}$ for some (unencoded) message $\mathbf{x}$. If there is an error in transmission there will be a value of j such that

$$\mathbf{B}\mathbf{z} = \mathbf{B}(\mathbf{A}\mathbf{x} + \mathbf{e}_j) = \mathbf{B}\mathbf{e}_j = \mathbf{b}_j.$$

The receiver can thus determine which bit of $\mathbf{z}$ should be changed to recover $\mathbf{y}$, and then read off the first eleven entries of $\mathbf{y}$ to recover $\mathbf{x}$.

2.6.11. (a) In order to get the encoded vector $\begin{bmatrix} \mathbf{A}\mathbf{x} \\ x_1 + x_2 + x_3 + x_4 \end{bmatrix}$, one would take the encoding matrix to be $\tilde{\mathbf{A}} = \begin{bmatrix} \mathbf{A} \\ 1\ 1\ 1\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$.

(b) The parity check matrix $\tilde{\mathbf{B}} \in \mathrm{M}_{4,8}(\mathbb{F}_2)$ should have the property that $\tilde{\mathbf{B}}\tilde{\mathbf{A}} = \mathbf{0}$. Our encoding matrix $\tilde{\mathbf{A}}$ has the form $\begin{bmatrix} \mathbf{I}_4 \\ \tilde{\mathbf{M}} \end{bmatrix}$, so if we take

$\tilde{\mathbf{B}} = \begin{bmatrix} \tilde{\mathbf{M}} & \mathbf{I}_4 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}$, then we get $\tilde{\mathbf{B}}\tilde{\mathbf{A}} = \tilde{\mathbf{M}} + \tilde{\mathbf{M}} = \mathbf{0}$.

Chapter 3

Linear independence, bases, and coordinates

3.1 Linear (in)dependence

3.1.1. (a) The RREF of $\begin{bmatrix} 1 & 2 & 1 \\ 2 & 3 & 1 \\ 3 & 1 & 2 \end{bmatrix}$ is $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$. According to Algorithm 3.4, the list is linearly independent.

(b) The RREF of $\begin{bmatrix} 1 & 2 & 1 \\ 2 & 3 & 1 \\ 3 & 1 & -2 \end{bmatrix}$ is $\begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$. According to Algorithm 3.4, the list is linearly dependent.

(c) The RREF of $\begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$ (over $\mathbb{R}$) is $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$. According to Algorithm 3.4, the list is linearly independent.

(d) The RREF of $\begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$ (over $\mathbb{F}_2$) is $\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$. According to Algorithm 3.4, the list is linearly dependent.

(e) The RREF of $\begin{bmatrix} 1 & 2 & 0 \\ -2 & 1 & 5 \\ -1 & 3 & 5 \\ 0 & -2 & -2 \end{bmatrix}$ is $\begin{bmatrix} 1 & 0 & -2 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$. According to Algorithm 3.4, the list is linearly dependent.

3.1.2. (a) $\begin{bmatrix} 1-i \\ 2 \end{bmatrix} = (1-i) \begin{bmatrix} 1 \\ 1+i \end{bmatrix}$, so these vectors are linearly dependent.

(b) By Corollary 3.3, any list of 4 vectors in $\mathbb{R}^3$ is linearly dependent.

(c) The RREF of $\begin{bmatrix} 1 & 2 & 1 \\ 1 & 1 & 0 \\ 2 & 3 & 1 \end{bmatrix}$ is $\begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$. According to Algorithm 3.4, the list is linearly dependent.

(d) The RREF of $\begin{bmatrix} 2 & 1 & 1 & 6 \\ -1 & 2 & -8 & 0 \\ 1 & -1 & 5 & -1 \\ 1 & 3 & -7 & 2 \end{bmatrix}$ is $\begin{bmatrix} 1 & 0 & 2 & 0 \\ 0 & 1 & -3 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$. According to Algorithm 3.4, the list is linearly dependent.

(e) The RREF of $\begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 \end{bmatrix}$ is $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$. According to Algorithm 3.4, the list is linearly independent.

3.1.3. We wanted to see the last equation as a linear combination of the first three, so we wanted a, b, c so that $2x + \frac{1}{4}y + z = a\left(x + \frac{1}{4}y\right) + b\left(2x + \frac{1}{4}y\right) + c(x + 8z)$. By equating coefficients of x, y and z we get $a + 2b + c = 2$, $\frac{1}{4}a + \frac{1}{4}b = \frac{1}{4}$, and $8c = 1$.

3.1.4. This follows directly from either version of the definition of linear dependence.

3.1.5. Suppose $v_1, \dots, v_n$ are all the vectors other than 0 in the given list. Then

$$0 = \sum_{i=1}^{n} 0v_i,$$

so one vector in the list, namely 0, is a linear combination of the other vectors in the list.

3.1.6. Suppose that $m \le n$, $v_1, \dots, v_n \in V$, and $v_1, \dots, v_m$ are linearly dependent. Then there exist scalars $a_1, \dots, a_m \in \mathbb{F}$, not all equal to 0, such that $\sum_{i=1}^{m} a_i v_i = 0$. Define $a_i = 0$ for $m + 1 \le i \le n$. Then it is still true that not all the a_i are 0, and $\sum_{i=1}^{n} a_i v_i = 0$. Therefore $v_1, \dots, v_n$ are also linearly dependent.

3.1.7. (a) This follows directly from the first definition of linear dependence.

(b) $\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

3.1.8. Since $\begin{bmatrix} i \\ -1 \end{bmatrix} = i \begin{bmatrix} 1 \\ i \end{bmatrix}$, the list is linearly dependent over $\mathbb{C}$. But since there is no $a \in \mathbb{R}$ such that $\begin{bmatrix} i \\ -1 \end{bmatrix} = a \begin{bmatrix} 1 \\ i \end{bmatrix}$, (the first equation would say that $a = i \notin \mathbb{R}$) the list is linearly independent over $\mathbb{R}$.

3.1.9. Suppose that $0 = c_1 \boldsymbol{T} v_1 + \dots + c_n \boldsymbol{T} v_n$ for some scalars $c_1, \dots, c_n$. By linearity, $\boldsymbol{T}\big(c_1 v_1 + \dots + c_n v_n\big) = 0$. Since $\boldsymbol{T}$ is injective, this implies that $c_1 v_1 + \dots + c_n v_n = 0$. Since the v_i are linearly independent, it follows that the c_i are all 0. Therefore $(\boldsymbol{T} v_1, \dots, \boldsymbol{T} v_n)$ is linearly independent.

3.1.10. (a) By Proposition 3.1, $\ker \mathbf{A} = \{\mathbf{0}\}$ iff the columns of $\mathbf{A}$ are linearly independent. By Algorithm 3.4, this is the case iff the RREF of $\mathbf{A}$ has a pivot in each column. Since $\mathbf{A}$ is $n \times n$, this is true iff the RREF has a pivot in each row. By Corollary 2.34, this is true iff $C(\mathbf{A}) = \mathbb{F}^n$.

(b) Let $\mathbf{A}$ be the matrix of $\boldsymbol{T}$. Then $\boldsymbol{T}$ is injective iff $\ker \mathbf{A} = \{\mathbf{0}\}$, and $\boldsymbol{T}$ is surjective iff $C(\mathbf{A}) = \mathbb{F}^n$. By part (a) these are equivalent.

3.1.11. When putting $\mathbf{A}$ into RREF, each diagonal entry will become a pivot, so there will be a pivot in each column of the RREF. According to Algorithm 3.4, this means that the columns of $\mathbf{A}$ are linearly independent.

3.1.12. For each $k = 2, \dots, n$, any linear combination of the first $k-1$ columns of $\mathbf{A}$ has k^{th} entry equal to 0, so the k^{th} column is not in the span of the earlier columns. By Corollary 3.7, the columns of $\mathbf{A}$ are linearly independent.

3.1.13. The matrix $[\mathbf{v}_1 \cdots \mathbf{v}_n] \in \mathrm{M}_{m,n}(\mathbb{F})$ has the same RREF when considered as a matrix in $\mathrm{M}_{m,n}(\mathbb{K})$. By Algorithm 3.4, this RREF determines whether $\mathbf{v}_1, \dots, \mathbf{v}_n$ are linearly independent over either field.

3.1.14. For any $k \in \{1, \dots, n\}$, the function $f_k(x) = \sin(kx)$ is an eigenvector (we usually say eigenfunction) of $\boldsymbol{D}^2$, with eigenvalue $-k^2$. Since $\sin(x), \dots, \sin(nx)$ are therefore all eigenvectors of the same linear map, with distinct eigenvalues, they are linearly independent by Theorem 3.8, and so if $\sum_{k=1}^n a_k \sin(kx)$ is the zero function, then $a_k = 0$ for each k.

3.1.15. For any $k \in \{1, \dots, n\}$, the function $f_k(x) = e^{c_k x}$ is an eigenvector (we usually say eigenfunction) of the derivative operator $\boldsymbol{D}$, with eigenvalue c_k. Since $e^{c_1 x}, \dots, e^{c_n x}$ are therefore all eigenvectors of the same linear map, with distinct eigenvalues, they are linearly independent by Theorem 3.8, and so if $\sum_{k=1}^n a_k e^{c_k x}$ is the zero function, then $a_k = 0$ for each k.

3.1.16. If $a_1 v_1 + \cdots + a_n v_n = 0$ for some $a_1, \dots, a_n \in \mathbb{Z}$ not all equal to 0, then $(v_1, \dots, v_n)$ is linearly dependent.

If $(v_1, \dots, v_n)$ are linearly dependent over $\mathbb{Q}$, then there exist $c_1, \dots, c_n \in \mathbb{Q}$, not all equal to 0, such that $c_1 v_1 + \cdots + c_n v_n = 0$. Let b be a common denominator for the nonzero c_i, and define $a_i = b c_i$. Then $a_i \in \mathbb{Z}$ for each i, not all the a_i are equal to 0, and $a_1 v_1 + \cdots + a_n v_n = 0$.

3.1.17. Let V be any nonzero real vector space, and let $\boldsymbol{T}$ be any linear map with an eigenvector v. (For example, $V = \mathbb{R}^2$ and $\boldsymbol{T}$ is the identity map.) Then v and $2v$ are linearly independent, and if $\boldsymbol{T}v = \lambda v$, then $\boldsymbol{T}(2v) = \lambda(2v)$, so $2v$ is also an eigenvector.

3.1.18. Suppose that $a_1 \log p_1 + \cdots + a_n \log p_n = 0$ for $a_1, \dots, a_n \in \mathbb{Q}$. Let b be a common denominator for the a_i, so that $c_i = b a_i \in \mathbb{Z}$. By properties of logarithms, this implies that $p_1^{c_1} \cdots p_n^{c_n} = 1$. By reordering if necessary, we may assume that $c_1, \dots, c_m \geq 0$ and $c_{m+1}, \dots, c_n < 0$, and then write $d_i = -c_i > 0$ for $i \geq m+1$. Then $p_1^{c_1} \cdots p_m^{c_m} = p_{m+1}^{d_{m+1}} \cdots p_n^{d_n}$. By the uniqueness of prime factorizations for integers, this implies that $c_i = 0$ for every i, and therefore $a_i = 0$ for every i. Thus $(\log p_1, \dots, \log p_n)$ is linearly independent over $\mathbb{Q}$.

3.1.19. Let $\mathbf{A}$ be the matrix of $\boldsymbol{T}$. By Theorem 2.37, $\boldsymbol{T}$ is injective iff $\ker \mathbf{A} = \ker \boldsymbol{T} = \{0\}$. By Proposition 3.1 this is the case iff the columns of $\mathbf{A}$ are linearly independent.

3.1.20. Let $(\mathbf{v}_1, \dots, \mathbf{v}_n)$ be a list of vectors in $\mathbb{F}^m$ with $n > m$. The matrix $\mathbf{A} = [\mathbf{v}_1 \cdots \mathbf{v_n}]$ has more columns than rows, so its REF cannot have a pivot in each column. By Algorithm 3.4, it follows that $(\mathbf{v}_1, \dots, \mathbf{v}_n)$ must be linearly dependent.

3.2 Bases

3.2.1. (a) The RREF of $\begin{bmatrix} 1 & 2 \\ -3 & 6 \end{bmatrix}$ is $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$. By Corollary 3.16, the list is therefore a basis for $\mathbb{R}^2$.

(b) No; by Quick Exercise #7, a basis of $\mathbb{R}^2$ must consist of 2 vectors.

(c) The RREF of $\begin{bmatrix} 1 & 1 & 2 \\ 0 & 2 & 1 \\ 2 & 3 & -1 \end{bmatrix}$ is $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$. By Corollary 3.16, the list is therefore a basis for $\mathbb{R}^3$.

(d) The RREF of $\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 2 & 2 \\ 1 & 2 & 3 & 3 \\ 1 & 2 & 3 & 4 \end{bmatrix}$ is $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$. By Corollary 3.16, the list is therefore a basis for $\mathbb{R}^4$.

3.2.2. (a) This list is linearly dependent, since $\begin{bmatrix} -2 \\ 6 \end{bmatrix} = -2 \begin{bmatrix} 1 \\ 3 \end{bmatrix}$, and therefore cannot be a basis.

(b) No; by Quick Exercise #7, a basis of $\mathbb{R}^3$ must consist of 3 vectors.

(c) The RREF of $\begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \\ 3 & 1 & 2 \end{bmatrix}$ is $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$. By Corollary 3.16, the list is therefore a basis for $\mathbb{R}^3$.

(d) The RREF of $\begin{bmatrix} 1 & 2 & 1 \\ 2 & 3 & 1 \\ 3 & 1 & -2 \end{bmatrix}$ is $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$. By Corollary 3.16, the list is therefore a basis for $\mathbb{R}^3$.

3.2.3. (a) The RREF of $\begin{bmatrix} 1 & 0 & 2 & 1 \\ 0 & -3 & 3 & -3 \\ 2 & 1 & 3 & 3 \end{bmatrix}$ is $\begin{bmatrix} 1 & 0 & 2 & 1 \\ 0 & 1 & -1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$. By Algorithm 3.13

$$\left(\begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 0 \\ -3 \\ 1 \end{bmatrix} \right)$$

is a basis for the space.

(b) The RREF of $\begin{bmatrix} 4 & -2 & 0 & 4 \\ -2 & 1 & 3 & 1 \\ 2 & -1 & -2 & 0 \end{bmatrix}$ is $\begin{bmatrix} 1 & -\frac{1}{2} & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$. By Algorithm 3.13

$$\left(\begin{bmatrix} 4 \\ -2 \\ 2 \end{bmatrix}, \begin{bmatrix} 0 \\ 3 \\ -2 \end{bmatrix} \right)$$

is a basis for the space.

(c) The RREF of $\begin{bmatrix} 2 & 1 & 0 & 2 & -1 \\ 0 & 1 & -2 & 1 & 0 \\ -1 & 0 & -1 & -2 & 2 \\ 3 & 1 & 1 & 0 & 1 \end{bmatrix}$ is $\begin{bmatrix} 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & -2 & 0 & 1 \\ 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$. By Algorithm 3.13

$$\left(\begin{bmatrix} 2 \\ 0 \\ -1 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ -2 \\ 0 \end{bmatrix} \right)$$

is a basis for the space.

(d) The RREF of $\begin{bmatrix} 1 & 4 & 1 & 0 & 1 \\ -2 & -3 & 0 & 1 & 1 \\ 3 & 2 & 1 & 0 & 1 \\ -4 & -1 & 0 & 1 & 1 \end{bmatrix}$ is $\begin{bmatrix} 1 & 0 & 0 & -\frac{1}{5} & -\frac{1}{5} \\ 0 & 1 & 0 & -\frac{1}{5} & -\frac{1}{5} \\ 0 & 0 & 1 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$. By Algorithm 3.13

$$\left(\begin{bmatrix} 1 \\ -2 \\ 3 \\ -4 \end{bmatrix}, \begin{bmatrix} 4 \\ -3 \\ 2 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix} \right)$$

is a basis for the space.

3.2.4. (a) The RREF of $\begin{bmatrix} 1 & 2 & 1 \\ 2 & 3 & 1 \\ 3 & 1 & -2 \end{bmatrix}$ is $\begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$. By Algorithm 3.13

$$\left(\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix} \right)$$

is a basis for the space.

(b) The RREF of $\begin{bmatrix} 1 & 2 & 1 & 0 & 1 \\ -2 & 0 & 2 & 3 & 1 \\ 1 & 3 & 2 & -1 & -3 \\ 0 & -1 & -1 & 2 & 2 \end{bmatrix}$ is $\begin{bmatrix} 1 & 0 & -1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$. By Algorithm 3.13

$$\left(\begin{bmatrix} 1 \\ -2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 3 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 3 \\ -1 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ -3 \\ 2 \end{bmatrix} \right)$$

is a basis for the space.

(c) The RREF of $\begin{bmatrix} 2 & 1 & 1 \\ -1 & 0 & -1 \\ 0 & 2 & -2 \\ -3 & -1 & -2 \end{bmatrix}$ is $\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$. By Algorithm 3.13

$$\left(\begin{bmatrix} 2 \\ -1 \\ 0 \\ -3 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 2 \\ -1 \end{bmatrix} \right)$$

is a basis for the space.

(d) The RREF of $\begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}$ is $\begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$. By Algorithm 3.13

$$\left(\begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix} \right)$$

is a basis for the space.

3.2.5. (a) The RREF of $\mathbf{A}$ is $\begin{bmatrix} 1 & 0 & -1 & -2 \\ 0 & 1 & 2 & 3 \end{bmatrix}$, so solutions of the system $\mathbf{Ax} = \mathbf{0}$ can be written $x = z + 2w$, $y = -2z - 3w$ for $z, w \in \mathbb{R}$, or

$$\begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix} = z \begin{bmatrix} 1 \\ -2 \\ 1 \\ 0 \end{bmatrix} + w \begin{bmatrix} 2 \\ -3 \\ 0 \\ 1 \end{bmatrix}$$

in vector form. Thus $\left(\begin{bmatrix} 1 \\ -2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ -3 \\ 0 \\ 1 \end{bmatrix} \right)$ is a basis for $\ker \mathbf{A}$.

(b) The RREF of $\mathbf{A}$ is $\begin{bmatrix} 1 & 0 & 2 & -3 \\ 0 & 1 & -5 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$, so solutions of the system $\mathbf{Ax} = \mathbf{0}$ can be written $x = -2z + 3w$, $y = 5z - 3w$ for $z, w \in \mathbb{R}$, or

$$\begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix} = z \begin{bmatrix} -2 \\ 5 \\ 1 \\ 0 \end{bmatrix} + w \begin{bmatrix} 3 \\ -3 \\ 0 \\ 1 \end{bmatrix}$$

in vector form. Thus $\left(\begin{bmatrix} -2 \\ 5 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 3 \\ -3 \\ 0 \\ 1 \end{bmatrix} \right)$ is a basis for $\ker \mathbf{A}$.

(c) The RREF of $\mathbf{A}$ is $\begin{bmatrix}1 & -2 & 0 & 0 & 1\\0 & 0 & 1 & 0 & 1\\0 & 0 & 0 & 1 & 2\\0 & 0 & 0 & 0 & 0\end{bmatrix}$, so solutions of the system $\mathbf{A}\mathbf{x} = \mathbf{0}$ can be written $x_1 = 2x_2 - x_5$, $x_3 = -x_5$, $x_4 = -2x_5$ for $x_2, x_5 \in \mathbb{R}$, or

$$\begin{bmatrix}x_1\\x_2\\x_3\\x_4\\x_5\end{bmatrix} = x_2\begin{bmatrix}2\\1\\0\\0\\0\end{bmatrix} + x_5\begin{bmatrix}-1\\0\\-1\\-2\\1\end{bmatrix}$$

in vector form. Thus $\left(\begin{bmatrix}2\\1\\0\\0\\0\end{bmatrix}, \begin{bmatrix}-1\\0\\-1\\-2\\1\end{bmatrix}\right)$ is a basis for $\ker \mathbf{A}$.

(d) The RREF of $\mathbf{A}$ is $\begin{bmatrix}1 & 0 & 0 & -\frac{1}{3} & \frac{1}{6} & \frac{7}{6}\\0 & 1 & 0 & -\frac{1}{3} & \frac{13}{6} & \frac{1}{6}\\0 & 0 & 1 & -\frac{1}{3} & -\frac{1}{3} & \frac{2}{3}\end{bmatrix}$, so solutions of the system $\mathbf{A}\mathbf{x} = \mathbf{0}$ can be written

$$\begin{bmatrix}x_1\\x_2\\x_3\\x_4\\x_5\\x_6\end{bmatrix} = \frac{x_4}{3}\begin{bmatrix}1\\1\\1\\3\\0\\0\end{bmatrix} + \frac{x_5}{6}\begin{bmatrix}-1\\-13\\2\\0\\6\\0\end{bmatrix} + \frac{x_6}{6}\begin{bmatrix}-7\\-1\\-4\\0\\0\\6\end{bmatrix}$$

for $x_4, x_5, x_6 \in \mathbb{R}$. Thus $\left(\begin{bmatrix}1\\1\\1\\3\\0\\0\end{bmatrix}, \begin{bmatrix}-1\\-13\\2\\0\\6\\0\end{bmatrix}, \begin{bmatrix}-7\\-1\\-4\\0\\0\\6\end{bmatrix}\right)$ is a basis for $\ker \mathbf{A}$.

3.2.6. (a) The RREF of $\mathbf{A}$ is $\begin{bmatrix}1 & 0 & \frac{17}{11} & -\frac{3}{22}\\0 & 1 & \frac{7}{11} & \frac{13}{22}\end{bmatrix}$, so solutions of the system $\mathbf{A}\mathbf{x} = \mathbf{0}$ can be written $x = -\frac{17}{11}z + \frac{3}{22}w$, $y = -\frac{7}{11}z - \frac{13}{22}w$ for $z, w \in \mathbb{R}$, or

$$\begin{bmatrix}x\\y\\z\\w\end{bmatrix} = \frac{z}{11}\begin{bmatrix}-17\\-7\\11\\0\end{bmatrix} + \frac{w}{22}\begin{bmatrix}3\\-13\\0\\22\end{bmatrix}$$

in vector form. Thus $\left(\begin{bmatrix}-17\\-7\\11\\0\end{bmatrix}, \begin{bmatrix}3\\-13\\0\\22\end{bmatrix}\right)$ is a basis for $\ker \mathbf{A}$.

(b) The RREF of $\mathbf{A}$ is $\begin{bmatrix} 1 & 0 & \frac{1}{2} & -\frac{1}{2} \\ 0 & 1 & \frac{1}{6} & -\frac{3}{2} \\ 0 & 0 & 0 & 0 \end{bmatrix}$, so solutions of the system $\mathbf{Ax} = \mathbf{0}$ can be written $x = -\frac{1}{2}z + \frac{1}{2}w$, $y = -\frac{1}{6}z + \frac{3}{2}w$ for $z, w \in \mathbb{R}$, or

$$\begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix} = \frac{z}{6}\begin{bmatrix} -3 \\ -1 \\ 6 \\ 0 \end{bmatrix} + \frac{w}{2}\begin{bmatrix} 1 \\ 3 \\ 0 \\ 2 \end{bmatrix}$$

in vector form. Thus $\left(\begin{bmatrix} -3 \\ -1 \\ 6 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \\ 0 \\ 2 \end{bmatrix}\right)$ is a basis for $\ker \mathbf{A}$.

(c) The RREF of $\mathbf{A}$ is $\begin{bmatrix} 1 & 0 & 0 & -\frac{18}{11} & \frac{7}{11} \\ 0 & 1 & 0 & -\frac{8}{11} & -\frac{3}{11} \\ 0 & 0 & 1 & \frac{24}{11} & -\frac{13}{11} \end{bmatrix}$, so solutions of the system $\mathbf{Ax} = \mathbf{0}$ can be written

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \frac{x_4}{11}\begin{bmatrix} 18 \\ 8 \\ -24 \\ 11 \\ 0 \end{bmatrix} + \frac{x_5}{11}\begin{bmatrix} -7 \\ 3 \\ 13 \\ 0 \\ 11 \end{bmatrix}$$

for $x_4, x_5 \in \mathbb{R}$. Thus $\left(\begin{bmatrix} 18 \\ 8 \\ -24 \\ 11 \\ 0 \end{bmatrix}, \begin{bmatrix} -7 \\ 3 \\ 13 \\ 0 \\ 11 \end{bmatrix}\right)$ is a basis for $\ker \mathbf{A}$.

(d) The RREF of $\mathbf{A}$ is $\begin{bmatrix} 1 & 0 & 0 & \frac{10}{3} & -\frac{4}{3} & \frac{17}{3} \\ 0 & 1 & 0 & \frac{2}{3} & -\frac{2}{3} & \frac{10}{3} \\ 0 & 0 & 1 & \frac{1}{3} & \frac{1}{6} & -\frac{4}{3} \end{bmatrix}$, so solutions of the system $\mathbf{Ax} = \mathbf{0}$ can be written

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = \frac{x_4}{3}\begin{bmatrix} -10 \\ -2 \\ -1 \\ 3 \\ 0 \\ 0 \end{bmatrix} + \frac{x_5}{6}\begin{bmatrix} 8 \\ 4 \\ -1 \\ 0 \\ 6 \\ 0 \end{bmatrix} + \frac{x_6}{3}\begin{bmatrix} -17 \\ -10 \\ 4 \\ 0 \\ 0 \\ 3 \end{bmatrix}$$

for $x_4, x_5, x_6 \in \mathbb{R}$. Thus $\left(\begin{bmatrix} -10 \\ -2 \\ -1 \\ 3 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 8 \\ 4 \\ -1 \\ 0 \\ 6 \\ 0 \end{bmatrix}, \begin{bmatrix} -17 \\ -10 \\ 4 \\ 0 \\ 0 \\ 3 \end{bmatrix}\right)$ is a basis for $\ker \mathbf{A}$.

3.2.7. (a) We wish to solve the linear system

$$x\begin{bmatrix}1\\1\end{bmatrix}+y\begin{bmatrix}1\\-1\end{bmatrix}=\begin{bmatrix}1\\2\end{bmatrix},$$

which has RREF $\left[\begin{array}{cc|c}1&0&\frac{3}{2}\\0&1&-\frac{1}{2}\end{array}\right]$. This yields the unique solution $x=\frac{3}{2}$, $y=-\frac{1}{2}$.

(b) By the same method, $\begin{bmatrix}1\\2\end{bmatrix}=\dfrac{3}{2}\begin{bmatrix}2\\3\end{bmatrix}-\dfrac{1}{2}\begin{bmatrix}4\\-5\end{bmatrix}$.

(c) By the same method, $\begin{bmatrix}1\\2\end{bmatrix}=\dfrac{3}{5}\begin{bmatrix}-1\\2\end{bmatrix}+\dfrac{4}{5}\begin{bmatrix}2\\1\end{bmatrix}$.

(d) By the same method, $\begin{bmatrix}1\\2\end{bmatrix}=-\dfrac{3}{4}\begin{bmatrix}2\\-1\end{bmatrix}+\dfrac{5}{8}\begin{bmatrix}4\\2\end{bmatrix}$.

3.2.8. (a) We wish to solve the linear system

$$x\begin{bmatrix}1\\1\\0\end{bmatrix}+y\begin{bmatrix}1\\0\\1\end{bmatrix}+z\begin{bmatrix}0\\1\\1\end{bmatrix}=\begin{bmatrix}1\\2\\3\end{bmatrix},$$

which has RREF $\left[\begin{array}{ccc|c}1&0&0&0\\0&1&0&1\\0&0&1&2\end{array}\right]$. This yields the unique solution $x=0$, $y=1$, $z=2$.

(b) By the same method, $\begin{bmatrix}1\\2\\3\end{bmatrix}=14\begin{bmatrix}1\\0\\0\end{bmatrix}-8\begin{bmatrix}2\\-1\\0\end{bmatrix}+3\begin{bmatrix}1\\-2\\1\end{bmatrix}$.

(c) In this case the corresponding linear system has RREF $\left[\begin{array}{ccc|c}1&0&1&0\\0&1&-3&0\\0&0&0&1\end{array}\right]$ and is thus inconsistent. Therefore $\begin{bmatrix}1\\2\\3\end{bmatrix}$ does not lie in the span of the given vectors.

3.2.9. Every element of $\mathbb{C}$ can be uniquely written as $a+ib$ with $a,b\in\mathbb{R}$.

3.2.10. **Solution 1:** Writing $\mathbf{A}=\begin{bmatrix}a&b\\c&d\end{bmatrix}$, the given information amounts to the linear system

$$\begin{aligned}a+2b&=0\\c+2d&=-1\\2a+b&=-3\\2c+d&=2\end{aligned}$$

which further breaks into two 2×2 systems, which can easily be solved to yield $a=-2$, $b=1$, $c=5/3$, $d=-4/3$.

Solution 2: As in Exercises 3.2.7 and 3.2.8, we can show that $\begin{bmatrix}1\\0\end{bmatrix} = -\frac{1}{3}\begin{bmatrix}1\\2\end{bmatrix} + \frac{2}{3}\begin{bmatrix}2\\1\end{bmatrix}$ and $\begin{bmatrix}0\\1\end{bmatrix} = \frac{2}{3}\begin{bmatrix}1\\2\end{bmatrix} - \frac{1}{3}\begin{bmatrix}2\\1\end{bmatrix}$. It follows that

$$\mathbf{A}\mathbf{e}_1 = -\frac{1}{3}\begin{bmatrix}0\\-1\end{bmatrix} + \frac{2}{3}\begin{bmatrix}-3\\2\end{bmatrix} = \begin{bmatrix}-2\\5/3\end{bmatrix} \qquad \text{and} \qquad \mathbf{A}\mathbf{e}_2 = \frac{2}{3}\begin{bmatrix}0\\-1\end{bmatrix} - \frac{1}{3}\begin{bmatrix}-3\\2\end{bmatrix} = \begin{bmatrix}1\\-4/3\end{bmatrix},$$

so $\mathbf{A} = \begin{bmatrix}-2 & 1\\ 5/3 & -4/3\end{bmatrix}$.

3.2.11. Solve the linear system

$$\left[\begin{array}{ccc|c}1 & 1 & 0 & 1\\1 & 0 & 1 & 2\\0 & 1 & 1 & 3\end{array}\right] \rightsquigarrow \left[\begin{array}{ccc|c}1 & 0 & 0 & 0\\0 & 1 & 0 & 1\\0 & 0 & 1 & 2\end{array}\right]$$

to see that $\begin{bmatrix}1\\2\\3\end{bmatrix} = 0\begin{bmatrix}1\\1\\0\end{bmatrix} + \begin{bmatrix}1\\0\\1\end{bmatrix} + 2\begin{bmatrix}0\\1\\1\end{bmatrix}$, so

$$\boldsymbol{T}\begin{bmatrix}1\\2\\3\end{bmatrix} = \boldsymbol{T}\begin{bmatrix}1\\0\\1\end{bmatrix} + 2\boldsymbol{T}\begin{bmatrix}0\\1\\1\end{bmatrix} = \begin{bmatrix}-4\\5\end{bmatrix}.$$

3.2.12. As illustrated in the example before Proposition 3.9, from the RREF of $\mathbf{A}$ one can find a list of vectors that spans $\ker \mathbf{A}$, one vector corresponding to each free variable in the system $\mathbf{A}\mathbf{x} = \mathbf{0}$. Since $\ker \mathbf{A}$ is spanned by a finite list of vectors, it is finite-dimensional.

3.2.13. For $1 \le j \le k$, let $(e_i^j)_{i\ge 1}$ be the sequence with $1 = e_j^j = e_{j+k}^j = e_{j+2k}^j = \cdots$ and $e_i^j = 0$ otherwise. If $(a_i)_{i\ge1} \in V$, then

$$(a_i) = a_1(e_i^1) + \cdots + a_k(e_i^k),$$

and so $(a_i)_{i\ge1} \in \left\langle (e_i^1)_{i\ge1}, \ldots, (e_i^k)_{i\ge1} \right\rangle$.

3.2.14. Let $\lambda \in \mathbb{R}$. A function $f \in C^\infty(\mathbb{R})$ is an eigenvector of $\boldsymbol{D}$ iff $\boldsymbol{D}f = \lambda f$, that is, iff $f'(x) = \lambda f(x)$. The solutions of this differential equation are all of the form $f(x) = ke^{\lambda x}$ for $k \in \mathbb{R}$. Therefore

$$\operatorname{Eig}_\lambda(\boldsymbol{D}) = \left\{ ke^{\lambda x} \,\middle|\, k \in \mathbb{R} \right\} = \left\langle e^{\lambda x} \right\rangle.$$

Each eigenspace is spanned by a single vector, and is therefore finite-dimensional.

3.2.15. Suppose that there are $c_1, \ldots, c_n \in \mathbb{F}$ so that $c_1(v_1+v_2)+c_2(v_2+v_3)+\cdots+c_{n-1}(v_{n-1}+v_n)+c_nv_n = 0$. Rearrange this to get $c_1v_1+(c_1+c_2)v_2+\cdots+(c_{n-2}+c_{n-1})v_{n-1}+(c_{n-1}+c_n)v_n = 0$. Since $(v_1, \ldots, v_n)$ is linearly independent, all the coefficients in this linear combination are 0. In particular, $c_1 = 0$. Then since $c_1 + c_2 = 0$, it follows that $c_2 = 0$. Continuing like this, we see that $c_i = 0$ for each i. Thus the new list is linearly independent.

Next observe that v_n is in both lists. Therefore $v_{n-1} = (v_{n-1} + v_n) - v_n$, so v_{n-1} is in the span of the new list, and then $v_{n-2} = (v_{n-2}+v_{n-1}) - v_{n-1}$ is also in the span of the new list.

Continuing like this, we see that v_i is in the span of the new list for each i. Since $(v_1, \dots, v_n)$ spans V, if $v \in V$ then v is a linear combination of $(v_1, \dots, v_n)$. Since each v_i is a linear combination of the new list, it follows that v is as well. Therefore the new list spans V.

3.2.16. As in the proof of Theorem 3.11, we will find a sublist in which no entry is in the span of the previous entries, which must then be linearly independent by Corollary 3.7.

Since $x^2 - 1$ has different degree than $x + 1$, it is not a multiple of $x + 1$. Next, $x^2 + 2x + 1 = (x^2 - 1) + 2(x + 1)$, so we omit that entry. Now suppose that $a(x^2 - x) + b(x^2 - 1) + c(x + 1) = 0$. This is equivalent to the linear system

$$\begin{aligned} a + b &= 0 \\ -a + c &= 0 \\ -b + c &= 0 \end{aligned}$$

which has the unique solution $a = b = c = 0$. Therefore $(x + 1, x^2 - 1, x^2 - x)$ is linearly independent.

Next observe that $1 = \frac{1}{2}[(x^2 - x) - (x^2 - 1) + (x + 1)]$, $x = \frac{1}{2}[-(x^2 - x) + (x^2 - 1) + (x + 1)]$, and $x^2 = \frac{1}{2}[(x^2 - x) + (x^2 - 1) + (x + 1)]$. Since 1, x, and x^2 span $\mathcal{P}_2(\mathbb{R})$ and are in the span of $(x + 1, x^2 - 1, x^2 - x)$, this list itself spans $\mathcal{P}_2(\mathbb{R})$.

3.2.17. $\mathcal{B}$ is linearly independent by assumption and spans $\langle \mathcal{B} \rangle$ by definition.

3.2.18. As in Exercises 3.1.11, the RREF of $\mathbf{A}$ has a pivot in each column; since $\mathbf{A}$ is $n \times n$ this implies that the RREF is $\mathbf{I}_n$. By Proposition 3.9, the columns of $\mathbf{A}$ form a basis of $\mathbb{F}^n$.

3.2.19. Since $(\mathbf{v}_1, \dots, \mathbf{v}_n)$ is a basis of $\mathbb{F}^n$, $\mathbf{V}$ is invertible by Corollary 3.16. Then $b = \mathbf{I}_n \mathbf{b} = \mathbf{V}\mathbf{V}^{-1}\mathbf{b} = \mathbf{V}\mathbf{x}$, where $\mathbf{x}$ is defined to be $\mathbf{V}^{-1}\mathbf{b}$. Then

$$\mathbf{b} = x_1 \mathbf{v}_1 + \cdots + x_n \mathbf{v}_n$$

by the definition of the matrix-vector product $\mathbf{V}\mathbf{x}$.

3.2.20. Define a linear map $\boldsymbol{T} : V \to V$ by requiring

$$\boldsymbol{T}(v_i) = w_i = \sum_{j=1}^{n} a_{ij} v_j$$

and extending by linearity (as in Theorem 3.14). By Theorem 3.15, it suffices to show that $\boldsymbol{T}$ is an isomorphism.

Let $\mathbf{B} = \mathbf{A}^{-1}$, and define a linear map $\boldsymbol{S}$ by setting

$$\boldsymbol{S}(v_j) = \sum_{k} b_{jk} v_k$$

and extending by linearity. Then $\boldsymbol{S}$ is an inverse to $\boldsymbol{T}$:

$$\boldsymbol{S}(\boldsymbol{T} v_i) = \boldsymbol{S}\left(\sum_{j=1}^{n} a_{ij} v_j\right) = \sum_{j=1}^{n} a_{ij} \boldsymbol{S} v_j = \sum_{j,k} a_{ij} b_{jk} v_k = \sum_{k} [\mathbf{A}\mathbf{B}]_{ik} v_k = v_i,$$

since $[\mathbf{A}\mathbf{B}]_{ik} = 1$ if $i = k$ and is zero otherwise. Similarly,

$$\boldsymbol{T}(\boldsymbol{S}v_j) = \boldsymbol{T}\left(\sum_{k=1}^{n} b_{jk}v_k\right) = \sum_{k=1}^{n} b_{jk}\boldsymbol{T}v_k = \sum_{i,k} b_{jk}a_{ki}v_i = \sum_{k}[\mathbf{B}\mathbf{A}]_{ji}v_i = v_j.$$

Therefore $\boldsymbol{T}$ is an isomorphism.

3.2.21. This is just the special case of Theorem 3.14, with $W = \mathbb{F} = \mathbb{F}^1$.

3.2.22. By Theorem 3.14 there is a unique linear map $\boldsymbol{T} \in \mathcal{L}(V)$ such that $\boldsymbol{T}v_i = w_i$ for each i. Since $(w_1, \ldots, w_n) = (\boldsymbol{T}v_1, \ldots, \boldsymbol{T}v_n)$ is a basis for V, Theorem 3.15 implies that this $\boldsymbol{T}$ is an isomorphism.

3.2.23. The vectors 1 and 2 span $\mathbb{R}$ as a vector space over itself. There is no linear map $\boldsymbol{T} : \mathbb{R} \to \mathbb{R}$ such that $\boldsymbol{T}(1) = \boldsymbol{T}(2) = 1$.

3.2.24. By Theorem 3.15, the assumptions tell us that $\boldsymbol{T}$ is an isomorphism. By Theorem 3.15 again, this means that $(\boldsymbol{T}u_1, \ldots, \boldsymbol{T}u_n)$ is a basis of W.

3.2.25. Suppose $(\mathbf{v}_1, \ldots, \mathbf{v}_m)$ is a basis for $\mathbb{F}^n$. By Proposition 3.9, the RREF of $\mathbf{V} = [\mathbf{v}_1 \cdots \mathbf{v}_m]$ is $\mathbf{I}_n$ (which implies that $m = n$). Therefore for any $\mathbf{w} \in \mathbb{F}^n$, the linear system $\mathbf{V}\mathbf{x} = \mathbf{w}$ has a unique solution $\mathbf{x}$, which means there is a unique way to represent $\mathbf{w}$ as a linear combination of the columns of $\mathbf{V}$, i.e., the $\mathbf{v}_i$.

Conversely, suppose that for every $\mathbf{w} \in \mathbb{F}^n$ there is a unique way to represent $\mathbf{w}$ as a linear combination of the $\mathbf{v}_i$. That means that the system $\mathbf{V}\mathbf{x} = \mathbf{w}$ has a unique solution for each $\mathbf{w} \in \mathbb{F}^n$, which implies that the RREF of $\mathbf{V}$ is $\mathbf{I}_n$. By Proposition 3.9, this implies that $(\mathbf{v}_1, \ldots, \mathbf{v}_m)$ is a basis of $\mathbb{F}^n$.

3.3 Dimension

3.3.1. (a) The RREF of $\begin{bmatrix} 2 & 1 & 0 \\ -1 & 0 & 1 \\ 3 & 1 & -1 \end{bmatrix}$ is $\begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix}$. Thus the first two vectors form a basis of the span, so the dimension is 2.

(b) By the same method, the dimension is 3.

(c) By the same method, the dimension is 2.

(d) By the same method, the dimension is 2.

(e) By the same method, the dimension is 3.

(f) By the same method, the dimension is 2.

(g) The RREF of $\mathbf{A}$ is $\begin{bmatrix} 1 & 0 & 2 & -1 \\ 0 & 1 & -4 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$, so the system $\mathbf{A}\mathbf{x} = \mathbf{0}$ has two free variables. Since each free variable corresponds to one basis vector for the solution space, the dimension of $\ker \mathbf{A}$ is 2.

3.3.2. (a) The second vector is -2 times the first vector, so the space is spanned by the first vector alone, and therefore has dimension 1.

(b) The RREF of $\begin{bmatrix} 1 & 3 & 1 & 1 \\ 2 & 2 & 1 & 1 \\ 3 & 1 & 1 & 1 \end{bmatrix}$ is $\begin{bmatrix} 1 & 0 & \frac{1}{4} & 0 \\ 0 & 1 & \frac{1}{4} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$. Thus the first, second, and fourth vectors form a basis of the span, so the dimension is 3.

(c) By the same method, the dimension is 3.

(d) By the same method, the dimension is 3.

(e) The RREF of $\mathbf{A}$ is $\begin{bmatrix} 1 & 0 & 0 & 5 \\ 0 & 1 & 0 & -7 \\ 0 & 0 & 1 & -1 \end{bmatrix}$, so the system $\mathbf{Ax} = \mathbf{0}$ has one free variable. Since each free variable corresponds to one basis vector for the solution space, the dimension of $\ker \mathbf{A}$ is 1.

(f) The second vector is i times the first one, so the space is spanned by the first vector alone, and therefore has dimension 1 over $\mathbb{C}$.

(g) In this case the second vector is not a multiple of the first by a *real* scalar, so the two vectors are linearly independent, and the space therefore has dimension 2 over $\mathbb{R}$.

3.3.3. Exercise 3.1.14 shows that the functions $\sin(kx)$ for for $k = 1, \dots, n$ are linearly independent, for each n. By Theorem 3.18, this implies that $C(\mathbb{R})$ is infinite-dimensional.

3.3.4. For each n, let $e_n \in \mathbb{F}^\infty$ denote the sequence whose n^{th} entry is 1, and whose other entries are all 0. Then it is easy to check that $(e_1, \dots, e_n)$ is a linearly independent list, so by Theorem 3.18, $\mathbb{F}^\infty$ is infinite-dimensional.

3.3.5. By the Linear Dependence Lemma (in the form of Corollary 3.7), the hypothesis implies that $(v_1, \dots, v_n)$ is linearly independent, and is therefore a basis of $\langle v_1, \dots, v_n \rangle$. Since this space has a basis consisting of n vectors, its dimension is n.

3.3.6. Each f_i is an eigenvector of $\boldsymbol{D}$ with eigenvalue λ_i. Since these eigenvalues are distinct, Theorem 3.8 implies that $(f_1, \dots, f_n)$ are linearly independent, and therefore form a basis of $\langle f_1, \dots, f_n \rangle$. Since this space has a basis consisting of n vectors, its dimension is n.

3.3.7. Let $(v_1, \dots, v_m)$ be a basis of U_1 and $(w_1, \dots, w_n)$ be a basis of U_2. Then if $u \in U_1 + U_2$, we can write

$$u = \sum_{i=1}^m a_i v_i + \sum_{j=1}^n w_j.$$

Thus $U_1 + U_2$ is spanned by $(v_1, \dots, v_m, w_1, \dots, w_n)$, so by Proposition 3.20,

$$\dim(U_1 + U_2) \leq m + n = \dim U_1 + \dim U_2.$$

3.3.8. Let $(u_1, \dots, u_k)$ be a basis for $U_1 \cap U_2$. Then since $(u_1, \dots, u_k)$ is a linearly independent list in U_1, there are vectors $w_1, \dots, w_\ell$ such that $(u_1, \dots, u_k, w_1, \dots, w_\ell)$ is a basis of U_1. Similarly, there are vectors $v_1, \dots, v_m$ such that $(u_1, \dots, u_k, v_1, \dots, v_m)$ is a basis of U_2. If we can show that $(u_1, \ldots, u_k, w_1, \ldots, w_\ell, v_1, \ldots, v_m)$ is a basis of $U_1 + U_2$, then we're done, because

$$\dim(U_1 + U_2) = k + m + \ell = k + \ell + k + m - k = \dim(U_1) + \dim(U_2) - \dim(U_1 \cap U_2).$$

Suppose that $a_1, \ldots, a_k, b_1, \ldots, b_\ell, c_1, \ldots, c_m$ are scalars such that

$$a_1 u_1 + \cdots + a_k u_k + b_1 w_1 + \cdots + b_\ell w_\ell + c_1 v_1 + \cdots + c_m v_m = 0.$$

Rewriting this as

$$a_1 u_1 + \cdots + a_k u_k + b_1 w_1 + \cdots + b_\ell w_\ell = -c_1 v_1 + \cdots - c_m v_m,$$

we can see that the expression on the left is in U_1 and the expression on the right is in U_2, so this vector is an element of $U_1 \cap U_2$. Since $(u_1, \ldots, u_k)$ is a basis of $U_1 \cap U_2$, there are scalars $d_1, \ldots, d_k$ such that

$$c_1 v_1 + \cdots + c_m v_m = d_1 u_1 + \cdots + d_k u_k.$$

But since $(u_1, \ldots, u_k, v_1, \ldots, v_m)$ is a basis for U_2 (hence linearly independent), this means that all of the c_i and all of the d_i are zero. Going back above, this means that

$$a_1 u_1 + \cdots + a_k u_k + b_1 w_1 + \cdots + b_\ell w_\ell = 0,$$

and since $(u_1, \ldots, u_k, w_1, \ldots, w_\ell)$ is a basis for U_1 (hence linearly independent), this means that all the a_i and all the b_i are zero. That is, $(u_1, \ldots, u_k, w_1, \ldots, w_\ell, v_1, \ldots, v_m)$ is linearly independent.

Now let $y \in U_1 + U_2$, so that there are $x_1 \in U_1$ and $x_2 \in U_2$ with $y = x_1 + x_2$. Then there are scalars $a_1, \ldots, a_k, b_1, \ldots, b_\ell$ and $c_1, \ldots, c_k, d_1, \ldots, d_m$ such that

$$\begin{aligned} x_1 &= a_1 u_1 + \cdots + a_k u_k + b_1 w_1 + \cdots + b_\ell w_\ell \\ x_2 &= c_1 u_1 + \cdots + c_k u_k + d_1 v_1 + \cdots + d_m v_m. \end{aligned}$$

Then

$$y = x_1 + x_2 = (a_1 + c_1) u_1 + \cdots + (a_k + c_k) u_k + b_1 w_1 + \cdots + b_\ell w_\ell + d_1 v_1 + \cdots + d_m v_m,$$

and so $(u_1, \ldots, u_k, w_1, \ldots, w_\ell, v_1, \ldots, v_m)$ spans $U_1 + U_2$.

3.3.9. Let $(v_1, \ldots, v_k)$ be a basis for V, and let $w \in \operatorname{range} \boldsymbol{T}$. Then there is a $v \in V$ with $\boldsymbol{T} v = w$. Write v in terms of the basis above:

$$v = c_1 v_1 + \cdots + c_k v_k.$$

Then

$$w = \boldsymbol{T} v = c_1 \boldsymbol{T} v_1 + \cdots + c_k \boldsymbol{T} v_k,$$

and so $(\boldsymbol{T} v_1, \ldots, \boldsymbol{T} v_k)$ spans $\operatorname{range} \boldsymbol{T}$. Then by Proposition 3.20, $\dim \operatorname{range} \boldsymbol{T} \leq k = \dim V$.

3.3.10. Let $n = \dim V$. By Theorem 3.27 there are vectors $v_{k+1}, \dots, v_n \in V$ such that $(v_1, \dots, v_n)$ is a basis of V. By Theorem 3.14 there is a linear map $\boldsymbol{T} : V \to W$ such that $\boldsymbol{T} v_j = w_j$ for $j = 1, \dots, k$ and $\boldsymbol{T} v_j = 0$ for $j = k+1, \dots, n$.

3.3.11. Over $\mathbb{C}$, $\mathbb{C}^2$ has the basis $\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right)$ (this is a special case of $\mathbb{F}^n$ as a vector space over $\mathbb{F}$ for any n and any field $\mathbb{F}$).

Over $\mathbb{R}$, $\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} i \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ i \end{bmatrix} \right)$ clearly spans $\mathbb{C}^2$. If

$$a \begin{bmatrix} 1 \\ 0 \end{bmatrix} + b \begin{bmatrix} i \\ 0 \end{bmatrix} + c \begin{bmatrix} 0 \\ 1 \end{bmatrix} + d \begin{bmatrix} 0 \\ i \end{bmatrix} = \begin{bmatrix} a + bi \\ c + di \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

for $a, b, c, d \in \mathbb{R}$, then $a = b = c = d = 0$, so this list is also linearly independent over $\mathbb{R}$, and is therefore a basis of $\mathbb{C}^2$ over $\mathbb{R}$.

3.3.12. Let m be the number of distinct eigenvalues of $\boldsymbol{T}$, and let $v_1, \dots, v_m$ be eigenvectors of $\boldsymbol{T}$, each corresponding to a different eigenvalue. Then by Theorem 3.8, the v_j are linearly independent, and so by Proposition 3.21, $m \le \dim V$.

3.3.13. Let $\mathbf{v}_1, \dots, \mathbf{v}_n$ be eigenvectors of $\mathbf{A}$ corresponding to distinct eigenvalues. By Theorem 3.8, these vectors are linearly independent. Since $\dim(\mathbb{F}^n) = n$, Theorem 3.26 implies that $(\mathbf{v}_1, \dots, \mathbf{v}_n)$ is a basis of $\mathbb{F}^n$.

3.3.14. Suppose that $f \in \mathbb{C}^\infty(\mathbb{R})$ and that $\boldsymbol{D} f = \lambda f$. This means that $f'(x) = \lambda f(x)$. Every solution of this differential equation is of the form $f(x) = c e^{\lambda x}$ for some $c \in \mathbb{R}$. Therefore $\mathrm{Eig}_\lambda(\boldsymbol{D}) = \left\langle e^{\lambda x} \right\rangle$, which is 1-dimensional.

3.3.15. A 1-dimensional subspace is of the form $\langle v \rangle$ for some $v = (p_1, \dots, p_n) \in V$. By the definition of scalar multiplication in V,

$$\langle v \rangle = \left\{ \frac{(p_1^\lambda, \dots, p_n^\lambda)}{p_1^\lambda + \cdots + p_n^\lambda} \;\middle|\; \lambda \in \mathbb{R} \right\}.$$

Each of these subspaces is a curve inside the simplex parametrized by λ. Each of these curves passes through the point $\left(\frac{1}{n}, \dots, \frac{1}{n}\right)$ (corresponding to $\lambda = 0$).

Suppose that $p_j < p_1$ for all $j > 1$. Then

$$\lim_{\lambda \to \infty} \frac{(p_1^\lambda, \dots, p_n^\lambda)}{p_1^\lambda + \cdots + p_n^\lambda} = \lim_{\lambda \to \infty} \frac{\left(1, \left(\frac{p_2}{p_1}\right)^\lambda, \dots, \left(\frac{p_n}{p_1}\right)^\lambda\right)}{1 + \left(\frac{p_2}{p_1}\right)^\lambda + \cdots + \left(\frac{p_n}{p_1}\right)^\lambda} = (1, 0, \dots, 0) = \mathbf{e}_1.$$

Similarly, if for some i, $p_j < p_i$ for all $j \neq k$, then the curve approaches the vertex $\mathbf{e}_i$ as $\lambda \to \infty$. In the same way, as $\lambda \to -\infty$, the curve approaches the vertex $\mathbf{e}_k$, where p_k is the smallest of the p_j, if there is one.

Suppose now that $p_1 = p_2 = \dots p_m$, and $p_j < p_1$ for all $j > m$. Then

$$\begin{aligned}\lim_{\lambda\to\infty} \frac{(p_1^\lambda, \dots, p_n^\lambda)}{p_1^\lambda + \cdots + p_n^\lambda} &= \lim_{\lambda\to\infty} \frac{\left(1, \dots, 1, \left(\frac{p_{m+1}}{p_1}\right)^\lambda, \dots, \left(\frac{p_n}{p_1}\right)^\lambda\right)}{m + \left(\frac{p_{m+1}}{p_1}\right)^\lambda + \cdots + \left(\frac{p_n}{p_1}\right)^\lambda)} \\ &= \frac{1}{m}(1, \dots, 1, 0, \dots, 0) = \frac{1}{m}(\mathbf{e}_1 + \cdots + \mathbf{e}_m).\end{aligned}$$

Similarly, it follows that, as $\lambda \to \infty$, the curve approaches the average of all the vertices corresponding to the largest entries of v, and as $\lambda \to -\infty$, the curve approaches the average of all the vertices corresponding to the smallest entries of v.

3.3.16. According to Exercise 3.1.18, if $p_1, \dots, p_n$ are distinct prime numbers, then $(\log p_1, \dots, \log p_n)$ is linearly independent over $\mathbb{Q}$. Since there are infinitely many prime numbers, Theorem 3.18 implies that $\mathbb{R}$ is infinite-dimensional over $\mathbb{Q}$.

3.3.17. A plane through the origin in $\mathbb{R}^3$ is a two-dimensional subspace. Since the sum of the dimensions of two such planes is $4 > 3$, according to Lemma 3.22 the intersection of the planes is a nonzero subspace of $\mathbb{R}^3$. Any nonzero subspace contains some nonzero vector $\mathbf{v}$, and hence contains $\langle \mathbf{v} \rangle$, which is a line through the origin.

3.3.18. By Corollary 3.24, if $n = \dim V$, then V is isomorphic to $\mathbb{F}^n$. Since $\mathbb{F}^n$ has q^n elements and there is a bijection between $\mathbb{F}^n$ and V, it follows that V has q^n elements.

3.3.19. (a) A linearly independent list of length 1 can be any nonzero vector zero, hence there are $q^n - 1$ such lists.

Suppose that $k \geq 2$ and $\mathcal{B} = (v_1, \dots, v_{k-1})$ is linearly independent. By the Linear Dependence Lemma, a list $(v_1, \dots, v_k)$ is linearly dependent iff $v_k \in \langle B \rangle$. Since $\mathcal{B}$ is a basis for $\langle \mathcal{B} \rangle$ (Exercise 3.2.17), $\langle \mathcal{B} \rangle$ is $(k-1)$-dimensional and therefore has q^{k-1} elements (Exercise 3.3.18). Therefore there are $q^n - q^{k-1}$ vectors v_k such that $(v_1, \dots, v_k)$ is linearly independent.

So if we make the induction hypothesis that there are exactly $(q^n - 1) \cdots (q^n - q^{k-2})$ linearly independent lists of length $k-1$ in $\mathbb{F}^n$, this implies that there are $(q^n - 1) \cdots (q^n - q^{k-1})$ linearly independent lists of length k. By induction, it follows that this holds for all $1 \leq k \leq n$.

Since a basis of $\mathbb{F}^n$ is precisely a linearly independent list of length n, it follows that there are $(q^n - 1) \cdots (q^n - q^{n-1})$ distinct bases of $\mathbb{F}^n$.

(b) It is clear that there are q^{n^2} matrices in $\mathrm{M}_n(\mathbb{F})$.

A matrix in $\mathrm{M}_n(\mathbb{F})$ is invertible if and only if its columns for a basis of $\mathbb{F}^n$. Therefore the number of invertible matrices in $\mathrm{M}_n(\mathbb{F})$ is the same as the number of bases of $\mathbb{F}^n$, which is the same as the number of linearly independent lists of n vectors in $\mathbb{F}^n$.

3.3.20. Suppose V is a finite-dimensional nonzero vector space. Then V contains a nonzero vector v. The list (v) is linearly independent, so by Theorem 3.27, it can be extended to a basis $\mathcal{B}$ of V. Thus V possesses a basis.

3.3.21. If V is infinite-dimensional, then V is nonzero, since otherwise it would be spanned by the finite list (0). Therefore there exists a nonzero vector $v_1 \in V$, and then the list (v_1) is linearly independent.

Suppose now that V contains a linearly independent list $(v_1, \dots, v_n)$. Since V is infinite-dimensional, $V \neq \langle v_1, \dots, v_n \rangle$, so there exists a vector $v_{n+1} \notin \langle v_1, \dots, v_n \rangle$. By the Linear Dependence Lemma, $(v_1, \dots, v_{n+1})$ is also linearly independent.

Therefore by induction, V contains a linearly independent list of length n for each $n \in \mathbb{N}$.

3.3.22. Suppose that U_1 and U_2 are subspaces of V such that $\dim U_1 + \dim U_2 > \dim V$. By Exercise 3.3.8,

$$\begin{aligned}\dim(U_1 \cap U_2) &= \dim(U_1) + \dim(U_2) - \dim(U_1 + U_2) \\ &> \dim V - \dim(U_1 + U_2) \geq 0,\end{aligned}$$

since $U_1 + U_2$ is a subspace of V. Since $\dim(U_1 \cap U_2) > 0$, $U_1 \cap U_2 \neq \{0\}$.

3.3.23. Since $\mathcal{B} = (v_1, \dots, v_n)$ is linearly independent, Theorem 3.27 implies that $\mathcal{B}$ can be extended to a basis $\mathcal{B}'$ of V. But since $\mathcal{B}$ already has n elements and every basis of V contains n elements, we must have $\mathcal{B}' = \mathcal{B}$, and hence $\mathcal{B}$ is a basis of V.

3.4 Rank and nullity

3.4.1. (a) The two rows are not multiples of each other, so they are linearly independent and thus the rank is 2, so the nullity is $3 - 2 = 1$.

(b) The RREF is $\begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$. There are three pivots and one free variable, so the rank is 3 and the nullity is 1.

(c) By the same method, the rank is 2 and the nullity is 3.

(d) By the same method, the rank is 3 and the nullity is 0.

(e) By the same method, the rank is 2 and the nullity is 1.

3.4.2. (a) The two rows are not multiples of each other, so they are linearly independent and thus the rank is 2, so the nullity is $4 - 2 = 2$.

(b) All the columns are multiples of the last one, so the rank is 1 and the nullity is therefore $3 - 1 = 2$.

(c) The RREF is $\begin{bmatrix} 1 & 0 & -\frac{1}{7} \\ 0 & 1 & -\frac{2}{7} \\ 0 & 0 & 0 \end{bmatrix}$. There are two pivots and one free variable, so the rank is 2 and the nullity is 1.

(d) By the same method, the rank is 3 and the nullity is 0.

(e) By the same method, the rank is 2 and the nullity is 2.

3.4.3. Recall that the dimension of $\mathcal{P}_n(\mathbb{R})$ is $n+1$. A polynomial $p(x)$ is in the kernel of $\boldsymbol{D}$ if and only if $p'(x) = 0$; that is, if and only if p is constant. The subspace of polynomials consisting of the constant is 1-dimensional (it is spanned by $p(x) = 1$), and so the nullity of $\boldsymbol{D}$ is 1, and the rank is therefore n by the Rank–Nullity Theorem. (You can also see the rank by noting that the range of $\boldsymbol{D}$ is exactly the polynomials of degree at most $n-1$).

3.4.4. The rank of the given matrix is 1: every row is a multiple of $\begin{bmatrix}1 & 2 & \cdots & n\end{bmatrix}$. By the Rank–Nullity Theorem, this means that the nullity must be $n-1$.

3.4.5. Each column of $\mathbf{AB}$ is a linear combination of columns of $\mathbf{A}$, so $C(\mathbf{AB}) \subseteq C(\mathbf{A})$, and thus $\operatorname{rank}\mathbf{AB} = \dim C(\mathbf{AB}) \leq \dim C(\mathbf{A}) = \operatorname{rank}\mathbf{A}$.

Next, if $\mathbf{B} \in \mathrm{M}_{n,p}(\mathbb{F})$, then

$$C(\mathbf{AB}) = \{(\mathbf{AB})\mathbf{x} \mid \mathbf{x} \in \mathbb{F}^p\} = \{\mathbf{Ay} \mid \mathbf{y} \in C(\mathbf{B})\}$$

and so $\dim C(\mathbf{AB}) \leq \dim C(\mathbf{B}) = \operatorname{rank}\mathbf{B}$ by Exercise 3.3.9.

Alternatively, either of the two halves implies the other, since

$$\operatorname{rank}\mathbf{AB} = \operatorname{rank}(\mathbf{AB})^{\mathrm{T}} = \operatorname{rank}\mathbf{B}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}},$$

$\operatorname{rank}\mathbf{A}^{\mathrm{T}} = \operatorname{rank}\mathbf{A}$, and $\operatorname{rank}\mathbf{B}^{\mathrm{T}} = \operatorname{rank}\mathbf{B}$.

3.4.6. By definition, $\boldsymbol{T}$ is surjective iff $\operatorname{range}\boldsymbol{T} = W$. By Theorem 3.29, this is true iff $\operatorname{rank}\boldsymbol{T} = \dim\operatorname{range}\boldsymbol{T} = \dim W$.

3.4.7. Suppose first that $\mathbf{x} = (x_1, \dots, x_5) \in S$ and $\mathbf{v}_0 = (y_1, \dots, y_5) \in S$. We need to show that if $\mathbf{z} = \mathbf{x} - \mathbf{v}_0$ then $\mathbf{z}$ satisfies $3z_1 + 2z_4 + z_5 = 0$ and $z_2 - z_3 - 5z_5 = 0$. We know that $3x_1 + 2x_4 + x_5 = 3y_1 + 2y_4 + y_5 = 10$ and $x_2 - x_3 - 5x_5 = y_2 - y_3 - 5y_5 = 7$, and so $\mathbf{z}$ does indeed have these properties.

Now suppose that $\mathbf{z}$ satisfies $3z_1+2z_4+z_5 = 0$ and $z_2-z_3-5z_5 = 0$, and $\mathbf{v}_0 = (y_1, \dots, y_5) \in S$. We need to show that $\mathbf{z} = \mathbf{x} - \mathbf{v}_0$ for some $\mathbf{x} \in S$. That is, we need to show that if we set $\mathbf{x} = \mathbf{z} + \mathbf{v}_0$, then $\mathbf{x} \in S$. Since $3y_1 + 2y_4 + y_5 = 10$ and $y_2 - y_3 - 5y_5 = 7$, it follows that $3x_1 + 2x_4 + x_5 = 10$ and $x_2 - x_3 - 5x_5 = 7$, ans so indeed $\mathbf{x} \in S$.

3.4.8. Write $\mathbf{A} = \begin{bmatrix} | & & | \\ \mathbf{a}_1 & \cdots & \mathbf{a}_n \\ | & & | \end{bmatrix}$, and let $(\mathbf{v}_1, \dots, \mathbf{v}_r)$ be a basis for $C(\mathbf{A})$. Then for each j, there are scalars $w_{1j}, \dots, w_{rj}$ such that

$$\mathbf{a}_j = \sum_{k=1}^{r} w_{kj}\mathbf{v}_k;$$

that is,

$$\begin{aligned}
\mathbf{A} &= \begin{bmatrix} | & & | \\ w_{11}\mathbf{v}_1 + \cdots + w_{r1}\mathbf{v}_r & \cdots & w_{1n}\mathbf{v}_1 + \cdots + w_{rn}\mathbf{v}_r \\ | & & | \end{bmatrix} \\
&= \begin{bmatrix} | & & | \\ w_{11}\mathbf{v}_1 & \cdots & w_{1n}\mathbf{v}_1 \\ | & & | \end{bmatrix} + \cdots + \begin{bmatrix} | & & | \\ w_{r1}\mathbf{v}_r & \cdots & w_{rn}\mathbf{v}_r \\ | & & | \end{bmatrix} \\
&= \mathbf{v}_1\mathbf{w}_1^{\mathrm{T}} + \cdots + \mathbf{v}_r\mathbf{w}_r^{\mathrm{T}},
\end{aligned}$$

where

$$\mathbf{w}_i = \begin{bmatrix} w_{i1} \\ \vdots \\ w_{in} \end{bmatrix}.$$

3.4.9. (a) By Exercise 2.5.11(a),

$$\operatorname{range}(\boldsymbol{S}+\boldsymbol{T}) \subseteq \operatorname{range}\boldsymbol{S} + \operatorname{range}\boldsymbol{T},$$

and so

$$\operatorname{rank}(\boldsymbol{S}+\boldsymbol{T}) = \dim\operatorname{range}(\boldsymbol{S}+\boldsymbol{T}) \le \dim\left(\operatorname{range}\boldsymbol{S} + \operatorname{range}\boldsymbol{T}\right).$$

By Exercise 3.3.7,

$$\dim\left(\operatorname{range}\boldsymbol{S} + \operatorname{range}\boldsymbol{T}\right) \le \dim\left(\operatorname{range}\boldsymbol{S}\right) + \dim\left(\operatorname{range}\boldsymbol{T}\right) = \operatorname{rank}\boldsymbol{S} + \operatorname{rank}\boldsymbol{T};$$

putting these two inequalities together completes the proof.

(b) This follows by applying part (a) to the linear maps from $\mathbb{F}^n$ to $\mathbb{F}^m$ represented by $\mathbf{A}$ and $\mathbf{B}$. Alternatively, it follows in the same way as part (a), using part (b) of Exercise 2.5.11.

3.4.10. By the Rank–Nullity Theorem, $\dim V = \operatorname{rank}\boldsymbol{T} + \operatorname{null}\boldsymbol{T}$, and so $\dim V = \operatorname{rank}\boldsymbol{T}$ if and only if $\operatorname{null}\boldsymbol{T} = 0$; by Theorem 2.37, this is true if and only if $\boldsymbol{T}$ is injective.

3.4.11. If $\mathbf{A}\mathbf{x} = \mathbf{b}$ is consistent but does not have a unique solution, then $\ker\mathbf{A} \neq \{\mathbf{0}\}$, so $\operatorname{null}\mathbf{A} > 0$. By the Rank–Nullity Theorem, this implies that $\operatorname{rank}\mathbf{A} = 5 - \operatorname{null}\mathbf{A} < 5$, so $C(\mathbf{A}) \neq \mathbb{F}^5$. Thus there is some $\mathbf{c} \in \mathbb{F}^5$ such that $\mathbf{c} \notin C(\mathbf{A})$, and so $\mathbf{A}\mathbf{x} = \mathbf{c}$ is inconsistent.

3.4.12. The system $\mathbf{A}\mathbf{x} = \mathbf{b}$ is consistent if and only if $\mathbf{b} \in C(\mathbf{A})$, so saying that $\mathbf{A}\mathbf{x} = \mathbf{b}$ is consistent if and only if $\mathbf{b} \in P$ is the same thing as saying that $C(\mathbf{A}) = P$. This means that the rank of $\mathbf{A}$ is 2, and (by the Rank–Nullity Theorem), this means that $\operatorname{null}(\mathbf{A}) = 1$. That is, the set of solutions to $\mathbf{A}\mathbf{x} = \mathbf{0}$ is a one-dimensional subspace of $\mathbb{R}^3$: a line through the origin.

3.4.13. If $\operatorname{rank}\boldsymbol{T} = \dim V$, then by the Rank–Nullity Theorem, $\operatorname{null}\boldsymbol{T} = 0$ and so $\boldsymbol{T}$ is injective. If $\operatorname{rank}\boldsymbol{T} = \dim(\operatorname{range}\boldsymbol{T}) = \dim W$, then $\operatorname{range}\boldsymbol{T} = \dim W$ and so $\boldsymbol{T}$ is surjective.

3.4.14. (a) Note that since $\dim(V) = n$, all operators on V have rank less than or equal to n. Suppose that $\operatorname{rank}(\boldsymbol{TS}) = n$. Then $\boldsymbol{TS}$ is surjective. But this means that both $\boldsymbol{S}$ and $\boldsymbol{T}$ are themselves surjective: firstly, $\operatorname{range} \boldsymbol{TS} \subseteq \operatorname{range} \boldsymbol{T}$, so $\boldsymbol{T}$ must be surjective. Also, if we think of $\boldsymbol{T}$ restricted to $\operatorname{range} \boldsymbol{S}$, we see that $\operatorname{rank} \boldsymbol{TS} \leq \operatorname{rank} \boldsymbol{S}$, so if $\operatorname{rank} \boldsymbol{TS} = n$ then $\operatorname{rank} \boldsymbol{S} = n$. But now that we know both $\boldsymbol{T}$ and $\boldsymbol{S}$ are surjective, so is $\boldsymbol{ST}$ (if $v \in V$, we can write $v = \boldsymbol{S}w$ for some $w \in W$, and then we can write $w = \boldsymbol{T}y$), and thus $\operatorname{rank} \boldsymbol{ST} = n$.

(b) Suppose that 0 is an eigenvalue of $\boldsymbol{ST}$. Then $\boldsymbol{ST}$ is not injective, so it is not surjective, so $\boldsymbol{TS}$ is also not surjective (by the previous part). So $\boldsymbol{TS}$ is not injective; that is, 0 is an eigenvalue of $\boldsymbol{TS}$.

3.4.15. Suppose that $\boldsymbol{T}$ has m distinct non-zero eigenvalues $\lambda_1, \ldots, \lambda_m$. Then the corresponding eigenvectors $v_1, \ldots, v_m$ are linearly independent, and hence so are the non-zero multiples $\lambda_1 v_1, \ldots, \lambda_m v_m$. Therefore $\operatorname{range} \boldsymbol{T}$ contains the m linearly independent vectors $\lambda_i v_i = \boldsymbol{T}(v_i)$ for $i = 1, \ldots, m$, and so $\operatorname{rank} \boldsymbol{T} \geq m$.

3.4.16. The set of $n \times n$ matrices is n^2-dimensional (take as a basis the matrices with one non-zero entry; there are n^2 such matrices and they are clearly linearly independent and span the space of all matrices). Now, the set in the problem has 2n linear constraints, but they are not independent. If we view the matrix entries a_{ij} as the entries of an n^2-dimensional vector by writing

$$(a_{11}, \ldots, a_{n1}, a_{12}, \ldots, a_{n2}, \ldots, a_{1n}, \ldots, a_{nn}),$$

then the constraints in the given set are exactly the requirement that this long vector be an element of the null space of the following matrix (I'm adding bars after every n entries and a vertical line separating the top n rows from the bottom n rows to make it easier to tell what's in the matrix; they don't mean anything)

$$\left[\begin{array}{cccc|cccc|c|cccc}
1 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 & \cdots & 1 & 0 & \cdots & 0 \\
0 & 1 & \cdots & 0 & 0 & 1 & \cdots & 0 & \cdots & 0 & 1 & \cdots & 0 \\
\vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & & \vdots & \vdots & & \vdots \\
0 & 0 & \cdots & 1 & 0 & 0 & \cdots & 1 & \cdots & 0 & 0 & \cdots & 1 \\
\hline
1 & 1 & \cdots & 1 & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 \\
0 & 0 & \cdots & 0 & 1 & 1 & \cdots & 1 & \cdots & 0 & 0 & \cdots & 0 \\
\vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & & \vdots & \vdots & & \vdots \\
0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & \cdots & 1 & 1 & \cdots & 1
\end{array}\right]$$

Now, the last row is equal to the sum of the first n rows minus the sum of the next $n-1$ rows, so the rank of this matrix is at most $2n - 1$. In fact, the rank is exactly $2n - 1$: firstly, the first n rows are clearly linearly independent, since the positions in which they have non-zero entries are all mutually disjoint. Moreover, the $n + 1$st row is not in the span of the first n rows: think of the rows as broken into n chunks of size n as illustrated with the bars. Any linear combination of the first n rows has every chunk the same (because this is true of all of the first n rows). But this is not true of the first row below the line, and so we now have $n + 1$ linearly independent rows. Similarly, any linear combination of the first $n+1$ rows must have the second and third chunks the same (this was true of the span of the first n, and the next had all zeroes from the $n+1$st entry on). Since this is not true of the $n+2$nd row of the matrix, it is independent of the rows above it. Continuing in this way, the span of the first $n+k-1$ rows always contains a pattern like this violated by the $n+k$th row, as long as $k<n$.

Since the rank of this matrix is $2n-1$, its nullity must be n^2-2n+1, and this is exactly our set of interest.

3.4.17. Row-reducing $\mathbf{A}$ is the same whether we do it in $\mathbb{F}$ or in $\mathbb{K}$, so the RREF of $\mathbf{A}$ is the same over either field. Since the rank and nullity of $\mathbf{A}$ depend only on the RREF, these quantities are the same whether we view $\mathbf{A}$ as an element of $\mathrm{M}_{m,n}(\mathbb{F})$ or $\mathrm{M}_{m,n}(\mathbb{K})$.

3.4.18. (a) Given $\mathbf{v}$ and $\mathbf{w}$, the matrix $\mathbf{v}\mathbf{w}^{\mathrm{T}}$ has rank 0 or 1, since each column is a scalar multiple of $\mathbf{v}$. (Note that if $\mathbf{v}$ or $\mathbf{w}$ is $\mathbf{0}$, then the rank is 0.) By part Exercise 3.4.9, this implies that

$$\operatorname{rank}\mathbf{A}=\operatorname{rank}\sum_{i=1}^{r}\mathbf{v}_i\mathbf{w}_i^{\mathrm{T}}\le\sum_{i=1}^{r}\operatorname{rank}\mathbf{v}_i\mathbf{w}_i^{\mathrm{T}}\le r.$$

(b) Suppose that $\operatorname{rank}\mathbf{A}=r$, and let $(\mathbf{v}_1,\ldots,\mathbf{v}_r)$ be a basis for $C(\mathbf{A})$. Then for each j, the jth column of $\mathbf{A}$ is some linear combination of these basis vectors, say $\sum_{i=1}^{r}w_{ij}\mathbf{v}_i$. Defining $\mathbf{w}_i$ to have entries $w_{i1},\ldots,w_{in}$, we then have

$$\mathbf{A}=\sum_{i=1}^{r}\mathbf{v}_i\mathbf{w}_i^{\mathrm{T}},$$

and so

$$\mathbf{A}^{\mathrm{T}}=\sum_{i=1}^{r}\mathbf{w}_i\mathbf{v}_i^{\mathrm{T}}.$$

Then by part (a), $\operatorname{rank}\mathbf{A}^{\mathrm{T}}\le r=\operatorname{rank}\mathbf{A}$.

Applying the same argument to $\mathbf{A}^{\mathrm{T}}$, we get $\operatorname{rank}\mathbf{A}=\operatorname{rank}\mathbf{A}^{\mathrm{TT}}\le\operatorname{rank}\mathbf{A}^{\mathrm{T}}$ as well, and so $\operatorname{rank}\mathbf{A}^{\mathrm{T}}=\operatorname{rank}\mathbf{A}$.

3.4.19. (a) Let $\mathbf{a}_j$ be the jth column of $\mathbf{A}$; since $\mathbf{a}_j\in C(\mathbf{A})$, there are coefficients b_{ij} so that $\mathbf{a}_j=b_{1j}\mathbf{c}_1+\cdots+b_{rj}\mathbf{c}_r$, which is the same as saying $\mathbf{a}_j=\mathbf{C}\mathbf{b}_j$, where $\mathbf{b}_j=\begin{bmatrix}b_{1j}&\cdots&b_{rj}\end{bmatrix}^{\mathrm{T}}$. This in turn is equivalent to $\mathbf{A}=\mathbf{C}\mathbf{B}$, where $\mathbf{B}\in\mathrm{M}_{r,n}(\mathbb{F})$ is the matrix whose jth column is $\mathbf{b}_j$.

(b) By Corollary 2.32, $\operatorname{rank}\mathbf{A}^{\mathrm{T}}=\operatorname{rank}(\mathbf{B}^{\mathrm{T}}\mathbf{C}^{\mathrm{T}})\le\operatorname{rank}(\mathbf{B}^{\mathrm{T}})$. By Theorem 3.32, $\operatorname{rank}(\mathbf{B}^{\mathrm{T}})\le r$ since $\mathbf{B}^{\mathrm{T}}\in\mathrm{M}_{n,r}(\mathbb{F})$. Combining these, $\operatorname{rank}\mathbf{A}^{\mathrm{T}}\le r=\operatorname{rank}\mathbf{A}$.

(c) By part (b) applied to $\mathbf{A}$, $\operatorname{rank}\mathbf{A}^{\mathrm{T}}\le\operatorname{rank}\mathbf{A}$. On the other hand, by part (b) applied to $\mathbf{A}^{\mathrm{T}}$, $\operatorname{rank}\mathbf{A}=\operatorname{rank}(\mathbf{A}^{\mathrm{T}})^{\mathrm{T}}\le\operatorname{rank}\mathbf{A}^{\mathrm{T}}$. Combining these, we have $\operatorname{rank}\mathbf{A}^{\mathrm{T}}=\operatorname{rank}\mathbf{A}$.

3.4.20. We need to show that if $\ker\boldsymbol{T}=\{0\}$, then $\operatorname{rank}\boldsymbol{T}=\dim V$. Let $\mathcal{B}=(v_1,\ldots,v_n)$ be a basis of V. We claim that $\mathcal{B}'=(\boldsymbol{T}v_1,\ldots,\boldsymbol{T}v_n)$ is a basis of $\operatorname{range}\boldsymbol{T}$.

Since $\boldsymbol{T}$ is injective, Exercise 3.1.9 shows that $\mathcal{B}'$ is linearly independent. If $w\in\operatorname{range}\boldsymbol{T}$, then $w=\boldsymbol{T}v$ for some $v\in V$. Then $v=a_1v_1+\cdots+a_nv_n$ for some $a_i\in\mathbb{F}$, so by linearity

$$w=\boldsymbol{T}v=a_1\boldsymbol{T}v_1+\cdots+a_n\boldsymbol{T}v_n.$$

Therefore $\mathcal{B}'$ spans range $\boldsymbol{T}$. Thus $\mathcal{B}'$ is a basis of range $\boldsymbol{T}$, and so rank $\boldsymbol{T} = n = \dim V$.

3.4.21. Statement: Let $\mathbf{A} \in \mathrm{M}_{m,n}(\mathbb{F})$. If $n > m$ then $\ker \mathbf{A} \neq \{\mathbf{0}\}$, and if $m > n$ then $C(\mathbf{A}) \subsetneq \mathbb{F}^m$.

Proof: If $n > m$, then by Theorem 3.32, $\operatorname{rank} \mathbf{A} \leq m$, and so by the Rank–Nullity Theorem, $\operatorname{null} \mathbf{A} = n - \operatorname{rank} \mathbf{A} \geq n - m > 0$.

If $m > n$, then by Theorem 3.32, $\operatorname{rank} \mathbf{A} \leq n < m$.

3.5 Coordinates

3.5.1. (a) We solve the linear system $x \begin{bmatrix} 2 \\ -3 \end{bmatrix} + y \begin{bmatrix} -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$: the RREF is $\left[\begin{array}{cc|c} 1 & 0 & 2 \\ 0 & 1 & 3 \end{array}\right]$, yielding the solution $x = 2$, $y = 3$. It follows that $\left[\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right]_{\mathcal{B}} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$.

(b) By the same method, $\left[\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right]_{\mathcal{B}} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$.

(c) By the same method, $\left[\begin{bmatrix} 2 \\ 3 \end{bmatrix}\right]_{\mathcal{B}} = \begin{bmatrix} 7 \\ 12 \end{bmatrix}$. Alternatively,

$$\left[\begin{bmatrix} 2 \\ 3 \end{bmatrix}\right]_{\mathcal{B}} = 2 \left[\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right]_{\mathcal{B}} + \left[\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right]_{\mathcal{B}},$$

and the answer follows from parts (a) and (b).

(d) By the same method, $\left[\begin{bmatrix} 4 \\ -5 \end{bmatrix}\right]_{\mathcal{B}} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$.

(e) By the same method, $\left[\begin{bmatrix} -6 \\ 1 \end{bmatrix}\right]_{\mathcal{B}} = \begin{bmatrix} -11 \\ -16 \end{bmatrix}$.

3.5.2. (a) We solve the linear system $x \begin{bmatrix} 1 \\ 2 \end{bmatrix} + y \begin{bmatrix} 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$: the RREF is $\left[\begin{array}{cc|c} 1 & 0 & -2 \\ 0 & 1 & 1 \end{array}\right]$, yielding the solution $x = 2$, $y = 3$. It follows that $\left[\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right]_{\mathcal{B}} = \begin{bmatrix} -2 \\ 1 \end{bmatrix}$.

(b) By the same method, $\left[\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right]_{\mathcal{B}} = \frac{1}{2} \begin{bmatrix} 3 \\ -1 \end{bmatrix}$.

(c) By the same method, $\left[\begin{bmatrix} 2 \\ 3 \end{bmatrix}\right]_{\mathcal{B}} = \frac{1}{2} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. Alternatively,

$$\left[\begin{bmatrix} 2 \\ 3 \end{bmatrix}\right]_{\mathcal{B}} = 2 \left[\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right]_{\mathcal{B}} + \left[\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right]_{\mathcal{B}},$$

and the answer follows from parts (a) and (b).

(d) By the same method, $\left[\begin{bmatrix} 4 \\ -5 \end{bmatrix}\right]_{\mathcal{B}} = \frac{1}{2} \begin{bmatrix} -31 \\ 13 \end{bmatrix}$.

(e) By the same method, $\left[\begin{bmatrix}-6\\1\end{bmatrix}\right]_{\mathcal{B}} = \frac{1}{2}\begin{bmatrix}27\\-13\end{bmatrix}$.

3.5.3. (a) We solve the linear system $x\begin{bmatrix}1\\-1\\0\end{bmatrix} + y\begin{bmatrix}1\\0\\1\end{bmatrix} + z\begin{bmatrix}0\\1\\-1\end{bmatrix} = \begin{bmatrix}1\\0\\0\end{bmatrix}$: the RREF is $\left[\begin{array}{ccc|c}1&0&0&\frac{1}{2}\\0&1&0&\frac{1}{2}\\0&0&1&\frac{1}{2}\end{array}\right]$, yielding the solution $x = y = z = \frac{1}{2}$. It follows that $\left[\begin{bmatrix}1\\0\\0\end{bmatrix}\right]_{\mathcal{B}} = \frac{1}{2}\begin{bmatrix}1\\1\\1\end{bmatrix}$.

(b) By the same method, $\left[\begin{bmatrix}0\\0\\1\end{bmatrix}\right]_{\mathcal{B}} = \frac{1}{2}\begin{bmatrix}-1\\1\\-1\end{bmatrix}$.

(c) By the same method, $\left[\begin{bmatrix}-4\\0\\2\end{bmatrix}\right]_{\mathcal{B}} = \begin{bmatrix}-3\\-1\\-3\end{bmatrix}$.

(d) By the same method, $\left[\begin{bmatrix}1\\2\\-3\end{bmatrix}\right]_{\mathcal{B}} = \begin{bmatrix}1\\0\\3\end{bmatrix}$.

(e) By the same method, $\left[\begin{bmatrix}6\\4\\-1\end{bmatrix}\right]_{\mathcal{B}} = \frac{1}{2}\begin{bmatrix}9\\3\\1\end{bmatrix}$.

3.5.4. (a) We solve the linear system $x\begin{bmatrix}1\\-1\\1\end{bmatrix} + y\begin{bmatrix}1\\2\\-1\end{bmatrix} + z\begin{bmatrix}0\\2\\-1\end{bmatrix} = \begin{bmatrix}1\\0\\0\end{bmatrix}$: the RREF is $\left[\begin{array}{ccc|c}1&0&0&0\\0&1&0&1\\0&0&1&-1\end{array}\right]$, yielding the solution $x = 0$, $y = 1$, $y = -1$. It follows that $\left[\begin{bmatrix}1\\0\\0\end{bmatrix}\right]_{\mathcal{B}} = \begin{bmatrix}0\\1\\-1\end{bmatrix}$.

(b) By the same method, $\left[\begin{bmatrix}0\\0\\1\end{bmatrix}\right]_{\mathcal{B}} = \begin{bmatrix}2\\-2\\3\end{bmatrix}$.

(c) By the same method, $\left[\begin{bmatrix}-4\\0\\2\end{bmatrix}\right]_{\mathcal{B}} = \begin{bmatrix}4\\-8\\10\end{bmatrix}$.

(d) By the same method, $\left[\begin{bmatrix}1\\2\\-3\end{bmatrix}\right]_{\mathcal{B}} = \begin{bmatrix}-4\\5\\6\end{bmatrix}$.

(e) By the same method, $\left[\begin{bmatrix}6\\4\\-1\end{bmatrix}\right]_{\mathcal{B}} = \begin{bmatrix}-2\\8\\-11\end{bmatrix}$.

3.5.5. (a) $S\begin{bmatrix}2\\1\end{bmatrix}=\begin{bmatrix}1\\3\\0\end{bmatrix}$ and $S\begin{bmatrix}1\\1\end{bmatrix}=\begin{bmatrix}1\\0\\-4\end{bmatrix}$, so $[S]_{\mathcal{B},\mathcal{E}}=\begin{bmatrix}1&1\\3&0\\0&-4\end{bmatrix}$.

(b) We need to express the columns $S\mathbf{e}_1$ and $S\mathbf{e}_2$ of $[S]_{\mathcal{E},\mathcal{E}}$ in terms of the basis $\mathbf{C}$:

$$\begin{bmatrix}0\\3\\4\end{bmatrix}=\begin{bmatrix}1\\3\\0\end{bmatrix}-\begin{bmatrix}1\\0\\4\end{bmatrix}\qquad\text{and}\qquad\begin{bmatrix}1\\-3\\8\end{bmatrix}=-\begin{bmatrix}1\\3\\0\end{bmatrix}+2\begin{bmatrix}1\\0\\-4\end{bmatrix},$$

so

$$\left[\begin{bmatrix}0\\3\\4\end{bmatrix}\right]_{\mathcal{C}}=\begin{bmatrix}1\\0\\-1\end{bmatrix}\qquad\text{and}\qquad\left[\begin{bmatrix}1\\-3\\8\end{bmatrix}\right]_{\mathcal{C}}=\begin{bmatrix}-1\\0\\2\end{bmatrix}.$$

Therefore $[S]_{\mathcal{E},\mathcal{C}}=\begin{bmatrix}1&-1\\0&0\\-1&2\end{bmatrix}$.

(c) We express the vectors found in part (a) in terms of the basis $\mathcal{C}$:

$$\left[S\begin{bmatrix}2\\1\end{bmatrix}\right]_{\mathcal{C}}=\begin{bmatrix}1\\0\\0\end{bmatrix}\qquad\text{and}\qquad\left[S\begin{bmatrix}1\\1\end{bmatrix}\right]_{\mathcal{C}}=\begin{bmatrix}0\\0\\1\end{bmatrix}.$$

Therefore $[S]_{\mathcal{B},\mathcal{C}}=\begin{bmatrix}1&0\\0&0\\0&1\end{bmatrix}$.

(d) By the same method as part (a), $[T]_{\mathcal{C},\mathcal{E}}=\begin{bmatrix}1&1&0\\0&1&-1\end{bmatrix}$.

(e) By the same method as part (b), $[T]_{\mathcal{E},\mathcal{B}}=\begin{bmatrix}1&0&0\\2&-1&1\end{bmatrix}$.

(f) By the same method as part (c), $[T]_{\mathcal{C},\mathcal{B}}=\begin{bmatrix}1&0&1\\-1&1&-2\end{bmatrix}$.

3.5.6. (a) $S\begin{bmatrix}2\\3\end{bmatrix}=\begin{bmatrix}1\\1\\0\end{bmatrix}$ and $S\begin{bmatrix}3\\5\end{bmatrix}\begin{bmatrix}1\\0\\1\end{bmatrix}$, so $[S]_{\mathcal{B},\mathcal{E}}=\begin{bmatrix}1&1\\1&0\\0&1\end{bmatrix}$.

(b) We need to express the columns $S\mathbf{e}_1$ and $S\mathbf{e}_2$ of $[S]_{\mathcal{E},\mathcal{E}}$ in terms of the basis $\mathbf{C}$.

$$S\mathbf{e_1}=\begin{bmatrix}2\\5\\-3\end{bmatrix}\qquad\text{and}\qquad S\mathbf{e_2}=\begin{bmatrix}-1\\-3\\2\end{bmatrix}.$$

We must now express these in terms of the basis $\mathcal{C}$:

$$\left[\begin{array}{ccc|c}1&1&0&2\\1&0&1&5\\0&1&1&-3\end{array}\right]\rightsquigarrow\left[\begin{array}{ccc|c}1&0&1&5\\0&1&1&-3\\0&0&-2&0\end{array}\right]\rightsquigarrow\left[\begin{array}{ccc|c}1&0&0&5\\0&1&0&-3\\0&0&1&0\end{array}\right],$$

so $[\boldsymbol{S}\mathbf{e_1}]_{\mathcal{C}} = \begin{bmatrix} 5 \\ -3 \\ 0 \end{bmatrix}$, and

$$\left[\begin{array}{ccc|c} 1 & 1 & 0 & -1 \\ 1 & 0 & 1 & -3 \\ 0 & 1 & 1 & 2 \end{array}\right] \rightsquigarrow \left[\begin{array}{ccc|c} 1 & 0 & 1 & -3 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & -2 & 0 \end{array}\right] \rightsquigarrow \left[\begin{array}{ccc|c} 1 & 0 & 0 & -3 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 0 \end{array}\right],$$

so $[\boldsymbol{S}\mathbf{e_2}]_{\mathcal{C}} = \begin{bmatrix} -3 \\ 2 \\ 0 \end{bmatrix}$, and finally

$$[\boldsymbol{S}]_{\mathcal{E},\mathcal{C}} = \begin{bmatrix} 5 & -3 \\ -3 & 2 \\ 0 & 0 \end{bmatrix}.$$

(c) From part (a),

$$\boldsymbol{S}\begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 2 & -1 \\ 5 & -3 \\ -3 & 2 \end{bmatrix}\begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \qquad \text{and} \qquad \boldsymbol{S}\begin{bmatrix} 3 \\ 5 \end{bmatrix} = \begin{bmatrix} 2 & -1 \\ 5 & -3 \\ -3 & 2 \end{bmatrix}\begin{bmatrix} 3 \\ 5 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix},$$

These vectors are easy to express in the basis $\mathcal{C}$:

$$\left[\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}\right]_{\mathcal{C}} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \qquad \text{and} \qquad \left[\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}\right]_{\mathcal{C}} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix},$$

so

$$[\boldsymbol{S}]_{\mathcal{B},\mathcal{C}} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}.$$

(d) By the same method as part (a), $[\boldsymbol{T}]_{\mathcal{C},\mathcal{E}} = \begin{bmatrix} 0 & 2 & 0 \\ 2 & 2 & 2 \end{bmatrix}$.

(e) By the same method as part (b), $[\boldsymbol{T}]_{\mathcal{E},\mathcal{B}} = \begin{bmatrix} 2 & -8 & 8 \\ -1 & 5 & -5 \end{bmatrix}$.

(f) By the same method as part (c), $[\boldsymbol{T}]_{\mathcal{C},\mathcal{B}} = \begin{bmatrix} -6 & 10 & -3 \\ 4 & -6 & 0 \end{bmatrix}$.

3.5.7. (a) i. $\begin{bmatrix} 1 \\ 2 \end{bmatrix} = \frac{3}{2}\begin{bmatrix} 1 \\ 1 \end{bmatrix} - \frac{1}{2}\begin{bmatrix} 1 \\ -1 \end{bmatrix}$, so $\left[\begin{bmatrix} 1 \\ 2 \end{bmatrix}\right]_{\mathcal{B}} = \frac{1}{2}\begin{bmatrix} 3 \\ -1 \end{bmatrix}$.

ii. $\begin{bmatrix} -2 \\ 3 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 1 \\ 1 \end{bmatrix} - \frac{5}{2}\begin{bmatrix} 1 \\ -1 \end{bmatrix}$, so $\left[\begin{bmatrix} -2 \\ 3 \end{bmatrix}\right]_{\mathcal{B}} = \frac{1}{2}\begin{bmatrix} 1 \\ -5 \end{bmatrix}$.

iii. $\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + 2\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$, so $\left[\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}\right]_{\mathcal{C}} = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}$.

iv. $\begin{bmatrix}3\\-2\\1\end{bmatrix} = 3\begin{bmatrix}1\\0\\1\end{bmatrix} - 2\begin{bmatrix}0\\1\\1\end{bmatrix}$, so $\left[\begin{bmatrix}3\\-2\\1\end{bmatrix}\right]_{\mathcal{C}} = \begin{bmatrix}0\\3\\-2\end{bmatrix}$.

v. We compute

$$\boldsymbol{T}\begin{bmatrix}1\\1\\0\end{bmatrix} = \begin{bmatrix}0\\1\end{bmatrix}, \qquad \boldsymbol{T}\begin{bmatrix}1\\0\\1\end{bmatrix} = \begin{bmatrix}1\\0\end{bmatrix}, \qquad \text{and} \qquad \boldsymbol{T}\begin{bmatrix}0\\1\\1\end{bmatrix} = \begin{bmatrix}1\\1\end{bmatrix},$$

so

$$[\boldsymbol{T}]_{\mathcal{C},\mathcal{E}} = \begin{bmatrix}0 & 1 & 1\\1 & 0 & 1\end{bmatrix}.$$

vi. We compute

$$\boldsymbol{T}\mathbf{e}_1 = \begin{bmatrix}0\\0\end{bmatrix},$$

$$\boldsymbol{T}\mathbf{e}_2 = \begin{bmatrix}0\\1\end{bmatrix} = \frac{1}{2}\begin{bmatrix}1\\1\end{bmatrix} - \frac{1}{2}\begin{bmatrix}1\\-1\end{bmatrix},$$

and

$$\boldsymbol{T}\mathbf{e}_3 = \begin{bmatrix}1\\0\end{bmatrix} = \frac{1}{2}\begin{bmatrix}1\\1\end{bmatrix} + \frac{1}{2}\begin{bmatrix}1\\-1\end{bmatrix},$$

so

$$[\boldsymbol{T}]_{\mathcal{E},\mathcal{B}} = \begin{bmatrix}0 & \frac{1}{2} & \frac{1}{2}\\0 & -\frac{1}{2} & \frac{1}{2}\end{bmatrix}.$$

vii. We compute

$$\boldsymbol{T}\begin{bmatrix}1\\1\\0\end{bmatrix} = \begin{bmatrix}0\\1\end{bmatrix} = \frac{1}{2}\begin{bmatrix}1\\1\end{bmatrix} - \frac{1}{2}\begin{bmatrix}1\\-1\end{bmatrix},$$

$$\boldsymbol{T}\begin{bmatrix}1\\0\\1\end{bmatrix} = \begin{bmatrix}1\\0\end{bmatrix} = \frac{1}{2}\begin{bmatrix}1\\1\end{bmatrix} + \frac{1}{2}\begin{bmatrix}1\\-1\end{bmatrix},$$

and

$$\boldsymbol{T}\begin{bmatrix}0\\1\\1\end{bmatrix} = \begin{bmatrix}1\\1\end{bmatrix},$$

so

$$[\boldsymbol{T}]_{\mathcal{C},\mathcal{B}} = \begin{bmatrix}\frac{1}{2} & \frac{1}{2} & 1\\-\frac{1}{2} & \frac{1}{2} & 0\end{bmatrix}.$$

(b) i. First method: $\left[\boldsymbol{T}\begin{bmatrix}1\\2\\3\end{bmatrix}\right]_{\mathcal{E}} = \left[\begin{bmatrix}3\\2\end{bmatrix}\right]_{\mathcal{E}} = \begin{bmatrix}3\\2\end{bmatrix}$.

Second method:

$$\left[\boldsymbol{T}\begin{bmatrix}1\\2\\3\end{bmatrix}\right]_{\mathcal{E}} = [\boldsymbol{T}]_{\mathcal{C},\mathcal{E}}\left[\begin{bmatrix}1\\2\\3\end{bmatrix}\right]_{\mathcal{C}} = \begin{bmatrix}0 & 1 & 1\\1 & 0 & 1\end{bmatrix}\begin{bmatrix}0\\1\\2\end{bmatrix} = \begin{bmatrix}3\\2\end{bmatrix}.$$

ii. First method:

$$\boldsymbol{T}\begin{bmatrix}3\\-2\\1\end{bmatrix}=\begin{bmatrix}1\\-2\end{bmatrix}=-\frac{1}{2}\begin{bmatrix}1\\1\end{bmatrix}+\frac{3}{2}\begin{bmatrix}1\\-1\end{bmatrix},$$

so $\left[\boldsymbol{T}\begin{bmatrix}3\\-2\\1\end{bmatrix}\right]_{\mathcal{B}}=\frac{1}{2}\begin{bmatrix}-1\\3\end{bmatrix}$.

Second method:

$$\left[\boldsymbol{T}\begin{bmatrix}3\\-2\\1\end{bmatrix}\right]_{\mathcal{B}}=[\boldsymbol{T}]_{\mathcal{C},\mathcal{B}}\left[\begin{bmatrix}3\\-2\\1\end{bmatrix}\right]_{\mathcal{C}}=\begin{bmatrix}\frac{1}{2}&\frac{1}{2}&1\\-\frac{1}{2}&\frac{1}{2}&0\end{bmatrix}\begin{bmatrix}0\\3\\-2\end{bmatrix}=\frac{1}{2}\begin{bmatrix}-1\\3\end{bmatrix}.$$

3.5.8. (a) i. $\begin{bmatrix}-2\\1\end{bmatrix}=-\frac{1}{2}\begin{bmatrix}1\\1\end{bmatrix}-\frac{3}{2}\begin{bmatrix}1\\-1\end{bmatrix}$, so $\left[\begin{bmatrix}-2\\1\end{bmatrix}\right]_{\mathcal{B}}=\frac{1}{2}\begin{bmatrix}-1\\-3\end{bmatrix}$.

ii. $\begin{bmatrix}2\\3\end{bmatrix}=\frac{5}{2}\begin{bmatrix}1\\1\end{bmatrix}-\frac{1}{2}\begin{bmatrix}1\\-1\end{bmatrix}$, so $\left[\begin{bmatrix}2\\3\end{bmatrix}\right]_{\mathcal{B}}=\frac{1}{2}\begin{bmatrix}5\\-1\end{bmatrix}$.

iii. $\begin{bmatrix}1\\-2\\3\end{bmatrix}=-2\begin{bmatrix}1\\0\\1\end{bmatrix}+3\begin{bmatrix}1\\0\\1\end{bmatrix}$, so $\left[\begin{bmatrix}1\\2\\3\end{bmatrix}\right]_{\mathcal{C}}=\begin{bmatrix}-2\\3\\0\end{bmatrix}$.

iv. $\begin{bmatrix}3\\2\\1\end{bmatrix}=2\begin{bmatrix}1\\0\\1\end{bmatrix}+\begin{bmatrix}0\\1\\1\end{bmatrix}$, so $\left[\begin{bmatrix}3\\-2\\1\end{bmatrix}\right]_{\mathcal{C}}=\begin{bmatrix}2\\1\\0\end{bmatrix}$.

v. We compute

$$\boldsymbol{T}\begin{bmatrix}1\\1\\0\end{bmatrix}=\begin{bmatrix}1\\1\end{bmatrix},\qquad \boldsymbol{T}\begin{bmatrix}1\\0\\1\end{bmatrix}=\begin{bmatrix}0\\1\end{bmatrix},\qquad\text{and}\qquad \boldsymbol{T}\begin{bmatrix}0\\1\\1\end{bmatrix}=\begin{bmatrix}1\\0\end{bmatrix},$$

so

$$[\boldsymbol{T}]_{\mathcal{C},\mathcal{E}}=\begin{bmatrix}1&0&1\\1&1&0\end{bmatrix}.$$

vi. We compute

$$\boldsymbol{T}\mathbf{e}_1=\begin{bmatrix}1\\1\end{bmatrix},$$

$$\boldsymbol{T}\mathbf{e}_2=\begin{bmatrix}1\\0\end{bmatrix}=\frac{1}{2}\begin{bmatrix}1\\1\end{bmatrix}+\frac{1}{2}\begin{bmatrix}1\\-1\end{bmatrix},$$

and

$$\boldsymbol{T}\mathbf{e}_3=\begin{bmatrix}0\\0\end{bmatrix},$$

so

$$[\boldsymbol{T}]_{\mathcal{E},\mathcal{B}}=\begin{bmatrix}1&\frac{1}{2}&0\\0&\frac{1}{2}&0\end{bmatrix}.$$

vii. We compute

$$\boldsymbol{T}\begin{bmatrix}1\\1\\0\end{bmatrix}=\begin{bmatrix}1\\1\end{bmatrix},$$

$$\boldsymbol{T}\begin{bmatrix}1\\0\\1\end{bmatrix}=\begin{bmatrix}0\\1\end{bmatrix}=\frac{1}{2}\begin{bmatrix}1\\1\end{bmatrix}-\frac{1}{2}\begin{bmatrix}1\\-1\end{bmatrix},$$

and

$$\boldsymbol{T}\begin{bmatrix}0\\1\\1\end{bmatrix}=\begin{bmatrix}1\\0\end{bmatrix}=\frac{1}{2}\begin{bmatrix}1\\1\end{bmatrix}+\frac{1}{2}\begin{bmatrix}1\\-1\end{bmatrix},$$

so

$$[\boldsymbol{T}]_{\mathcal{C},\mathcal{B}}=\begin{bmatrix}1&\frac{1}{2}&\frac{1}{2}\\0&-\frac{1}{2}&\frac{1}{2}\end{bmatrix}.$$

(b) i. First method: $\left[\boldsymbol{T}\begin{bmatrix}1\\-2\\3\end{bmatrix}\right]_{\mathcal{E}}=\left[\begin{bmatrix}-2\\1\end{bmatrix}\right]_{\mathcal{E}}=\begin{bmatrix}-2\\1\end{bmatrix}.$

Second method:

$$\left[\boldsymbol{T}\begin{bmatrix}1\\-2\\3\end{bmatrix}\right]_{\mathcal{E}}=[\boldsymbol{T}]_{\mathcal{C},\mathcal{E}}\left[\begin{bmatrix}1\\-2\\3\end{bmatrix}\right]_{\mathcal{C}}=\begin{bmatrix}1&0&1\\1&1&0\end{bmatrix}\begin{bmatrix}-2\\3\\0\end{bmatrix}=\begin{bmatrix}-2\\1\end{bmatrix}.$$

ii. First method:

$$\left[\boldsymbol{T}\begin{bmatrix}3\\2\\1\end{bmatrix}\right]_{\mathcal{B}}=\left[\begin{bmatrix}2\\3\end{bmatrix}\right]_{\mathcal{B}}=\frac{1}{2}\begin{bmatrix}5\\-1\end{bmatrix}.$$

Second method:

$$\left[\boldsymbol{T}\begin{bmatrix}3\\2\\1\end{bmatrix}\right]_{\mathcal{B}}=[\boldsymbol{T}]_{\mathcal{C},\mathcal{B}}\left[\begin{bmatrix}3\\2\\1\end{bmatrix}\right]_{\mathcal{C}}=\begin{bmatrix}1&\frac{1}{2}&\frac{1}{2}\\0&-\frac{1}{2}&\frac{1}{2}\end{bmatrix}\begin{bmatrix}2\\1\\0\end{bmatrix}=\frac{1}{2}\begin{bmatrix}5\\-1\end{bmatrix}.$$

3.5.9. (a) The plane P is the set of solutions to the 1×3 linear system $x-2y+3z=0$, which has pivot variable x and free variables y,z. We can therefore right points in P as

$$\begin{bmatrix}x\\y\\z\end{bmatrix}=\begin{bmatrix}2y-3z\\y\\z\end{bmatrix}=y\begin{bmatrix}2\\1\\0\end{bmatrix}+z\begin{bmatrix}-3\\0\\1\end{bmatrix},$$

so $\mathcal{B}=\left(\begin{bmatrix}2\\1\\0\end{bmatrix},\begin{bmatrix}-3\\0\\1\end{bmatrix}\right)$ is a basis for P.

(b) i. This vector lies in P. The linear system

$$a\begin{bmatrix}2\\1\\0\end{bmatrix}+b\begin{bmatrix}-3\\0\\1\end{bmatrix}=\begin{bmatrix}1\\-1\\-1\end{bmatrix}$$

has solution $a=b=-1$, so $\left[\begin{bmatrix}1\\-1\\-1\end{bmatrix}\right]_{\mathcal{B}}=\begin{bmatrix}-1\\-1\end{bmatrix}$.

ii. This vector is not in P.

iii. This vector lies in P. The linear system

$$a\begin{bmatrix}2\\1\\0\end{bmatrix}+b\begin{bmatrix}-3\\0\\1\end{bmatrix}=\begin{bmatrix}5\\-2\\-3\end{bmatrix}$$

has solution $a=-2$, $b=-3$, so $\left[\begin{bmatrix}1\\-1\\-1\end{bmatrix}\right]_{\mathcal{B}}=\begin{bmatrix}-2\\-3\end{bmatrix}$.

3.5.10. (a) The plane P is the set of solutions to the 1×3 linear system $4x+y-2z=0$. Solving slightly unconventionally for y (because it avoids fractions) gives that $y=-4x+2z$, and so the plane can also be written

$$P=\left\{x\begin{bmatrix}1\\-4\\0\end{bmatrix}+z\begin{bmatrix}0\\2\\1\end{bmatrix}\;\middle|\;x,z\in\mathbb{R}\right\},$$

from which we can see that $\left(\begin{bmatrix}1\\-4\\0\end{bmatrix},\begin{bmatrix}0\\2\\1\end{bmatrix}\right)$ is a basis for P.

(b) i. Not in P: $4(1)+(1)-2(1)=3\neq 0$.

ii. Yes, this is in P: $4(0)+(2)-2(1)=0$. Indeed, this is one of basis vectors, and so its coordinates with respect to the basis are $\begin{bmatrix}0\\1\end{bmatrix}$.

iii. Yes, this is in P: $4(1)+(-2)-2(1)=0$. Finding the coordinates in our basis requires solving the linear system

$$\left[\begin{array}{cc|c}1&0&1\\-4&2&-2\\0&1&1\end{array}\right]\rightsquigarrow\left[\begin{array}{cc|c}1&0&1\\0&2&2\\0&1&1\end{array}\right]\rightsquigarrow\left[\begin{array}{cc|c}1&0&1\\0&1&1\\0&0&0\end{array}\right],$$

from which we deduce that $\left[\begin{bmatrix}1\\-2\\1\end{bmatrix}\right]_{\mathcal{B}}=\begin{bmatrix}1\\1\end{bmatrix}$.

3.5.11. Since $\boldsymbol{T}x^j = x^{j+1}$, the jth column of $[\boldsymbol{T}]$ is $\mathbf{e}_{j+1}$:

$$[\boldsymbol{T}] = \begin{bmatrix} 0 & 0 & & 0 \\ 1 & 0 & & 0 \\ 0 & 1 & & 0 \\ \vdots & & \ddots & \\ 0 & 0 & & 1 \end{bmatrix} \in \mathrm{M}_{n+1,n}(\mathbb{R}).$$

The matrix representation of $[\boldsymbol{D}]$ was found in Example 2 on pages 188–189:

$$[\boldsymbol{D}] = \begin{bmatrix} 0 & 1 & 0 & & 0 \\ 0 & 0 & 2 & & 0 \\ & & & \ddots & \\ 0 & 0 & 0 & & n \end{bmatrix}.$$

We can then compute $[\boldsymbol{DT}]$ and $[\boldsymbol{TD}]$ by matrix multiplication:

$$[\boldsymbol{DT}] = [\boldsymbol{D}][\boldsymbol{T}] = \begin{bmatrix} 1 & 0 & 0 & & 0 \\ 0 & 2 & 0 & & 0 \\ 0 & 0 & 3 & & 0 \\ \vdots & & & \ddots & \\ 0 & 0 & & & n \end{bmatrix} \in \mathrm{M}_n(\mathbb{R})$$

and

$$[\boldsymbol{TD}] = [\boldsymbol{T}][\boldsymbol{D}] = \begin{bmatrix} 0 & 0 & 0 & & 0 \\ 0 & 1 & 0 & & 0 \\ 0 & 0 & 2 & & 0 \\ \vdots & & & \ddots & \\ 0 & 0 & & & n \end{bmatrix} \in \mathrm{M}_{n+1}(\mathbb{R}).$$

3.5.12. (a) We know that $\mathcal{P}_2(\mathbb{R})$ is 3-dimensional, so it suffices to show that the elements of $\mathcal{B}$ are linearly independent. This is clear by the linear dependence lemma: $1 \neq 0$, $x \notin \langle 1 \rangle$, and $\frac{3}{2}x^2 - \frac{1}{2} \notin \langle 1, x \rangle$.

(b) We must write x^2 as a linear combination of the basis elements: $x^2 = a\left(\frac{3}{2}x^2 - \frac{1}{2}\right) + bx + c$. By what is essentially back-substitution, we can see first that $a = \frac{2}{3}$, then $b = 0$, and finally $c = \frac{1}{3}$. The $\mathcal{B}$-coordinates of x^2 are thus $\left(\frac{1}{3}, 0, \frac{2}{3}\right)$.

(c) For each basis element $p(x)$, we need to express $\boldsymbol{D}p(x)$ in $\mathcal{B}$-coordinates. We have

$$\boldsymbol{D}1 = 0 \qquad \boldsymbol{D}x = 1 \qquad \boldsymbol{D}\left(\frac{3}{2}x^2 - \frac{1}{2}\right) = 3x,$$

and so

$$[\boldsymbol{D}1]_{\mathcal{B}} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \qquad [\boldsymbol{D}x]_{\mathcal{B}} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \qquad \left[\boldsymbol{D}\left(\frac{3}{2}x^2 - \frac{1}{2}\right)\right]_{\mathcal{B}} = \begin{bmatrix} 0 \\ 3 \\ 0 \end{bmatrix},$$

and thus

$$[\boldsymbol{D}]_{\mathcal{B}} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 3 \\ 0 & 0 & 0 \end{bmatrix}.$$

(d)

$$\left[\frac{d}{dx}(x^2)\right]_{\mathcal{B}} = [\boldsymbol{D}]_{\mathcal{B}} \left[x^2\right]_{\mathcal{B}} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 3 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{3} \\ 0 \\ \frac{2}{3} \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix},$$

and so $\frac{d}{dx}(x^2) = 0 \cdot 1 + 2 \cdot x + 0\left(\frac{3}{2}x^2 - \frac{1}{2}\right) = 2x$, as expected.

3.5.13. Let $\mathcal{B} = (v_1, \ldots, v_n)$. By Proposition 3.45, v_j is an eigenvector of $\boldsymbol{T}$ with eigenvalue λ_j. Since $\boldsymbol{T}$ is invertible, each $\lambda_j \neq 0$. Then

$$\boldsymbol{T}^{-1}(v_j) = \boldsymbol{T}^{-1}(\lambda_j^{-1}\lambda_j v_j) = \lambda_j^{-1}\boldsymbol{T}^{-1}(\lambda_j v_j) = \lambda_j^{-1}\boldsymbol{T}^{-1}(\boldsymbol{T}v_j) = \lambda_j^{-1}v_j,$$

and so $\left[\boldsymbol{T}^{-1}\right]_{\mathcal{B}} = \mathbf{diag}(\lambda_1^{-1}, \ldots, \lambda_n^{-1})$.

3.5.14. (a) For any nonzero $\mathbf{x} \in \mathbb{R}^2$, $\boldsymbol{R}(\mathbf{x})$ points in a different direction from $\mathbf{x}$, so $\mathbf{x}$ cannot be an eigenvector of $\boldsymbol{R}$. By Proposition 3.45, $\boldsymbol{R}$ is not diagonalizable.

(b) $\boldsymbol{R}^2$ is rotation by π radians, so $\boldsymbol{R}^2 = -\boldsymbol{I}$. It follows that the matrix of $\boldsymbol{R}^2$ with respect to any basis of $\mathbb{R}^2$ is $-\mathbf{I}_2$.

3.5.15. $\boldsymbol{P}(\mathbf{e}_1) = \mathbf{e_1}$, $\boldsymbol{P}(\mathbf{e}_2) = \mathbf{e_2}$, and $\boldsymbol{P}(\mathbf{e}_3) = \mathbf{0}$, so $[\boldsymbol{P}]_{\mathcal{E}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$, which is diagonal.

3.5.16. Let $\mathbf{v}$ be a nonzero vector lying on L, and let $\mathbf{w}$ be any nonzero vector pointing in the perpendicular direction. By Exercise 3.1.7, $\mathbf{v}$ and $\mathbf{w}$ are linearly independent, and so by Theorem 3.28, $\mathcal{B} = (\mathbf{v}, \mathbf{w})$ is a basis for $\mathbb{R}^2$. Now $\boldsymbol{P}\mathbf{v} = \mathbf{v}$ and $\boldsymbol{P}\mathbf{w} = \mathbf{0} = 0\mathbf{w}$, so $[\boldsymbol{P}]_{\mathcal{B}} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$, which is diagonal.

3.5.17. We can write $\mathbf{u} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n$ for some $c_1, \ldots, c_n \in \mathbb{F}$. Then $[\mathbf{u}]_{\mathcal{B}} = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}$, and so

$\mathbf{A} \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n = \mathbf{u}$ by the definition of matrix-vector multiplication.

3.5.18. Firstly, since $\mathcal{B}$ is a basis for $\mathbb{F}^n$, we know that the matrix $\mathbf{A} = \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{bmatrix}$ is invertible, so $\mathbf{A}^{-1}$ does exist. In particular, showing that for any $\mathbf{x} \in \mathbb{F}^n$, $[\mathbf{x}]_{\mathcal{B}} = \mathbf{A}^{-1}\mathbf{x}$ is equivalent to showing that for any $\mathbf{x} \in \mathbb{F}^n$, $\mathbf{A}\,[\mathbf{x}]_{\mathcal{B}} = \mathbf{x}$.

Now, let $[\mathbf{x}]_{\mathcal{B}} = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}$, which is the same thing as saying that

$$\mathbf{x} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n.$$

But also,

$$\mathbf{A}\begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} = \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{bmatrix}\begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n,$$

and so indeed $\mathbf{A}\,[\mathbf{x}]_{\mathcal{B}} = \mathbf{x}$ for all $\mathbf{x} \in \mathbb{F}^n$.

3.5.19. Suppose that $\boldsymbol{S}$ and $\boldsymbol{T}$ are both diagonalized by $\mathcal{B}$, say $[\boldsymbol{S}]_{\mathcal{B}} = \mathbf{diag}(\lambda_1, \ldots, \lambda_n)$ and $[\boldsymbol{T}]_{\mathcal{B}} = \mathbf{diag}(\mu_1, \ldots, \mu_n)$. Then $[\boldsymbol{ST}]_{\mathcal{B}} = [\boldsymbol{S}]_{\mathcal{B}}\,[\boldsymbol{T}]_{\mathcal{B}} = \mathbf{diag}(\lambda_1\mu_1, \ldots, \lambda_n\mu_n)$, so $\boldsymbol{ST}$ is also diagonalized by $\mathcal{B}$.

3.5.20. By Theorem 3.47, $[\boldsymbol{I}]_{\mathcal{B}_1,\mathcal{B}_2}\,[\boldsymbol{I}]_{\mathcal{B}_2,\mathcal{B}_1} = [\boldsymbol{I}]_{\mathcal{B}_2} = \mathbf{I}_n$ (where $n = \dim V$), so by Corollary 3.37, $[\boldsymbol{I}]_{\mathcal{B}_1,\mathcal{B}_2}$ is invertible and $[\boldsymbol{I}]^{-1}_{\mathcal{B}_1,\mathcal{B}_2} = [\boldsymbol{I}]_{\mathcal{B}_2,\mathcal{B}_1}$.

3.6 Change of basis

3.6.1. (a) By Lemma 3.50,

$$[\boldsymbol{I}]_{\mathcal{B},\mathcal{E}} = \begin{bmatrix} 2 & 3 \\ 5 & 7 \end{bmatrix} \qquad \text{and} \qquad [\boldsymbol{I}]_{\mathcal{C},\mathcal{E}} = \begin{bmatrix} -1 & 1 \\ 2 & -4 \end{bmatrix},$$

and so

$$[\boldsymbol{I}]_{\mathcal{E},\mathcal{B}} = [\boldsymbol{I}]^{-1}_{\mathcal{B},\mathcal{E}} = \begin{bmatrix} -7 & 3 \\ 5 & -2 \end{bmatrix} \qquad \text{and} \qquad [\boldsymbol{I}]_{\mathcal{E},\mathcal{C}} = [\boldsymbol{I}]^{-1}_{\mathcal{C},\mathcal{E}} = \frac{1}{2}\begin{bmatrix} -4 & -1 \\ -2 & -1 \end{bmatrix}.$$

From this, we can compute

$$[\boldsymbol{I}]_{\mathcal{B},\mathcal{C}} = [\boldsymbol{I}]_{\mathcal{E},\mathcal{C}}\,[\boldsymbol{I}]_{\mathcal{B},\mathcal{E}} = \frac{1}{2}\begin{bmatrix} -4 & -1 \\ -2 & -1 \end{bmatrix}\begin{bmatrix} 2 & 3 \\ 5 & 7 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} -13 & -19 \\ -9 & -13 \end{bmatrix}$$

and

$$[\boldsymbol{I}]_{\mathcal{C},\mathcal{B}} = [\boldsymbol{I}]_{\mathcal{E},\mathcal{B}}\,[\boldsymbol{I}]_{\mathcal{C},\mathcal{E}} = \begin{bmatrix} -7 & 3 \\ 5 & -2 \end{bmatrix}\begin{bmatrix} -1 & 1 \\ 2 & -4 \end{bmatrix} = \begin{bmatrix} 13 & -19 \\ -9 & 13 \end{bmatrix}.$$

(b) By the same method, $[\boldsymbol{I}]_{\mathcal{B},\mathcal{C}} = \frac{1}{2}\begin{bmatrix} 0 & -4 \\ 1 & 5 \end{bmatrix}$ and $[\boldsymbol{I}]_{\mathcal{C},\mathcal{B}} = \frac{1}{2}\begin{bmatrix} 5 & 4 \\ -1 & 0 \end{bmatrix}$.

(c) By the same method, $[\boldsymbol{I}]_{\mathcal{B},\mathcal{C}} = \begin{bmatrix} 1 & 2 & 0 \\ -2 & -3 & -1 \\ -4 & -6 & -1 \end{bmatrix}$ and $[\boldsymbol{I}]_{\mathcal{C},\mathcal{B}} = \begin{bmatrix} -3 & 2 & -2 \\ 2 & -1 & 1 \\ 0 & -2 & 1 \end{bmatrix}$.

(d) By the same method, $[\boldsymbol{I}]_{\mathcal{B},\mathcal{C}} = \begin{bmatrix} 2 & 0 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix}$ and $[\boldsymbol{I}]_{\mathcal{C},\mathcal{B}} = \frac{1}{2}\begin{bmatrix} 1 & 0 & 0 \\ 1 & 2 & 2 \\ 1 & 2 & 4 \end{bmatrix}$.

3.6.2. (a) First,

$$[\boldsymbol{I}]_{\mathcal{B},\mathcal{E}} = \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix} \qquad \text{and} \qquad [\boldsymbol{I}]_{\mathcal{C},\mathcal{E}} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix},$$

and so

$$[\boldsymbol{I}]_{\mathcal{E},\mathcal{B}} = [\boldsymbol{I}]_{\mathcal{B},\mathcal{E}}^{-1} = \begin{bmatrix} -3 & 2 \\ 2 & -1 \end{bmatrix} \qquad \text{and} \qquad [\boldsymbol{I}]_{\mathcal{E},\mathcal{C}} = [\boldsymbol{I}]_{\mathcal{C},\mathcal{E}}^{-1} = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}.$$

From this, we can compute

$$[\boldsymbol{I}]_{\mathcal{B},\mathcal{C}} = [\boldsymbol{I}]_{\mathcal{E},\mathcal{C}}\,[\boldsymbol{I}]_{\mathcal{B},\mathcal{E}} = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 3 & 5 \\ -1 & -1 \end{bmatrix}$$

and

$$[\boldsymbol{I}]_{\mathcal{C},\mathcal{B}} = [\boldsymbol{I}]_{\mathcal{E},\mathcal{B}}\,[\boldsymbol{I}]_{\mathcal{C},\mathcal{E}} = \begin{bmatrix} -3 & 2 \\ 2 & -1 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} -1 & -5 \\ 1 & 3 \end{bmatrix}.$$

(b) Directly this time, we can express the elements of $\mathcal{B}$ in terms of $\mathcal{C}$:

$$\left[\begin{array}{cc|cc} 1 & 0 & 2 & 3 \\ 2 & 1 & 3 & 5 \end{array}\right] \rightsquigarrow \left[\begin{array}{cc|cc} 1 & 0 & 2 & 3 \\ 0 & 1 & -1 & -1 \end{array}\right],$$

and so

$$\left[\begin{bmatrix} 2 \\ 3 \end{bmatrix}\right]_{\mathcal{C}} = \begin{bmatrix} 2 \\ -1 \end{bmatrix} \qquad \text{and} \qquad \left[\begin{bmatrix} 3 \\ 5 \end{bmatrix}\right]_{\mathcal{C}} = \begin{bmatrix} 3 \\ -1 \end{bmatrix},$$

which means $[\boldsymbol{I}]_{\mathcal{B},\mathcal{C}} = \begin{bmatrix} 2 & 3 \\ -1 & -1 \end{bmatrix}$. Finally, $[\boldsymbol{I}]_{\mathcal{C},\mathcal{B}} = [\boldsymbol{I}]_{\mathcal{B},\mathcal{C}}^{-1} = \begin{bmatrix} -1 & -3 \\ 1 & 2 \end{bmatrix}$.

(c) By either of the above methods, $[\boldsymbol{I}]_{\mathcal{B},\mathcal{C}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{3} & \frac{2}{3} \\ 0 & \frac{2}{3} & \frac{1}{3} \end{bmatrix}$ and $[\boldsymbol{I}]_{\mathcal{C},\mathcal{B}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{3} & \frac{2}{3} \\ 0 & \frac{2}{3} & \frac{1}{3} \end{bmatrix}$.

(d) By either of the above methods, $[\boldsymbol{I}]_{\mathcal{B},\mathcal{C}} = \begin{bmatrix} 1 & 2 & 3 \\ -2 & -3 & -4 \\ 1 & 0 & 0 \end{bmatrix}$ and $[\boldsymbol{I}]_{\mathcal{C},\mathcal{B}} = \begin{bmatrix} 0 & 0 & 1 \\ -4 & -3 & -2 \\ 3 & 2 & 1 \end{bmatrix}$.

3.6.3. (a) By Lemma 3.50, $[\boldsymbol{I}]_{\mathcal{B},\mathcal{E}} = \begin{bmatrix} 2 & -1 \\ -3 & 2 \end{bmatrix}$, and then $[\boldsymbol{I}]_{\mathcal{E},\mathcal{B}} = \begin{bmatrix} 2 & -1 \\ -3 & 2 \end{bmatrix}^{-1} = \begin{bmatrix} 2 & 1 \\ 3 & 2 \end{bmatrix}$.

(b) Multiply each of the given vectors by $[\boldsymbol{I}]_{\mathcal{E},\mathcal{B}} = \begin{bmatrix} 2 & 1 \\ 3 & 2 \end{bmatrix}$ to obtain:

(i) $\begin{bmatrix} 2 \\ 3 \end{bmatrix}$ (ii) $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$ (iii) $\begin{bmatrix} 7 \\ 12 \end{bmatrix}$ (iv) $\begin{bmatrix} 3 \\ 2 \end{bmatrix}$ (v) $\begin{bmatrix} -11 \\ -16 \end{bmatrix}$

3.6.4. (a) By Lemma 3.50, $[\boldsymbol{I}]_{\mathcal{B},\mathcal{E}} = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}$, and then $[\boldsymbol{I}]_{\mathcal{E},\mathcal{B}} = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}^{-1} = \frac{1}{2}\begin{bmatrix} -4 & 3 \\ 2 & -1 \end{bmatrix}$.

(b) Multiply each of the given vectors by $[\boldsymbol{I}]_{\mathcal{E},\mathcal{B}} = \frac{1}{2}\begin{bmatrix} -4 & 3 \\ 2 & -1 \end{bmatrix}$ to obtain:

(i) $\begin{bmatrix} -2 \\ 1 \end{bmatrix}$ (ii) $\frac{1}{2}\begin{bmatrix} 3 \\ -1 \end{bmatrix}$ (iii) $\frac{1}{2}\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ (iv) $\frac{1}{2}\begin{bmatrix} -31 \\ 13 \end{bmatrix}$ (v) $\frac{1}{2}\begin{bmatrix} 27 \\ -13 \end{bmatrix}$

3.6.5. (a) By Lemma 3.50, $[\boldsymbol{I}]_{\mathcal{B},\mathcal{E}} = \begin{bmatrix} 1 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & 1 & -1 \end{bmatrix}$, and then $[\boldsymbol{I}]_{\mathcal{E},\mathcal{B}} = \begin{bmatrix} 1 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & 1 & -1 \end{bmatrix}^{-1} = \frac{1}{2}\begin{bmatrix} 1 & -1 & -1 \\ 1 & 1 & 1 \\ 1 & 1 & -1 \end{bmatrix}$.

(b) Multiply each of the given vectors by $[\boldsymbol{I}]_{\mathcal{E},\mathcal{B}} = \frac{1}{2}\begin{bmatrix} 1 & -1 & -1 \\ 1 & 1 & 1 \\ 1 & 1 & -1 \end{bmatrix}$ to obtain:

(i) $\frac{1}{2}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ (ii) $\frac{1}{2}\begin{bmatrix} -1 \\ 1 \\ -1 \end{bmatrix}$ (iii) $\begin{bmatrix} -3 \\ -1 \\ -3 \end{bmatrix}$ (iv) $\begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix}$ (v) $\frac{1}{2}\begin{bmatrix} 9 \\ 3 \\ 1 \end{bmatrix}$

3.6.6. (a) By Lemma 3.50, $[\boldsymbol{I}]_{\mathcal{B},\mathcal{E}} = \begin{bmatrix} 1 & 1 & 0 \\ -1 & 2 & 2 \\ 1 & -1 & -1 \end{bmatrix}$, and then $[\boldsymbol{I}]_{\mathcal{E},\mathcal{B}} = \begin{bmatrix} 1 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & 1 & -1 \end{bmatrix}^{-1} = \begin{bmatrix} 0 & 1 & 2 \\ 1 & -1 & -2 \\ -1 & 2 & 3 \end{bmatrix}$.

(b) Multiply each of the given vectors by $[\boldsymbol{I}]_{\mathcal{E},\mathcal{B}} = \begin{bmatrix} 0 & 1 & 2 \\ 1 & -1 & -2 \\ -1 & 2 & 3 \end{bmatrix}$ to obtain:

(i) $\begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}$ (ii) $\begin{bmatrix} 2 \\ -2 \\ 3 \end{bmatrix}$ (iii) $\begin{bmatrix} 4 \\ -8 \\ 10 \end{bmatrix}$ (iv) $\begin{bmatrix} -4 \\ 5 \\ -6 \end{bmatrix}$ (v) $\begin{bmatrix} -2 \\ 8 \\ -11 \end{bmatrix}$

3.6.7. (a) By Lemma 3.50,

$$[\boldsymbol{I}]_{\mathcal{B},\mathcal{E}} = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \qquad \text{and} \qquad [\boldsymbol{I}]_{\mathcal{C},\mathcal{E}} = \begin{bmatrix} 1 & 0 & 1 \\ 3 & 2 & 0 \\ 0 & 3 & -4 \end{bmatrix},$$

and then

$$[\boldsymbol{I}]_{\mathcal{E},\mathcal{B}} = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \qquad \text{and} \qquad [\boldsymbol{I}]_{\mathcal{E},\mathcal{C}} = \begin{bmatrix} 1 & 0 & 1 \\ 3 & 2 & 0 \\ 0 & 3 & -4 \end{bmatrix}^{-1} = \begin{bmatrix} -8 & 3 & -2 \\ 12 & -4 & 3 \\ 9 & -3 & 2 \end{bmatrix}.$$

(b) i. $[\boldsymbol{S}]_{\mathcal{B},\mathcal{E}} = [\boldsymbol{S}]_{\mathcal{E},\mathcal{E}}\,[\boldsymbol{I}]_{\mathcal{B},\mathcal{E}} = \begin{bmatrix} 1 & 1 \\ 3 & 0 \\ 0 & -4 \end{bmatrix}$

ii. $[\boldsymbol{S}]_{\mathcal{E},\mathcal{C}} = [\boldsymbol{I}]_{\mathcal{E},\mathcal{C}}\,[\boldsymbol{S}]_{\mathcal{E},\mathcal{E}} = \begin{bmatrix} 1 & -1 \\ 0 & 0 \\ -1 & 2 \end{bmatrix}$

iii. $[\boldsymbol{S}]_{\mathcal{B},\mathcal{C}} = [\boldsymbol{I}]_{\mathcal{E},\mathcal{C}}\,[\boldsymbol{S}]_{\mathcal{E},\mathcal{E}}\,[\boldsymbol{I}]_{\mathcal{B},\mathcal{E}} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}$

iv. $[\boldsymbol{T}]_{\mathcal{C},\mathcal{E}} = [\boldsymbol{T}]_{\mathcal{E},\mathcal{E}}\,[\boldsymbol{I}]_{\mathcal{C},\mathcal{E}} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & -1 \end{bmatrix}$

v. $[\boldsymbol{T}]_{\mathcal{E},\mathcal{B}} = [\boldsymbol{I}]_{\mathcal{E},\mathcal{B}}\,[\boldsymbol{T}]_{\mathcal{E},\mathcal{E}} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & -1 & 1 \end{bmatrix}$

vi. $[\boldsymbol{T}]_{\mathcal{C},\mathcal{B}} = [\boldsymbol{I}]_{\mathcal{E},\mathcal{B}}\,[\boldsymbol{T}]_{\mathcal{E},\mathcal{E}}\,[\boldsymbol{I}]_{\mathcal{C},\mathcal{E}} = \begin{bmatrix} 1 & 0 & 1 \\ -1 & 1 & -2 \end{bmatrix}$

3.6.8. (a) By Lemma 3.50,

$$[\boldsymbol{I}]_{\mathcal{B},\mathcal{E}} = \begin{bmatrix} 2 & 3 \\ 3 & 5 \end{bmatrix} \qquad \text{and} \qquad [\boldsymbol{I}]_{\mathcal{C},\mathcal{E}} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix},$$

and then

$$[\boldsymbol{I}]_{\mathcal{E},\mathcal{B}} = \begin{bmatrix} 2 & 3 \\ 3 & 5 \end{bmatrix}^{-1} = \begin{bmatrix} 5 & -3 \\ -3 & 2 \end{bmatrix} \qquad \text{and} \qquad [\boldsymbol{I}]_{\mathcal{E},\mathcal{C}} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}^{-1} = \frac{1}{2}\begin{bmatrix} 1 & 1 & -1 \\ 1 & -1 & 1 \\ -1 & 1 & 1 \end{bmatrix}.$$

(b) i. $[\boldsymbol{S}]_{\mathcal{B},\mathcal{E}} = [\boldsymbol{S}]_{\mathcal{E},\mathcal{E}}\,[\boldsymbol{I}]_{\mathcal{B},\mathcal{E}} = \begin{bmatrix} 1 & 1 \\ 19 & 30 \\ 0 & 1 \end{bmatrix}$

ii. $[\boldsymbol{S}]_{\mathcal{E},\mathcal{C}} = [\boldsymbol{I}]_{\mathcal{E},\mathcal{C}}\,[\boldsymbol{S}]_{\mathcal{E},\mathcal{E}} = \begin{bmatrix} 5 & 0 \\ -3 & 1 \\ 0 & 3 \end{bmatrix}$

iii. $[\boldsymbol{S}]_{\mathcal{B},\mathcal{C}} = [\boldsymbol{I}]_{\mathcal{E},\mathcal{C}}\,[\boldsymbol{S}]_{\mathcal{E},\mathcal{E}}\,[\boldsymbol{I}]_{\mathcal{B},\mathcal{E}} = \begin{bmatrix} 10 & 15 \\ -9 & -14 \\ 9 & 15 \end{bmatrix}$

iv. $[\boldsymbol{T}]_{\mathcal{C},\mathcal{E}} = [\boldsymbol{T}]_{\mathcal{E},\mathcal{E}}\,[\boldsymbol{I}]_{\mathcal{C},\mathcal{E}} = \begin{bmatrix} 0 & 2 & 0 \\ 2 & 0 & 0 \end{bmatrix}$

v. $[\boldsymbol{T}]_{\mathcal{E},\mathcal{B}} = [\boldsymbol{I}]_{\mathcal{E},\mathcal{B}}\,[\boldsymbol{T}]_{\mathcal{E},\mathcal{E}} = \begin{bmatrix} 2 & -8 & 8 \\ -1 & 5 & -5 \end{bmatrix}$

vi. $[\boldsymbol{T}]_{\mathcal{C},\mathcal{B}} = [\boldsymbol{I}]_{\mathcal{E},\mathcal{B}}\,[\boldsymbol{T}]_{\mathcal{E},\mathcal{E}}\,[\boldsymbol{I}]_{\mathcal{C},\mathcal{E}} = \begin{bmatrix} -6 & 10 & 0 \\ 4 & -6 & 0 \end{bmatrix}$

3.6.9. (a) Since $\boldsymbol{T}(1,-1,0) = (1,0,-1)$ and $\boldsymbol{T}(1,0,-1) = (1,-1,0)$,

$$[\boldsymbol{T}]_{\mathcal{B}} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

(b) Writing the elements of $\mathcal{B}$ in terms of $\mathcal{C}$:

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

so

$$\left[\begin{bmatrix}1\\-1\\0\end{bmatrix}\right]_{\mathcal{C}} = \begin{bmatrix}1\\-2\end{bmatrix} \qquad \text{and} \qquad \left[\begin{bmatrix}1\\0\\-1\end{bmatrix}\right]_{\mathcal{C}} = \begin{bmatrix}1\\-1\end{bmatrix},$$

and $[\boldsymbol{I}]_{\mathcal{B},\mathcal{C}} = \begin{bmatrix}1 & 1\\-2 & -1\end{bmatrix}$. Then $[\boldsymbol{I}]_{\mathcal{C},\mathcal{B}} = [\boldsymbol{I}]_{\mathcal{B},\mathcal{C}}^{-1} = \begin{bmatrix}-1 & -1\\2 & 1\end{bmatrix}$.

(c) $[\boldsymbol{T}]_{\mathcal{C}} = [\boldsymbol{I}]_{\mathcal{B},\mathcal{C}}\,[\boldsymbol{T}]_{\mathcal{B}}\,[\boldsymbol{I}]_{\mathcal{C},\mathcal{B}} = \begin{bmatrix}1 & 0\\-3 & -1\end{bmatrix}$.

3.6.10. (a) Since $\boldsymbol{T}(2,-1,0) = (0,-3,2) = -(0,3,-2)$ and $\boldsymbol{T}(0,-3,2) = (-18,9,0) = -9(2,-1,0)$,

$$[\boldsymbol{T}]_{\mathcal{B}} = \begin{bmatrix}0 & -9\\-1 & 0\end{bmatrix}.$$

(b) Writing the elements of $\mathcal{B}$ in terms of $\mathcal{C}$:

$$\left[\begin{array}{cc|cc}1 & 3 & 2 & 0\\1 & 0 & -1 & 3\\-1 & -1 & 0 & -2\end{array}\right] \rightsquigarrow \left[\begin{array}{cc|cc}1 & 3 & 2 & 0\\0 & -3 & -3 & 3\\0 & 2 & 2 & -2\end{array}\right] \rightsquigarrow \left[\begin{array}{cc|cc}1 & 0 & -1 & 3\\0 & 1 & 1 & -1\\0 & 0 & 0 & 0\end{array}\right],$$

so

$$\left[\begin{bmatrix}2\\-1\\0\end{bmatrix}\right]_{\mathcal{C}} = \begin{bmatrix}-1\\1\end{bmatrix} \qquad \text{and} \qquad \left[\begin{bmatrix}0\\3\\-2\end{bmatrix}\right]_{\mathcal{C}} = \begin{bmatrix}3\\-1\end{bmatrix},$$

and $[\boldsymbol{I}]_{\mathcal{B},\mathcal{C}} = \begin{bmatrix}-1 & 3\\1 & -1\end{bmatrix}$. Then $[\boldsymbol{I}]_{\mathcal{C},\mathcal{B}} = [\boldsymbol{I}]_{\mathcal{B},\mathcal{C}}^{-1} = \frac{1}{2}\begin{bmatrix}1 & 3\\1 & 1\end{bmatrix}$.

(c) $[\boldsymbol{T}]_{\mathcal{C}} = [\boldsymbol{I}]_{\mathcal{B},\mathcal{C}}\,[\boldsymbol{T}]_{\mathcal{B}}\,[\boldsymbol{I}]_{\mathcal{C},\mathcal{B}} = \begin{bmatrix}3 & 0\\-4 & -3\end{bmatrix}$.

3.6.11. (a) $\mathbf{v} = \begin{bmatrix}3\\1\end{bmatrix}$ lies on L so $\boldsymbol{P}\mathbf{v} = \mathbf{v}$. $\mathbf{w} = \begin{bmatrix}1\\-3\end{bmatrix}$ is perpendicular to L, so $\boldsymbol{P}\mathbf{w} = \mathbf{0}$. So if $\mathcal{B} = (\mathbf{v},\mathbf{w})$ then $[\boldsymbol{P}]_{\mathcal{B}} = \mathbf{diag}(1,0)$.

(b) By Lemma 3.50, $[\boldsymbol{I}]_{\mathcal{B},\mathcal{E}} = \begin{bmatrix}3 & 1\\1 & -3\end{bmatrix}$.

(c) $[\boldsymbol{I}]_{\mathcal{E},\mathcal{B}} = [\boldsymbol{I}]_{\mathcal{B},\mathcal{E}}^{-1} = \frac{1}{10}\begin{bmatrix}3 & 1\\1 & -3\end{bmatrix}$.

(d) $[\boldsymbol{P}]_{\mathcal{E}} = [\boldsymbol{I}]_{\mathcal{B},\mathcal{E}}\,[\boldsymbol{R}]_{\mathcal{B}}\,[\boldsymbol{I}]_{\mathcal{E},\mathcal{B}} = \frac{1}{10}\begin{bmatrix}3 & 1\\1 & -3\end{bmatrix}\begin{bmatrix}1 & 0\\0 & 0\end{bmatrix}\begin{bmatrix}3 & 1\\1 & -3\end{bmatrix} = \frac{1}{10}\begin{bmatrix}9 & 3\\3 & 1\end{bmatrix}$.

3.6.12. (a) If $\mathcal{B} = (\mathbf{v}_1,\mathbf{v}_2) = \left(\begin{bmatrix}2\\1\end{bmatrix}, \begin{bmatrix}-1\\2\end{bmatrix}\right)$, then $\boldsymbol{R}\mathbf{v}_1 = \mathbf{v}_1$ and $\boldsymbol{R}\mathbf{v}_2 = -\mathbf{v}_2$, so $[\boldsymbol{R}]_{\mathcal{B}} = \begin{bmatrix}1 & 0\\0 & -1\end{bmatrix}$.

(b) $[\boldsymbol{I}]_{\mathcal{B},\mathcal{E}} = \begin{bmatrix}2 & -1\\1 & 2\end{bmatrix}$ by Lemma 3.50.

(c) $[\boldsymbol{I}]_{\mathcal{E},\mathcal{B}} = [\boldsymbol{I}]_{\mathcal{B},\mathcal{E}}^{-1} = \frac{1}{5}\begin{bmatrix} 2 & 1 \\ -1 & 2 \end{bmatrix}$.

(d) $[\boldsymbol{R}]_{\mathcal{E}} = [\boldsymbol{I}]_{\mathcal{B},\mathcal{E}}\,[\boldsymbol{R}]_{\mathcal{B}}\,[\boldsymbol{I}]_{\mathcal{E},\mathcal{B}} = \frac{1}{5}\begin{bmatrix} 2 & -1 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}\begin{bmatrix} 2 & 1 \\ -1 & 2 \end{bmatrix} = \frac{1}{5}\begin{bmatrix} 3 & 4 \\ 4 & -3 \end{bmatrix}$.

3.6.13. It's simpler to first find $\mathbf{S} = [\boldsymbol{I}]_{\mathcal{B}',\mathcal{B}}$, since it's trivial to see how to write the vectors in $\mathcal{B}'$ in terms of the vectors in $\mathcal{B}$:

$$\begin{aligned} 1 &= 1\cdot 1 + 0x + 0x^2, \\ x &= 0\cdot 1 + 1x + 0x^2, \\ \frac{3}{2}x^2 - \frac{1}{2} &= \left(-\frac{1}{2}\right) 1 + 0x + \frac{3}{2}x^2. \end{aligned}$$

Therefore

$$\mathbf{S} = \begin{bmatrix} 1 & 0 & -\frac{1}{2} \\ 0 & 1 & 0 \\ 0 & 0 & \frac{3}{2} \end{bmatrix}.$$

We can then invert this using two easy row operations to get

$$[\boldsymbol{I}]_{\mathcal{B},\mathcal{B}'} = \mathbf{S}^{-1} = \begin{bmatrix} 1 & 0 & \frac{1}{3} \\ 0 & 1 & 0 \\ 0 & 0 & \frac{2}{3} \end{bmatrix}.$$

3.6.14. (a) By formula (3.14), $[\boldsymbol{I}]_{\mathcal{C},\mathcal{B}} = [\boldsymbol{I}]_{\mathcal{E},\mathcal{B}}\,[\boldsymbol{I}]_{\mathcal{C},\mathcal{E}}$. As shown in the example on page 201, $[\boldsymbol{I}]_{\mathcal{E},\mathcal{B}} = [\boldsymbol{I}]_{\mathcal{C},\mathcal{E}} = \mathbf{B}$, so

$$[\boldsymbol{I}]_{\mathcal{C},\mathcal{B}} = \mathbf{B}^2 = \begin{bmatrix} 1 & -2 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & -2 & 1 & 0 & & \vdots \\ 0 & 0 & 1 & -2 & 1 & & \vdots \\ 0 & 0 & 0 & 1 & -2 & \ddots & \vdots \\ \vdots & & & & \ddots & \ddots & 1 \\ \vdots & & & & & \ddots & -2 \\ 0 & \cdots & \cdots & \cdots & \cdots & \cdots & 1 \end{bmatrix}.$$

(b) $[\mathbf{v}]_{\mathcal{C}} = [\boldsymbol{I}]_{\mathcal{B},\mathcal{C}}\,[\mathbf{v}]_{\mathcal{B}} = -\mathbf{e}_{n-1} + \mathbf{e}_n$.

(c) It follows directly from the definition of $\boldsymbol{T}$ that

$$\mathbf{C} = [\boldsymbol{T}]_{\mathcal{B}} = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & 0 & & & \vdots \\ 0 & 1 & 0 & & & \vdots \\ \vdots & & \ddots & \ddots & & \vdots \\ \vdots & & & \ddots & 0 & 0 \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{bmatrix},$$

and so

$$\mathbf{D} = [\boldsymbol{T}]_{\mathcal{E}} = [\boldsymbol{I}]_{\mathcal{B},\mathcal{E}} [\boldsymbol{T}]_{\mathcal{B}} [\boldsymbol{I}]_{\mathcal{E},\mathcal{B}} = \mathbf{ACB} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & -1 \\ 1 & 0 & 0 & \cdots & 0 & -1 \\ 0 & 1 & 0 & \cdots & 0 & -1 \\ 0 & 0 & 1 & \ddots & & \vdots \\ \vdots & & & \ddots & 0 & -1 \\ 0 & 0 & 0 & \cdots & 1 & -1 \end{bmatrix}.$$

(d) We can compute $[\boldsymbol{T}]_{\mathcal{C}}$ as either

$$[\boldsymbol{I}]_{\mathcal{B},\mathcal{C}} [\boldsymbol{T}]_{\mathcal{B}} [\boldsymbol{I}]_{\mathcal{C},\mathcal{B}} = \mathbf{A}^2\mathbf{C}\mathbf{B}^2$$

or

$$[\boldsymbol{I}]_{\mathcal{E},\mathcal{C}} [\boldsymbol{T}]_{\mathcal{E}} [\boldsymbol{I}]_{\mathcal{C},\mathcal{E}} = \mathbf{ADB}$$

(since we have all the matrices $\mathbf{A}$, $\mathbf{A}^2$, $\mathbf{B}$, $\mathbf{B}^2$, $\mathbf{C}$, and $\mathbf{D}$ on hand). Either way we obtain

$$[\boldsymbol{T}]_{\mathcal{C}} = \begin{bmatrix} 2 & -1 & 0 & \cdots & \cdots & 0 & -(n+1) \\ 1 & 0 & 0 & & & 0 & -n \\ 0 & 1 & 0 & & & 0 & -(n-1) \\ 0 & 0 & 1 & \ddots & & 0 & -(n-2) \\ \vdots & & & \ddots & \ddots & & \vdots \\ \vdots & & & & 1 & 0 & -3 \\ 0 & \cdots & \cdots & \cdots & 0 & 1 & -2 \end{bmatrix}$$

3.6.15. The trace of the first matrix is -1 and the trace of the second matrix is 13. By Corollary 3.61, this means that the matrices cannot be similar.

3.6.16. The trace of the first matrix is 6 and the trace of the second matrix is 0. By Corollary 3.61, this means that the matrices cannot be similar.

3.6.17. As in Exercises 2.1.9 and 2.1.10, we find that $\mathbf{A}$ has two distinct eigenvalues (1 and 2), so by Quick Exercise #29 in Chapter 3, the matrix is diagonalizable.

3.6.18. As in Exercises 2.1.9 and 2.1.10, we find that $\mathbf{A}$ has two distinct eigenvalues (-1 and 4), so by Quick Exercise #29 in Chapter 3, the matrix is diagonalizable.

3.6.19. The Example itself shows that $\mathbf{A}$ is similar to $\mathbf{diag}(i, -i)$, which implies that $\pm i$ are eigenvalues of $\mathbf{A}$, hence they are the only eigenvalues of $\mathbf{A}$ in $\mathbb{C}$. Therefore $\mathbf{A}$ has no eigenvalues in $\mathbb{R}$, and so $\mathbf{A}$ cannot be diagonalizable over $\mathbb{R}$.

3.6.20. (a) It is clear that $\mathbf{A} - \lambda\mathbf{I} = \begin{bmatrix} 1-\lambda & 1 \\ 0 & 1-\lambda \end{bmatrix}$ has rank 2 (since the columns are linearly independent) unless $\lambda = 1$, so this is the only eigenvalue of $\mathbf{A}$. An eigenvector is a solution to

$$\left[\begin{array}{cc|c} 0 & 1 & 0 \\ 0 & 0 & 0 \end{array}\right],$$

and so the 1-eigenspace is spanned by $\begin{bmatrix}1\\0\end{bmatrix}$.

(b) The matrix $\mathbf{A}$ is diagonalizable if and only if there is a basis of $\mathbb{F}^2$ consisting of eigenvectors of $\mathbf{A}$. But this is impossible because we just found that the dimension of the only eigenspace is 1. This entire argument works in any field (the only field elements appearing here are 0 and 1).

3.6.21. Let $\mathbf{u}$ and $\mathbf{v}$ be any two noncollinear vectors in P; then they are linearly independent. Now let $\mathbf{w}$ be any vector which is perpendicular to P. Then $\mathbf{w}$ is not in the span of $\mathbf{u}$ and $\mathbf{v}$, so by the Linear Dependence Lemma, $\mathcal{B} = (\mathbf{u}, \mathbf{v}, \mathbf{w})$ is linearly independent, and hence is a basis of $\mathbb{R}^3$. Since $\boldsymbol{P}\mathbf{u} = \mathbf{u}$, $\boldsymbol{p}\mathbf{v} = \mathbf{v}$, and $\boldsymbol{P}\mathbf{w} = -\mathbf{w}$, we have $[\boldsymbol{P}]_{\mathcal{B}} = \mathbf{diag}(1, 1, -1)$. Then $\operatorname{tr} \boldsymbol{P} = \operatorname{tr} [\boldsymbol{P}]_{\mathcal{B}} = 1 + 1 - 1 = 1$.

3.6.22. (a) By Proposition 3.60, $\operatorname{tr} \mathbf{A}(\mathbf{BC}) = \operatorname{tr}(\mathbf{BC})\mathbf{A} = \operatorname{tr} \mathbf{B}(\mathbf{CA}) = \operatorname{tr}(\mathbf{CA})\mathbf{B}$.

(b) There are infinitely many possibilities. One of the simplest is:

$$\mathbf{A} = \begin{bmatrix}1 & 0\\0 & 0\end{bmatrix}, \qquad \mathbf{B} = \begin{bmatrix}0 & 0\\1 & 0\end{bmatrix}, \qquad \mathbf{C} = \begin{bmatrix}0 & 1\\0 & 0\end{bmatrix},$$

so that $\operatorname{tr} \mathbf{ABC} = \operatorname{tr} \begin{bmatrix}0 & 0\\0 & 0\end{bmatrix} = 0$, but $\operatorname{tr} \mathbf{ACB} = \operatorname{tr} \begin{bmatrix}1 & 0\\0 & 0\end{bmatrix} = 1$.

3.6.23. If $\boldsymbol{T}$ has n distinct eigenvalues $\lambda_1, \ldots, \lambda_n$, then there are corresponding eigenvectors $v_1, \ldots, v_n$, which are linearly independent by Theorem 3.8. Then $\mathcal{B} = (v_1, \ldots, v_n)$ is a basis of V by Theorem 3.28, and since $\boldsymbol{T}v_j = \lambda v_j$ for every j, we have $[\boldsymbol{T}]_{\mathcal{B}} = \mathbf{diag}(\lambda_1, \ldots, \lambda_n)$. Therefore $\operatorname{tr} \boldsymbol{T} = \operatorname{tr} \mathbf{diag}(\lambda_1, \ldots, \lambda_n) = \lambda_1 + \cdots + \lambda_n$.

3.6.24. (a) This is just a restatement of Theorem 2.41.

(b) If $\mathbf{A} = \mathbf{SBS}^{-1}$, then $\mathbf{A} - \lambda\mathbf{I}_n = \mathbf{S}(\mathbf{B} - \lambda\mathbf{I}_n)\mathbf{S}^{-1}$. Then Corollary 3.59 implies that $\operatorname{null}(\mathbf{A} - \lambda\mathbf{I}_n) = \operatorname{null}(\mathbf{B} - \lambda\mathbf{I}_n)$.

3.6.25. Let $\mathbf{A} = \begin{bmatrix}1 & 1\\0 & 2\end{bmatrix}$. Then $\mathbf{A}$ has distinct eigenvalues 1 and 2 and is therefore similar to $\begin{bmatrix}1 & 0\\0 & 2\end{bmatrix} = \mathbf{D}$. But $\operatorname{sum} \mathbf{A} = 4 \neq 3 = \operatorname{sum} \mathbf{D}$.

3.6.26. Let $\mathbf{S} = \begin{bmatrix}0 & 1\\1 & 0\end{bmatrix}$ and $\mathbf{A} = \begin{bmatrix}1 & 0\\0 & 2\end{bmatrix}$. Then $\mathbf{S}^{-1} = \mathbf{S}$ and $\mathbf{SAS}^{-1} = \begin{bmatrix}2 & 0\\0 & 1\end{bmatrix}$, which has a different top-left entry from $\mathbf{A}$.

3.6.27. (a) Take $\mathbf{I}_n\mathbf{A}\mathbf{I_n}^{-1} = \mathbf{A}$, so $\mathbf{A}$ is similar to $\mathbf{A}$.

(b) Suppose that $\mathbf{B} = \mathbf{SAS}^{-1}$ and $\mathbf{C} = \mathbf{TBT}^{-1}$. Then $\mathbf{C} = \mathbf{TSAS}^{-1}\mathbf{T}^{-1} = (\mathbf{TS})\mathbf{A}(\mathbf{TS})^{-1}$ and thus $\mathbf{C}$ is similar to $\mathbf{A}$.

3.6.28. Suppose that V and W are both finite dimensional. Let $\mathcal{B}_V$ and $\mathcal{B}_W$ be bases of V and W, respectively. By Theorem 3.58 and the matrix version of the Rank–Nullity theorem,

$$\operatorname{rank} \boldsymbol{T} + \operatorname{null} \boldsymbol{T} = \operatorname{rank} [\boldsymbol{T}]_{\mathcal{B}_V,\mathcal{B}_W} + \operatorname{null} [\boldsymbol{T}]_{\mathcal{B}_V,\mathcal{B}_W} = \dim V,$$

since $[\boldsymbol{T}]_{\mathcal{B}_V,\mathcal{B}_W}$ is a $(\dim W) \times (\dim V)$ matrix.

If W is not finite dimensional, we may replace W with $W' = \operatorname{range} \boldsymbol{T}$, which is finite dimensional since it is spanned by $(\boldsymbol{T}v_1, \dots, \boldsymbol{T}v_n)$, where $(v_1, \dots, v_n)$ is any list of vectors which spans V.

3.6.29. Suppose $\mathbf{A}$ is diagonalizable, and let $\boldsymbol{T}$ be the operator with $[\boldsymbol{T}]_{\mathcal{E}} = \mathbf{A}$. By definition of diagonalizability for a matrix, $\mathbf{A}$ is similar to some diagonal matrix $\mathbf{D}$. By Theorem 3.54, there is a basis $\mathcal{B}$ of $\mathbb{F}^n$ such that $[\boldsymbol{T}]_{\mathcal{B}} = \mathbf{D}$. Then by Proposition 3.45, each vector in $\mathcal{B}$ is an eigenvector of $\boldsymbol{T}$, and hence of $\mathbf{A}$.

Now suppose that $\mathcal{B} = (\mathbf{v}_1, \dots, \mathbf{v}_n)$ is a basis of $\mathbb{F}^n$ consisting of eigenvectors of $\mathbf{A}$. Again let $\boldsymbol{T}$ be the operator with $[\boldsymbol{T}]_{\mathcal{E}} = \mathbf{A}$, so each $\mathbf{v}_j$ is an eigenvector of $\boldsymbol{T}$. Then by Proposition 3.45, $[\boldsymbol{T}]_{\mathcal{B}} = \mathbf{diag}(\lambda_1, \dots, \lambda_n)$, where $\mathbf{A}\mathbf{v}_j = \lambda_j \mathbf{v}_j$. By Theorem 3.54, $\mathbf{A} = [\boldsymbol{T}]_{\mathcal{E}}$ is similar to the diagonal matrix $[\boldsymbol{T}]_{\mathcal{B}}$, and hence is diagonalizable.

Moreover, by the change of basis formula, in this case $\mathbf{A} = [\boldsymbol{T}]_{\mathcal{E}} = \mathbf{S}\,[\boldsymbol{T}]_{\mathcal{B}}\,\mathbf{S}^{-1}$, where $\mathbf{S} = [\boldsymbol{I}]_{\mathcal{B},\mathcal{E}} = [\mathbf{v}_1 \cdots \mathbf{v}_n]$.

3.7 Triangularization

3.7.1. Since all the matrices are triangular, the eigenvalues are equal to the diagonal entries, and then $\operatorname{Eig}_\lambda(\mathbf{A}) = \ker(\mathbf{A} - \lambda \mathbf{I}_n)$ can be found as usual via Gaussian elimination.

(a) $\left(1, \left\langle \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\rangle\right), \left(3, \left\langle \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\rangle\right)$

(b) $\left(1, \left\langle \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \right\rangle\right), \left(4, \left\langle \begin{bmatrix} 2 \\ 3 \\ 0 \end{bmatrix} \right\rangle\right), \left(6, \left\langle \begin{bmatrix} 16 \\ 25 \\ 10 \end{bmatrix} \right\rangle\right)$

(c) $\left(-2, \left\langle \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \right\rangle\right), \left(3, \left\langle \begin{bmatrix} 1 \\ 5 \\ 0 \end{bmatrix} \right\rangle\right)$

(d) $\left(1, \left\langle \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \end{bmatrix} \right\rangle\right), \left(2, \left\langle \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \right\rangle\right)$

3.7.2. Since all the matrices are triangular, the eigenvalues are equal to the diagonal entries, and then $\operatorname{Eig}_\lambda(\mathbf{A}) = \ker(\mathbf{A} - \lambda \mathbf{I}_n)$ can be found as usual via Gaussian elimination.

(a) $\left(1, \left\langle \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\rangle\right)$

(b) $\left(3, \left\langle \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\rangle\right), \left(1, \left\langle \begin{bmatrix} 1 \\ -2 \\ 0 \end{bmatrix} \right\rangle\right), \left(9, \left\langle \begin{bmatrix} 37 \\ 30 \\ 48 \end{bmatrix} \right\rangle\right)$

(c) $\left(2, \left\langle \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \right\rangle \right)$

(d) $\left(1, \left\langle \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \right\rangle \right), \left(2, \left\langle \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \right\rangle \right)$

3.7.3. (a) Yes: this matrix has two distinct eigenvalues (8 and $\sqrt{5}$), which therefore have linearly independent eigenvectors, which then are a basis for $\mathbb{R}^2$.

(b) Yes. This matrix has two distinct eigenvalues: 2 and 1. The 1-eigenspace is the null space of the matrix

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

which hass rank one and thus has a two-dimensional null space. We can therefore choose two linearly independent eigenvectors with eigenvalue 1. Because 2 is also an eigenvalue, we can choose a corresponding eigenvector, which is linearly independent with the first two by Theorem 3.8 (any linear combination of the first two is still an eigenvector with eigenvalue 1). Since we have produced 3 linearly independent eigenvectors, $\mathbb{R}^3$ has a basis of eigenvectors and so the matrix is diagonalizable.

(c) No. The only eigenvalues of this matrix are 2 and 1. The eigenspace for each eigenvalue is only one-dimensional. Therefore any linearly independent list of eigenvectors of this matrix has length at most two, so there is no basis of $\mathbb{R}^3$ consisting of eigenvectors of this matrix.

(d) Yes: this matrix has three distinct eigenvalues (9, 5, and 1), which therefore have linearly independent eigenvectors, which then are a basis for $\mathbb{R}^3$.

3.7.4. (a) Yes: this matrix has two distinct eigenvalues (i and $\sqrt{17} - 4i$), which therefore have linearly independent eigenvectors, which then are a basis for $\mathbb{C}^2$.

(b) No. The only eigenvalue of this matrix is 1, and the corresponding eigenspace is the null space of

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

Since this matrix has rank 2, its nullity is 1 and this means that all eigenvectors lie in a one-dimensional subspace of $\mathbb{R}^3$: there cannot be a basis of $\mathbb{R}^3$ consisting of eigenvectors.

(c) Yes. This matrix has two distinct eigenvalues: 3 and 1. The 3-eigenspace is the null space of the matrix

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

which is rank one and thus has a two-dimensional null space. We can therefore choose two linearly independent eigenvectors with eigenvalue 3. Because 1 is also an eigenvalue, we can choose a corresponding eigenvector, which is linearly independent with the first two by Theorem 3.8 (any linear combination of the first two is still an eigenvector with eigenvalue 3). Since we have produced 3 linearly independent eigenvectors, $\mathbb{R}^3$ has a basis of eigenvectors and so the matrix is diagonalizable.

(d) No. This is similar to (b): the 2-eigenspace is the null space of

$$\begin{bmatrix} 0 & 0 & 7 \\ 0 & -1 & 8 \\ 0 & 0 & 0 \end{bmatrix},$$

and since this is a rank 2 matrix, its null space is one-dimensional. The 1-eigenspace is also one-dimensional, and so we cannot find a basis of eigenvectors.

3.7.5. Let $\mathbf{v}_j = \mathbf{e}_{n-j+1}$, so $(\mathbf{v}_1, \ldots, \mathbf{v}_n) = (\mathbf{e}_n, \ldots, \mathbf{e}_1)$. If $\mathbf{A}$ is lower triangular, then

$$\mathbf{A}\mathbf{v}_j = \mathbf{A}\mathbf{e}_{n-j+1} \in \langle \mathbf{e}_{n-j+1}, \mathbf{e}_{n-j+2}, \ldots, \mathbf{e}_n \rangle = \langle \mathbf{v}_1, \ldots, \mathbf{v}_j \rangle,$$

so by Lemma 3.68, the linear map represented by $\mathbf{A}$ (with respect to the standard basis) has an upper triangular matrix $\mathbf{B}$ with respect to the basis $(\mathbf{v}_1, \ldots, \mathbf{v}_n)$; $\mathbf{A}$ is then similar to $\mathbf{B}$. More concretely, $\mathbf{A} = \mathbf{S}\mathbf{B}\mathbf{S}^{-1} = \mathbf{S}\mathbf{B}\mathbf{S}$, where

$$\mathbf{S} = \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \\ 0 & & & 1 & 0 \\ \vdots & & \iddots & & \vdots \\ 0 & 1 & & & 0 \\ 1 & 0 & \cdots & 0 & 0 \end{bmatrix}.$$

3.7.6. In the LUP decomposition, the matrix $\mathbf{L}$ is square and lower triangular with 1s on the diagonal; by Exercise 3.7.5 and Lemma 3.63, $\mathbf{L}$ is invertible. Also $\mathbf{P}$ is invertible, since it is the product of (invertible) elementary matrices. Therefore $\mathbf{A} = \mathbf{S}\mathbf{U}$, where $\mathbf{S} = \mathbf{P}^{-1}\mathbf{L}$ is invertible. If $\mathbf{x} \in \ker \mathbf{U}$, then $\mathbf{A}\mathbf{x} = \mathbf{S}(\mathbf{U}\mathbf{x}) = \mathbf{0}$, so $x \in \ker \mathbf{U}$. Similarly, if $\mathbf{x} \in \ker \mathbf{A}$, then $\mathbf{U}\mathbf{x} = \mathbf{S}^{-1}(\mathbf{A}\mathbf{x}) = \mathbf{0}$.

3.7.7. (a) By Theorem 3.64, the eigenvalues of $\mathbf{A}$ and $\mathbf{B}$ are exactly their diagonal entries. Since $\mathbf{A} + \mathbf{B}$ is also upper triangular, its eigenvalues are its diagonal entries, which are all sums of the corresponding entries from $\mathbf{A}$ and $\mathbf{B}$.

(b) By Exercise 2.3.12, $\mathbf{AB}$ is upper triangular, and its diagonal entries are $a_{ii}b_{ii}$. Therefore by Theorem 3.64, the eigenvalues of $\mathbf{AB}$ are just the diagonal entries are $a_{ii}b_{ii}$ for $i = 1, \ldots, n$, which are all products of the eigenvalues of the triangular matrices $\mathbf{A}$ and $\mathbf{B}$.

(c) Let $\mathbf{A} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$. The only eigenvalues of $\mathbf{A}$ and $\mathbf{B}$ are 0, but the eigenvalues of

$$\mathbf{A} + \mathbf{B} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

are ± 1, and

$$\mathbf{AB} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

has eigenvalues 0 and 1.

3.7.8. Each column of $\mathbf{A}$ with a nonzero diagonal entry does not lie in the span of the preceding columns (all of which have entry 0 in the position of that column's diagonal entry), so by the Linear Dependence Lemma those k columns are linearly independent. Therefore $\operatorname{rank} \mathbf{A} = \dim C(\mathbf{A}) \geq k$.

3.7.9. By Theorem 3.64, the eigenvalues of $\mathbf{A}$ are its diagonal entries, so $\mathbf{A}$ has n distinct eigenvalues. By Quick Exercise #29 in Chapter 3, $\mathbf{A}$ is diagonalizable.

3.7.10. (a) We know that $\mathbf{Ax} = \lambda \mathbf{x}$, and so for any $m \in \mathbb{N}$,

$$\mathbf{A}^m \mathbf{x} = \mathbf{A}^{m-1}(\mathbf{Ax}) = \mathbf{A}^{m-1}(\lambda \mathbf{x}) = \lambda \mathbf{A}^{m-1}\mathbf{x} = \cdots = \lambda^m \mathbf{x};$$

that is, $\mathbf{x}$ is an eigenvector of $\mathbf{A}^m$ with eigenvalue λ^m. If $p(x) = a_0 + a_1 x + \cdots + a_m x^m$, we thus have

$$p(\mathbf{A})\mathbf{x} = (a_0 \mathbf{I} + a_1 \mathbf{A} + \cdots + a_m \mathbf{A}^m)\, \mathbf{x} = a_0 \mathbf{x} + a_1 \lambda \mathbf{x} + \cdots + a_n \lambda^n \mathbf{x} = p(\lambda)\mathbf{x}.$$

That is, $\mathbf{x}$ is an eigenvector of $p(\mathbf{A})$ with eigenvalue $p(\lambda)$.

(b) By part (a), $p(\mathbf{A})\mathbf{x} = p(\lambda)\mathbf{x}$, so if $p(\mathbf{A})\mathbf{x} = \mathbf{0}$, then $p(\lambda)\mathbf{x} = \mathbf{0}$. Since $\mathbf{x} \neq \mathbf{0}$ (because $\mathbf{x}$ is an eigenvector), it follows that $p(\lambda) = 0$.

(c) If $p(\mathbf{A}) = \mathbf{0}$, then for any eigenvalue λ with corresponding eigenvector $\mathbf{x}$, we have that $p(\mathbf{A})\mathbf{x} = 0$, and so $p(\lambda) = 0$ by part (b).

3.7.11. Since $\mathbb{F}$ is algebraically closed, Theorem 3.67 implies that $\boldsymbol{T}$ can be represented in some basis by an upper triangular matrix $\mathbf{A}$. Exercise 2.3.12 shows that the product of upper triangular matrices is upper triangular, so $\mathbf{A}^m$ is upper triangular for each $m \in \mathbb{N}$. Moreover, that exercise also shows that that if $\mathbf{A}$ and $\mathbf{B}$ are upper triangular, then the diagonal entries of $\mathbf{AB}$ are $a_{11}b_{11}, \ldots, a_{nn}b_{nn}$; this means that the diagonal entries of $\mathbf{A}^m$ are $a_{11}^m, \ldots, a_{nn}^m$. From there it follows that $p(\mathbf{A})$ is upper triangular with diagonal entries $p(a_{11}), \ldots, p(a_{nn})$. Since the eigenvalues of an upper triangular matrix are exactly the diagonal entries, this means that $a_{11}, \ldots, a_{nn}$ are the eigenvalues of $\boldsymbol{T}$, and the eigenvalues of $p(\boldsymbol{T})$ are then exactly $p(\lambda)$ for λ an eigenvalue of $\boldsymbol{T}$.

3.7.12. Suppose that there is a fixed k such that $a_{ij} = 0$ whenever $i > j - k$. (That is, the non-zero entries of $\mathbf{A}$ lie in a triangle in the top-right corner, whose sides have length $n-k$). Let $\mathbf{B}$ be strictly upper triangular. Then

$$[\mathbf{AB}]_{ij} = \sum_{\ell} a_{i\ell} b_{\ell j},$$

which is zero if $i > j - k + 1$, since if $\ell < i + k$, then $a_{i\ell} = 0$, and if $\ell \geq i + k > j + 1 \geq j$, then $b_{\ell j} = 0$. It follows that if $\mathbf{A}$ is strictly upper triangular, then $\mathbf{A}^2$ has non-zero entries only in a triangle of side lengths $n-2$, $\mathbf{A}^3$ has non-zero entries in a triangle of side lengths $n-3$, and so on: $\mathbf{A}^{n-1}$ can only have something non-zero in the top-right entry, and $\mathbf{A}^n$ itself has no non-zero entries.

3.7.13. By Theorem 3.67, $\mathbf{A}$ is similar to an upper triangular matrix $\mathbf{B}$, whose diagonal entries are all equal to λ by Theorem 3.64. Therefore $\mathbf{B} - \lambda\mathbf{I}_n$ is upper triangular, so by Exercise 3.7.12, $(\mathbf{B} - \lambda\mathbf{I}_n)^n = \mathbf{0}$. Since $\mathbf{B}$ is similar to $\mathbf{A}$, $\mathbf{B} - \lambda\mathbf{I}_n$ is similar to $\mathbf{A} - \lambda\mathbf{I}_n$, and it follows that $(\mathbf{A} - \lambda\mathbf{I}_n)^n = \mathbf{0}$.

3.7.14. If $\dim V = n$, then $\dim(\mathcal{L}(V)) = n^2$ by Corollary 3.43. Given $\boldsymbol{T} \in \mathcal{L}(V)$, consider the list

$$\left(\boldsymbol{I}, \boldsymbol{T}, \boldsymbol{T}^2, \dots, \boldsymbol{T}^{n^2}\right).$$

Since this list has length $n^2 + 1 > \dim(\mathcal{L}(V))$, it must be linearly dependent, and so there are scalars $a_0, \dots, a_{n^2}$ such that

$$a_0\boldsymbol{I} + a_1\boldsymbol{T} + \cdots + a_{n^2}\boldsymbol{T}^{n^2} = \mathbf{0}.$$

If we take $p(x) = a_0 + a_1 x + \cdots + a_{n^2} x^{n^2}$, the this says that $p(\boldsymbol{T}) = \mathbf{0}$.

3.7.15. The only roots of the polynomial $p(x) = x^2 - 2$ in $\mathbb{C}$ are $\pm\sqrt{2}$, which do not lie in $\mathbb{F}$. Therefore $p(x)$ cannot be factored into linear factors over $\mathbb{F}$.

3.7.16. Row-reducing $\mathbf{A} - \lambda\mathbf{I}_2 = \mathbf{0}$ shows that λ is an eigenvalue of $\mathbf{A}$ iff λ is a root of $p(x) = x^2 + x + 1$. Since $p(0) = p(1) = 1$, $\mathbf{A}$ has no root in $\mathbb{F}_2$.

Alternatively, one can check that $\mathbf{A}^2 + \mathbf{A} + \mathbf{I}_2 = \mathbf{0}$ (using $\mathbb{F}_2$ arithmetic), so by Exercise 3.7.10(c), each eigenvalue of $\mathbf{A}$ is a root of $p(x)$, and the proof is completed as before.

3.7.17. Let $a_1, \dots, a_n$ be all the distinct elements of $\mathbb{F}$, and define $p(x) = (x - a_1)\cdots(x - a_n) + 1$. Then $p(a) = 1$ for each $a \in \mathbb{F}$, so $p(x)$ has no roots in $\mathbb{F}$, and therefore cannot be factored into linear factors over $\mathbb{F}$.

Chapter 4

Inner products

4.1 Inner product spaces

4.1.1. (a) This list is orthogonal with respect to the standard inner product on $\mathbb{R}^4$.

(b) This list is orthogonal with respect to the standard inner product on $\mathbb{C}^3$.

(c) This list is orthogonal with respect to the Frobenius inner on $\mathrm{M}_{2,3}(\mathbb{R})$.

(d) This list is linearly indepedent with respect to the L^2 inner product from Example 4 on p. 228.

4.1.2. We are given that

$$3 = \|v-w\|^2 = \langle v-w, v-w\rangle = \langle v,v\rangle - \langle v,w\rangle - \langle w,v\rangle + \langle w,w\rangle = \|v\|^2 + \|w\|^2 - 2\langle v,w\rangle$$

and similarly

$$7 = \|v\|^2 + \|w\|^2 + 2\langle v,w\rangle$$

Adding these equations gives $2(\|v\|^2 + \|w\|^2) = 10$, so $\|v\|^2 + \|w\|^2 = 5$. Subtracting them gives $4\langle v,w\rangle = 4$, so $\langle v,w\rangle = 1$.

4.1.3. We are given that

$$10 = \|v+w\|^2 = \langle v-w, v-w\rangle = \langle v,v\rangle + \langle v,w\rangle + \langle w,v\rangle + \langle w,w\rangle = \|v\|^2 + \|w\|^2 + 2\langle v,w\rangle$$

and similarly

$$16 = \|v\|^2 + \|w\|^2 - 2\langle v,w\rangle$$

Adding these equations gives $2(\|v\|^2 + \|w\|^2) = 26$, so $\|v\|^2 + \|w\|^2 = 13$. Subtracting them gives $4\langle v,w\rangle = -6$, so $\langle v,w\rangle = -3/2$.

4.1.4. By definition,

$$\langle \mathbf{A}, \mathbf{B}\rangle_F = \sum_{j=1}^{n}\sum_{k=1}^{n} a_{kj}\overline{b_{kj}}.$$

The inner sum is exactly $\langle \mathbf{a}_j, \mathbf{b}_j\rangle$, where $\mathbf{a}_j$ and $\mathbf{b}_j$ denote the jth columns of $\mathbf{A}$ and $\mathbf{B}$, respectively. By applying this when $\mathbf{B} = \mathbf{A}$, we get that

$$\|\mathbf{A}\|_F^2 = \sum_{j=1}^{n} \langle \mathbf{a}_j, \mathbf{a}_j\rangle = \sum_{j=1}^{n} \|\mathbf{a}_j\|^2.$$

4.1.5. (a) By the linearity of the trace and Proposition 3.60,

$$\begin{aligned}\langle \operatorname{Re}\mathbf{A}, \operatorname{Im}\mathbf{A}\rangle_F &= \frac{i}{4}\operatorname{tr}(\mathbf{A}+\mathbf{A}^*)(\mathbf{A}-\mathbf{A}^*)^* \\ &= \frac{i}{4}\operatorname{tr}(\mathbf{A}+\mathbf{A}^*)(\mathbf{A}^*-\mathbf{A}) \\ &= \frac{i}{4}(\operatorname{tr}\mathbf{A}\mathbf{A}^* + \operatorname{tr}(\mathbf{A}^*)^2 - \operatorname{tr}\mathbf{A}^2 - \operatorname{tr}\mathbf{A}^*\mathbf{A}) \\ &= \frac{i}{4}(\operatorname{tr}(\mathbf{A}^*)^2 - \operatorname{tr}\mathbf{A}^2).\end{aligned}$$

Since $\mathbf{A}$ has real entries,

$$\operatorname{tr}(\mathbf{A}^*)^2 = \operatorname{tr}(\mathbf{A}^{\mathrm{T}})^2 = \operatorname{tr}(\mathbf{A}^2)^{\mathrm{T}} = \operatorname{tr}\mathbf{A}^2,$$

and so $\langle \operatorname{Re}\mathbf{A}, \operatorname{Im}\mathbf{A}\rangle_F = 0$.

(b) This follows immediately from part (a) and Theorem 4.4.

4.1.6. The jth row of $\mathbf{A}^*$ is $\mathbf{a}_j^*$. The first expression for the (j,k) entry follows from Lemma 2.13, and the second from Quick Exercise #2.

4.1.7. This can be done via integration by parts and trig identities, but it's easier via complex exponentials. (See Appendix A.2 for details.) From Euler's formula (Theorem A.4), it follows that

$$\cos x = \frac{e^{ix}+e^{-ix}}{2} \qquad \text{and} \qquad \sin x = \frac{e^{ix}-e^{-ix}}{2i}.$$

(a) By definition of the norm,

$$\begin{aligned}\|f\|^2 &= \int_0^{2\pi} \sin^2 x\,dx = -\frac{1}{4}\int_0^{2\pi}(e^{ix}-e^{-ix})^2\,dx \\ &= -\frac{1}{4}\int_0^{2\pi}(e^{2ix}-2+e^{-2ix})\,dx = -\frac{1}{4}\left[\frac{1}{2i}e^{2ix} - 2x + \frac{1}{-2i}e^{2ix}\right]_0^{2\pi} = \pi,\end{aligned}$$

so $\|f\| = \sqrt{\pi}$.

(b) Similarly,

$$\|g\|^2 = \int_0^{2\pi}\cos^2 x\,dx = \frac{1}{4}\int_0^{2\pi}(e^{2ix}+2+e^{-2ix})\,dx = \pi,$$

so $\|g\| = \sqrt{\pi}$.

(c)

$$\begin{aligned}\langle f, g\rangle &= \int_0^{2\pi}\sin x\cos x\,dx = \frac{1}{4i}\int_0^{2\pi}(e^{2ix}-e^{-2ix})\,dx \\ &= \frac{1}{4i}\left[\frac{1}{2i}e^{2ix} - \frac{1}{-2i}e^{2ix}\right]_0^{2\pi} = 0.\end{aligned}$$

(d) By part (c) and Theorem 4.4, $\|af+bg\|^2 = a^2\|f\|^2 + b^2\|g\|^2$, so by parts (a) and (b),

$$\|af+bg\| = \sqrt{\pi(a^2+b^2)}.$$

4.1.8. Suppose there were such a w. By applying the triangle inequality twice, we get a contradiction:

$$9 = \|v\| - \|u\| \le \|v-u\| \le \|v-w\| + \|w-u\| \le 8.$$

4.1.9. **The inner product of two vectors is a scalar:** For each $v_1, v_2 \in V$, $\langle v_1, v_2\rangle_{\boldsymbol{T}} = \langle \boldsymbol{T}v_1, \boldsymbol{T}v_2\rangle \in \mathbb{F}$ because the inner product of two vectors in W is known to be a scalar.

Distributive law: For each $v_1, v_2, v_3 \in V$,

$$\begin{aligned}\langle v_1+v_2, v_3\rangle_{\boldsymbol{T}} &= \langle \boldsymbol{T}(v_1+v_2), \boldsymbol{T}v_3\rangle \\ &= \langle \boldsymbol{T}v_1, \boldsymbol{T}v_3\rangle + \langle \boldsymbol{T}v_2, \boldsymbol{T}v_3\rangle \\ &= \langle v_1, v_3\rangle_{\boldsymbol{T}} + \langle v_2, v_3\rangle_{\boldsymbol{T}}\end{aligned}$$

by the linearity of $\boldsymbol{T}$ and the distributive law for the inner product on W.

Homogeneity: This is similar to the distributive law.

Symmetry: For each $v_1, v_2 \in V$,

$$\langle v_1, v_2\rangle_{\boldsymbol{T}} = \langle \boldsymbol{T}v_1, \boldsymbol{T}v_2\rangle = \overline{\langle \boldsymbol{T}v_2, \boldsymbol{T}v_1\rangle} = \overline{\langle v_2, v_1\rangle_{\boldsymbol{T}}}$$

by the symmetry of the inner product on W.

Nonnegativity: For each $v \in V$,

$$\langle v, v\rangle_{\boldsymbol{T}} = \langle \boldsymbol{T}v, \boldsymbol{T}v\rangle \ge 0$$

by the nonnegativity property of the inner product on W.

Definiteness: If $\langle v, v\rangle_{\boldsymbol{T}} = 0$, then $\langle \boldsymbol{T}v, \boldsymbol{T}v\rangle = 0$. By the definiteness of the inner product on W, this means that $\boldsymbol{T}v = 0$, and so $v = \boldsymbol{T}^{-1}0 = 0$.

4.1.10. This follows from Exercise 4.1.9, applied to $\boldsymbol{T}^{-1}$.

4.1.11. (a) Using the dot product notation for the standard inner product on $\mathbb{R}^2$, the given formula states that

$$\langle \mathbf{x}, \mathbf{y}\rangle = \mathbf{A}\mathbf{x} \cdot \mathbf{A}\mathbf{y},$$

where $\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$. Since $\mathbf{A}$ is invertible (it's diagonal with only nonzero diagonal entries), it represents an injective linear map, so Exercise 4.1.9 implies that $\langle \cdot, \cdot\rangle$ is an inner product on $\mathbb{R}^2$.

(b) $\begin{bmatrix} 1 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ -1 \end{bmatrix} = 0$, but $\left\langle \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right\rangle = 1 - 4 = -3.$

(c) $\left\langle \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ -1 \end{bmatrix} \right\rangle = 0$, but $\begin{bmatrix} 2 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ -1 \end{bmatrix} = 3.$

4.1.12. (a) For any $w \in V$,

$$\langle 0, w\rangle = \langle 0+0, w\rangle = \langle 0, w\rangle + \langle 0, w\rangle,$$

and so $\langle 0, w\rangle = 0$. Conversely, if $\langle v, w\rangle = 0$ for all $w \in V$, take $w = v$ to see that $\langle v, v\rangle = 0$, which by definiteness means that $v = 0$.

(b) First note that $v = w$ if and only if $v = w = 0$. By part (a), this is the case if and only if, for every $u \in V$, $\langle v - w, u\rangle = 0$, or equivalently, if $\langle v, u\rangle = \langle w, u\rangle$.

(c) We have $\boldsymbol{S} = \boldsymbol{T}$ if and only if $\boldsymbol{S}v_1 = \boldsymbol{T}v_1$ for all $v_1 \in V$. For a given $v_1 \in V$, it follows from part (c) that $\boldsymbol{S}v_1 = \boldsymbol{T}v_1$ if and only if $\langle \boldsymbol{S}v_1, v_2\rangle = \langle \boldsymbol{T}v_1, v_2\rangle$ for all $v_2 \in V$.

4.1.13. If $\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ and $\mathbf{D} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$. Then $\mathbf{A}$ has the distinct eigenvalues ± 1, so it is diagonalizable and hence similar to $\mathbf{D}$. But $\|\mathbf{A}\|_F = \sqrt{3}$ and $\|\mathbf{D}\|_F = \sqrt{2}$.

4.1.14. (a) Simply expand the squared norms in terms of the inner product and use linearity and symmetry:

$$\begin{aligned} \|v+w\|^2 - \|v-w\|^2 &= \langle v+w, v+w\rangle - \langle v-w, v-w\rangle \\ &= \langle v, v\rangle + \langle v, w\rangle + \langle w, v\rangle + \langle w, w\rangle \\ &\qquad - \big(\langle v, v\rangle - \langle v, w\rangle - \langle w, v\rangle + \langle w, w\rangle\big) \\ &= 4\langle v, w\rangle. \end{aligned}$$

(b) In the complex case we use conjugate linearity in the second slot:

$$\begin{aligned} &\|v+w\|^2 - \|v-w\|^2 + i\,\|v+iw\|^2 - i\,\|v-iw\|^2 \\ &\quad= \langle v+w, v+w\rangle - \langle v-w, v-w\rangle + i\,\langle v+iw, v+iw\rangle - i\,\langle v-iw, v-iw\rangle \\ &\quad= \langle v, v\rangle + \langle v, w\rangle + \langle w, v\rangle + \langle w, w\rangle \\ &\qquad\quad - \big(\langle v, v\rangle - \langle v, w\rangle - \langle w, v\rangle + \langle w, w\rangle\big) \\ &\qquad\quad + i\big(\langle v, v\rangle - i\,\langle v, w\rangle + i\,\langle w, v\rangle + \langle w, w\rangle\big) \\ &\qquad\quad - i\big(\langle v, v\rangle + i\,\langle v, w\rangle - i\,\langle w, v\rangle + \langle w, w\rangle\big) \\ &\quad= 4\langle v, w\rangle. \end{aligned}$$

4.1.15. Let $\mathbf{x} = (a_1, \dots, a_n) \in \mathbb{R}^n$ and $\mathbf{y} = (b_1, \dots, b_n) \in \mathbb{R}^n$. Then the result is immediate from the Cauchy–Schwarz inequality with the standard inner product on $\mathbb{R}^n$.

4.1.16. This is immediate from the Cauchy–Schwarz inequality, applied to the L^2 inner product

$$\langle f, g\rangle = \int_a^b f(x)g(x)\,dx$$

on $C([a,b])$.

4.1.17. Define $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ by $v_i = \sqrt{i}a_i$ and $w_i = \frac{b_i}{\sqrt{i}}$. Then by the Cauchy–Schwarz inequality,

$$\left(\sum_{i=1}^n a_i b_i\right)^2 = \left(\sum_{i=1}^n v_i w_i\right)^2 = \langle \mathbf{v}, \mathbf{w}\rangle^2 \le \|\mathbf{v}\|^2 \|\mathbf{w}\|^2 = \left(\sum_{i=1}^n i a_i^2\right)\left(\sum_{i=1}^n \frac{b_i^2}{i}\right).$$

4.1.18. Examining the proof of Theorem 4.7, have equality if and only if $\operatorname{Re}\langle v,w\rangle = |\langle v,w\rangle| = \|v\|\,\|w\|$. The second equality is the equality case of the Cauchy–Schwarz inequality, and holds if and only if v and w are collinear; i.e., there is $a \in \mathbb{F}$ such that either $v = aw$ or $w = 0$. Suppose $v = aw$. Then $\langle v,w\rangle = \langle aw,w\rangle = a\|w\|^2$. We have $\operatorname{Re}\langle v,w\rangle = |\langle v,w\rangle|$ if and only if $\langle v,w\rangle$ is real and nonnegative, which means that we must have a real and nonnegative. That is, equality holds in the triangle inequality if and only if $v = aw$ for $a \geq 0$ or else $w = 0$.

4.1.19. **The inner product of two vectors is a scalar:** Since we are thinking of V as a *real* vector space, this must be interpreted as saying that $\langle v,w\rangle_{\mathbb{R}}$ is a real number. Since the real part of any complex number is a real number, this follows immediately from the definition of $\langle v,w\rangle_{\mathbb{R}}$.

Distributive law: For each $v_1, v_2, v_3 \in V$,

$$\begin{aligned}\langle v_1+v_2, v_3\rangle_{\mathbb{R}} &= \operatorname{Re}\langle v_1+v_2, v_3\rangle \\ &= \operatorname{Re}(\langle v_1,v_3\rangle + \langle v_2,v_3\rangle) \\ &= \operatorname{Re}\langle v_1,v_3\rangle + \operatorname{Re}\langle v_2,v_3\rangle \\ &= \langle v_1,v_3\rangle_{\mathbb{R}} + \langle v_2,v_3\rangle_{\mathbb{R}}\end{aligned}$$

by the linearity of Re and the distributive law for the complex inner product on V.

Homogeneity: This is similar to the distributive law.

Symmetry: For each $v_1, v_2 \in V$,

$$\langle v_1,v_2\rangle_{\mathbb{R}} = \operatorname{Re}\langle v_1,v_2\rangle = \operatorname{Re}\overline{\langle v_2,v_1\rangle} = \operatorname{Re}\langle v_2,v_1\rangle = \langle v_2,v_1\rangle_{\mathbb{R}}$$

by the symmetry of the complex inner product on V.

Nonnegativity: For each $v \in V$,

$$\langle v,v\rangle_{\mathbb{R}} = \operatorname{Re}\langle v,v\rangle \geq 0$$

by the nonnegativity property of the complex inner product on V.

Definiteness: If $\langle v,v\rangle_{\mathbb{R}} = 0$, then $\operatorname{Re}\langle v,v\rangle = 0$. Since $\langle v,v\rangle \in \mathbb{R}$ in any case, this implies that $\langle v,v\rangle = 0$. By the definiteness of the complex inner product on V, this means that $v = 0$.

4.1.20. For each $v \in V$, $\langle 0,v\rangle = \langle 00,v\rangle = 0\,\langle 0,v\rangle = 0$, and $\langle v,0\rangle = \langle v,00\rangle = 0\,\langle v,0\rangle = 0$ by part 2 of Proposition 4.2.

4.1.21. If $w = 0$, then any $v \in V$ is already orthogonal to w and so taking $u = v$ and any choice of $a \in \mathbb{F}$ works.

4.1.22. Expanding the norm-squared,

$$f(t) = \langle v+tw, v+tw\rangle = \|v\|^2 + (2\operatorname{Re}\langle v,w\rangle)t + \|w\|^2 t^2.$$

Assuming that $w \neq 0$ (otherwise the Cauchy–Schwarz inequality holds trivially), then completing the square, we have

$$f(t) = \left(\|w\| t + \frac{\operatorname{Re}\langle v, w\rangle}{\|w\|} \right)^2 + \|v\|^2 - \frac{(\operatorname{Re}\langle v, w\rangle)^2}{\|w\|^2}.$$

Then the minimum value of f, achieved when the first term on the right hand side is 0, is $\|v\|^2 - \frac{(\operatorname{Re}\langle v,w\rangle)^2}{\|w\|^2}$. Since $f(t) \geq 0$ for every t, this implies that

$$\|v\|^2 - \frac{(\operatorname{Re}\langle v, w\rangle)^2}{\|w\|^2} \geq 0,$$

so $(\operatorname{Re}\langle v, w\rangle)^2 \leq \|v\|^2 \|w\|^2$ and therefore $|\operatorname{Re}\langle v, w\rangle| \leq \|v\| \|w\|$.

If $\langle v, w\rangle \in \mathbb{R}$ then we are done. Otherwise, let $\omega = \frac{|\langle v,w\rangle|}{\langle v,w\rangle}$. Then $\langle \omega v, w\rangle = \omega \langle v, w\rangle = |\langle v, w\rangle|$ and $|\omega| = 1$, so the above argument implies that

$$|\langle v, w\rangle| = |\operatorname{Re}\langle \omega v, w\rangle| \leq \|\omega v\| \|w\| = \|v\| \|w\|.$$

4.2 Orthonormal bases

4.2.1. In each part, the list is orthonormal with respect to the standard inner product on $\mathbb{F}^n$, hence linearly independent, and has length n, and is therefore a basis by Theorem 3.28.

4.2.2. In each part, the list is orthonormal (with respect to the standard inner product in parts (a) and (c), and with respect to the Frobenius inner product in part (b)), hence linearly independent. Each linearly independent list has length equal to the dimension of the named space (in part (a), a plane in $\mathbb{R}^3$, and in the other parts the entire space), and is therefore a basis by Theorem 3.28.

4.2.3. (a) $\left\langle \begin{bmatrix} -10 \\ 3 \end{bmatrix}, \frac{1}{\sqrt{13}} \begin{bmatrix} 2 \\ -3 \end{bmatrix} \right\rangle = -\dfrac{29}{\sqrt{13}}$ and $\left\langle \begin{bmatrix} -10 \\ 3 \end{bmatrix}, \frac{1}{\sqrt{13}} \begin{bmatrix} 3 \\ 2 \end{bmatrix} \right\rangle = -\dfrac{24}{\sqrt{13}}$, so $\left[\begin{bmatrix} -10 \\ 3 \end{bmatrix} \right]_{\mathcal{B}} = \begin{bmatrix} -\frac{29}{\sqrt{13}} \\ -\frac{24}{\sqrt{13}} \end{bmatrix}$.

(b) $\left\langle \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}, \frac{1}{\sqrt{30}} \begin{bmatrix} 1 \\ 2 \\ -5 \end{bmatrix} \right\rangle = \dfrac{2}{\sqrt{30}}$, $\left\langle \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}, \frac{1}{2\sqrt{5}} \begin{bmatrix} 4 \\ -2 \\ 0 \end{bmatrix} \right\rangle = \dfrac{4}{\sqrt{5}}$, and $\left\langle \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}, \frac{1}{2\sqrt{6}} \begin{bmatrix} 2 \\ 4 \\ 2 \end{bmatrix} \right\rangle = \dfrac{8}{\sqrt{6}}$, so $\left[\begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix} \right]_{\mathcal{B}} = \begin{bmatrix} \frac{2}{\sqrt{30}} \\ \frac{4}{\sqrt{5}} \\ \frac{8}{\sqrt{6}} \end{bmatrix}$.

(c) $\left\langle \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\rangle = \dfrac{2}{\sqrt{3}}$, $\left\langle \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ e^{2\pi i/3} \\ e^{4\pi i/3} \end{bmatrix} \right\rangle = \dfrac{1 + e^{2\pi i/3}}{\sqrt{3}}$, and $\left\langle \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ e^{4\pi i/3} \\ e^{2\pi i/3} \end{bmatrix} \right\rangle = \dfrac{1 + e^{4\pi i/3}}{\sqrt{3}}$, so $\left[\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \right]_{\mathcal{B}} = \begin{bmatrix} \frac{2}{\sqrt{3}} \\ \frac{1+e^{4\pi i/3}}{\sqrt{3}} \\ \frac{1+e^{2\pi i/3}}{\sqrt{3}} \end{bmatrix}$.

(d) $\left\langle \begin{bmatrix} -2 \\ 1 \\ 0 \\ 4 \end{bmatrix}, \frac{1}{\sqrt{3}} \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix} \right\rangle = \frac{5}{\sqrt{3}}$, $\left\langle \begin{bmatrix} -2 \\ 1 \\ 0 \\ 4 \end{bmatrix}, \frac{1}{\sqrt{15}} \begin{bmatrix} 3 \\ -2 \\ 1 \\ 1 \end{bmatrix} \right\rangle = -\frac{4}{\sqrt{3}}$, $\left\langle \begin{bmatrix} -2 \\ 1 \\ 0 \\ 4 \end{bmatrix}, \frac{1}{\sqrt{35}} \begin{bmatrix} 3 \\ 3 \\ -4 \\ 1 \end{bmatrix} \right\rangle =$ $\frac{1}{\sqrt{35}}$, and $\left\langle \begin{bmatrix} -2 \\ 1 \\ 0 \\ 4 \end{bmatrix}, \frac{1}{\sqrt{7}} \begin{bmatrix} 1 \\ 1 \\ 1 \\ -2 \end{bmatrix} \right\rangle = -\frac{9}{\sqrt{3}}$, so $\left[\begin{bmatrix} -2 \\ 1 \\ 0 \\ 4 \end{bmatrix} \right]_{\mathcal{B}} = \begin{bmatrix} \frac{5}{\sqrt{3}} \\ -\frac{4}{\sqrt{15}} \\ \frac{1}{\sqrt{35}} \\ -\frac{9}{\sqrt{7}} \end{bmatrix}$.

4.2.4. (a) $\left\langle \begin{bmatrix} 1 \\ 2 \\ -3 \end{bmatrix}, \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \right\rangle = 2\sqrt{2}$ and $\left\langle \begin{bmatrix} 1 \\ 2 \\ -3 \end{bmatrix}, \frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} \right\rangle = -\sqrt{6}$, so $\left[\begin{bmatrix} 1 \\ 2 \\ -3 \end{bmatrix} \right]_{\mathcal{B}} = \begin{bmatrix} 2\sqrt{2} \\ -\sqrt{6} \end{bmatrix}$.

(b) $\left\langle \begin{bmatrix} 5 & 7 \\ 7 & 2 \end{bmatrix}, \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \right\rangle = \frac{21}{2}$, $\left\langle \begin{bmatrix} 5 & 7 \\ 7 & 2 \end{bmatrix}, \frac{1}{2} \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix} \right\rangle = \frac{3}{2}$, $\left\langle \begin{bmatrix} 5 & 7 \\ 7 & 2 \end{bmatrix}, \frac{1}{2} \begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix} \right\rangle =$ $\frac{3}{2}$, and $\left\langle \begin{bmatrix} 5 & 7 \\ 7 & 2 \end{bmatrix}, \frac{1}{2} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \right\rangle = -\frac{7}{2}$, so $\left[\begin{bmatrix} 5 & 7 \\ 7 & 2 \end{bmatrix} \right]_{\mathcal{B}} = \frac{1}{2} \begin{bmatrix} 21 \\ 3 \\ 3 \\ -7 \end{bmatrix}$.

(c) $\left\langle \begin{bmatrix} -3 \\ 0 \\ 1 \\ 2 \end{bmatrix}, \frac{1}{2} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \right\rangle = 0$, $\left\langle \begin{bmatrix} -3 \\ 0 \\ 1 \\ 2 \end{bmatrix}, \frac{1}{2} \begin{bmatrix} 1 \\ i \\ -1 \\ -i \end{bmatrix} \right\rangle = -2 + i$, $\left\langle \begin{bmatrix} -3 \\ 0 \\ 1 \\ 2 \end{bmatrix}, \frac{1}{2} \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \end{bmatrix} \right\rangle = -2$, and $\left\langle \begin{bmatrix} -3 \\ 0 \\ 1 \\ 2 \end{bmatrix}, \frac{1}{2} \begin{bmatrix} 1 \\ -i \\ -1 \\ i \end{bmatrix} \right\rangle = -2 - i$, so $\left[\begin{bmatrix} -3 \\ 0 \\ 1 \\ 2 \end{bmatrix} \right]_{\mathcal{B}} = \begin{bmatrix} 0 \\ -2+i \\ -2 \\ -2-i \end{bmatrix}$.

4.2.5. (a) Write $\mathcal{B} = (\mathbf{v}_1, \mathbf{v}_2)$. Since $\mathbf{v}_2$ lies on the line and $\mathbf{v}_1$ is orthogonal to it, $\boldsymbol{T}\mathbf{v}_1 = -\mathbf{v}_1$ and $\boldsymbol{T}\mathbf{v}_2 = \mathbf{v}_2$, so $[\boldsymbol{T}]_{\mathcal{B}} = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$.

(b) Write $\mathcal{B} = (\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)$. Then $\boldsymbol{T}\mathbf{v}_1 = \frac{1}{\sqrt{30}} \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}$, $\boldsymbol{T}\mathbf{v}_2 = \frac{1}{2\sqrt{5}} \begin{bmatrix} 4 \\ -2 \\ 0 \end{bmatrix} = \mathbf{v}_2$, $\boldsymbol{T}\mathbf{v}_3 =$ $\frac{1}{2\sqrt{6}} \begin{bmatrix} 2 \\ 4 \\ 0 \end{bmatrix}$, and

$$[\boldsymbol{T}]_{\mathcal{B}} = \begin{bmatrix} \langle \boldsymbol{T}\mathbf{v}_1, \mathbf{v}_1 \rangle & \langle \boldsymbol{T}\mathbf{v}_2, \mathbf{v}_1 \rangle & \langle \boldsymbol{T}\mathbf{v}_3, \mathbf{v}_1 \rangle \\ \langle \boldsymbol{T}\mathbf{v}_1, \mathbf{v}_2 \rangle & \langle \boldsymbol{T}\mathbf{v}_2, \mathbf{v}_2 \rangle & \langle \boldsymbol{T}\mathbf{v}_3, \mathbf{v}_2 \rangle \\ \langle \boldsymbol{T}\mathbf{v}_1, \mathbf{v}_3 \rangle & \langle \boldsymbol{T}\mathbf{v}_2, \mathbf{v}_3 \rangle & \langle \boldsymbol{T}\mathbf{v}_3, \mathbf{v}_3 \rangle \end{bmatrix} = \begin{bmatrix} \frac{1}{6} & 0 & \frac{\sqrt{5}}{6} \\ 0 & 1 & 0 \\ \frac{\sqrt{5}}{6} & 0 & \frac{5}{6} \end{bmatrix}.$$

(c) Write $\mathcal{B} = (\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)$. Then $\boldsymbol{T}\mathbf{v}_1 = \mathbf{v}_1$, $\boldsymbol{T}\mathbf{v}_2 = \frac{1}{\sqrt{3}}\begin{bmatrix} e^{2\pi i/3} \\ e^{4\pi i/3} \\ 1 \end{bmatrix} = e^{2\pi i/3}\mathbf{v}_2$, $\boldsymbol{T}\mathbf{v}_3 = \frac{1}{\sqrt{3}}\begin{bmatrix} e^{4\pi i/3} \\ 1 \\ e^{2\pi i/3} \end{bmatrix} = e^{4\pi i/3}\mathbf{v}_3$, so $[\boldsymbol{T}]_{\mathcal{B}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & e^{2\pi i/3} & 0 \\ 0 & 0 & e^{4\pi i/3} \end{bmatrix}$.

(d) Write $\mathcal{B} = (\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4)$. Then $\boldsymbol{T}\mathbf{v}_1 = \frac{1}{\sqrt{3}}\begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \end{bmatrix}$, $\boldsymbol{T}\mathbf{v}_2 = \frac{1}{\sqrt{15}}\begin{bmatrix} 1 \\ 1 \\ 3 \\ -2 \end{bmatrix}$, $\boldsymbol{T}\mathbf{v}_3 = \frac{1}{\sqrt{35}}\begin{bmatrix} -4 \\ 1 \\ 3 \\ 3 \end{bmatrix}$, $\boldsymbol{T}\mathbf{v}_4 = \frac{1}{\sqrt{7}}\begin{bmatrix} 1 \\ -2 \\ 1 \\ 1 \end{bmatrix}$, and

$$[\boldsymbol{T}]_{\mathcal{B}} = \begin{bmatrix} \langle \boldsymbol{T}\mathbf{v}_1, \mathbf{v}_1\rangle & \langle \boldsymbol{T}\mathbf{v}_2, \mathbf{v}_1\rangle & \langle \boldsymbol{T}\mathbf{v}_3, \mathbf{v}_1\rangle & \langle \boldsymbol{T}\mathbf{v}_4, \mathbf{v}_1\rangle \\ \langle \boldsymbol{T}\mathbf{v}_1, \mathbf{v}_2\rangle & \langle \boldsymbol{T}\mathbf{v}_2, \mathbf{v}_2\rangle & \langle \boldsymbol{T}\mathbf{v}_3, \mathbf{v}_2\rangle & \langle \boldsymbol{T}\mathbf{v}_4, \mathbf{v}_2\rangle \\ \langle \boldsymbol{T}\mathbf{v}_1, \mathbf{v}_3\rangle & \langle \boldsymbol{T}\mathbf{v}_2, \mathbf{v}_3\rangle & \langle \boldsymbol{T}\mathbf{v}_3, \mathbf{v}_3\rangle & \langle \boldsymbol{T}\mathbf{v}_4, \mathbf{v}_3\rangle \\ \langle \boldsymbol{T}\mathbf{v}_1, \mathbf{v}_4\rangle & \langle \boldsymbol{T}\mathbf{v}_2, \mathbf{v}_4\rangle & \langle \boldsymbol{T}\mathbf{v}_3, \mathbf{v}_4\rangle & \langle \boldsymbol{T}\mathbf{v}_4, \mathbf{v}_4\rangle \end{bmatrix} = \begin{bmatrix} \frac{2}{3} & \frac{2}{3\sqrt{5}} & \frac{\sqrt{7}}{\sqrt{15}} & 0 \\ \frac{2}{3\sqrt{5}} & \frac{2}{15} & -\frac{8}{5\sqrt{21}} & \frac{9}{\sqrt{105}} \\ \frac{\sqrt{7}}{\sqrt{15}} & -\frac{8}{5\sqrt{21}} & -\frac{18}{35} & -\frac{6}{7\sqrt{5}} \\ 0 & \frac{9}{\sqrt{105}} & -\frac{6}{7\sqrt{5}} & -\frac{2}{7} \end{bmatrix}.$$

4.2.6. (a) Write $\mathcal{B} = (\mathbf{v}_1, \mathbf{v}_2)$. Then $\boldsymbol{T}\mathbf{v}_1 = \frac{1}{\sqrt{2}}\begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}$, $\boldsymbol{T}\mathbf{v}_2 = \frac{1}{\sqrt{6}}\begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}$, and

$$[\boldsymbol{T}]_{\mathcal{B}} = \begin{bmatrix} \langle \boldsymbol{T}\mathbf{v}_1, \mathbf{v}_1\rangle & \langle \boldsymbol{T}\mathbf{v}_2, \mathbf{v}_1\rangle \\ \langle \boldsymbol{T}\mathbf{v}_1, \mathbf{v}_2\rangle & \langle \boldsymbol{T}\mathbf{v}_2, \mathbf{v}_2\rangle \end{bmatrix} = \begin{bmatrix} -\frac{1}{2} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{bmatrix}.$$

(b) Write $\mathcal{B} = (\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3, \mathbf{A}_4)$. Then $\boldsymbol{T}\mathbf{A}_1 = \mathbf{A}_1$, $\boldsymbol{T}\mathbf{A}_2 = \mathbf{A}_3$, $\boldsymbol{T}\mathbf{A}_3 = \mathbf{A}_2$, $\boldsymbol{T}\mathbf{A}_4 = \mathbf{A}_4$, so

$$[\boldsymbol{T}]_{\mathcal{B}} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

(c) Write $\mathcal{B} = (\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4)$. Then $\boldsymbol{T}\mathbf{v}_1 = \mathbf{v}_1$, $\boldsymbol{T}\mathbf{v}_2 = \frac{1}{2}\begin{bmatrix} i \\ 1 \\ -i \\ -1 \end{bmatrix} = i\mathbf{v}_4$, $\boldsymbol{T}\mathbf{v}_3 =$

$\frac{1}{2}\begin{bmatrix}-1\\1\\-1\\1\end{bmatrix} = -\mathbf{v}_3$, $\boldsymbol{T}\mathbf{v}_4 = \frac{1}{2}\begin{bmatrix}-i\\1\\i\\-1\end{bmatrix} = -i\mathbf{v}_2$, so

$$[\boldsymbol{T}]_{\mathcal{B}} = \begin{bmatrix}1 & 0 & 0 & 0\\ 0 & 0 & 0 & -i\\ 0 & 0 & -1 & 0\\ 0 & i & 0 & 0\end{bmatrix}.$$

4.2.7. (a) Write

$$\mathbf{v}_1 = \frac{1}{\sqrt{2}}\begin{bmatrix}1\\-1\\0\end{bmatrix} \qquad \text{and} \qquad \mathbf{v}_2 = \frac{1}{\sqrt{6}}\begin{bmatrix}1\\1\\-2\end{bmatrix}.$$

It is straightforward to check that $\langle \mathbf{v}_1, \mathbf{v}_1\rangle = \langle \mathbf{v}_2, \mathbf{v}_2\rangle = 0$ and $\langle \mathbf{v}_1, \mathbf{v}_2\rangle = 0$, and that $\mathbf{v}_1, \mathbf{v}_2 \in U$. Thus $(\mathbf{v}_1, \mathbf{v}_2)$ is an orthonormal list, hence linearly independent and thus a basis of its span. Since $U \neq \mathbb{R}^3$ (since, e.g., $\mathbf{e}_1 \notin U$), $\dim U < 3$. Therefore $(\mathbf{v}_1, \mathbf{v}_2)$ must span U, so it is an orthonormal basis of U.

(b) $\left\langle \begin{bmatrix}1\\2\\-3\end{bmatrix}, \mathbf{v}_1\right\rangle = -\frac{1}{\sqrt{2}}$ and $\left\langle \begin{bmatrix}1\\2\\-3\end{bmatrix}, \mathbf{v}_1\right\rangle = \frac{9}{\sqrt{6}}$, so $\left[\begin{bmatrix}1\\2\\-3\end{bmatrix}\right]_{\mathcal{B}} = \frac{1}{\sqrt{2}}\begin{bmatrix}-1\\3\sqrt{3}\end{bmatrix}$.

(c) We compute

$$\begin{aligned}
\langle \boldsymbol{T}\mathbf{v}_1, \mathbf{v}_1\rangle &= \left\langle \frac{1}{\sqrt{2}}\begin{bmatrix}-1\\0\\1\end{bmatrix}, \frac{1}{\sqrt{2}}\begin{bmatrix}1\\-1\\0\end{bmatrix}\right\rangle = -\frac{1}{2},\\
\langle \boldsymbol{T}\mathbf{v}_2, \mathbf{v}_1\rangle &= \left\langle \frac{1}{\sqrt{6}}\begin{bmatrix}1\\-2\\1\end{bmatrix}, \frac{1}{\sqrt{2}}\begin{bmatrix}1\\-1\\0\end{bmatrix}\right\rangle = \frac{3}{\sqrt{12}} = \frac{\sqrt{3}}{2},\\
\langle \boldsymbol{T}\mathbf{v}_1, \mathbf{v}_2\rangle &= \left\langle \frac{1}{\sqrt{2}}\begin{bmatrix}-1\\0\\1\end{bmatrix}, \frac{1}{\sqrt{6}}\begin{bmatrix}1\\1\\-2\end{bmatrix}\right\rangle = -\frac{3}{\sqrt{12}} = -\frac{\sqrt{3}}{2},\\
\langle \boldsymbol{T}\mathbf{v}_2, \mathbf{v}_2\rangle &= \left\langle \frac{1}{\sqrt{6}}\begin{bmatrix}1\\-2\\1\end{bmatrix}, \frac{1}{\sqrt{6}}\begin{bmatrix}1\\1\\-2\end{bmatrix}\right\rangle = -\frac{1}{2}.
\end{aligned}$$

Therefore

$$[\boldsymbol{T}]_{\mathcal{B}} = \begin{bmatrix}\langle \boldsymbol{T}\mathbf{v}_1, \mathbf{v}_1\rangle & \langle \boldsymbol{T}\mathbf{v}_2, \mathbf{v}_1\rangle\\ \langle \boldsymbol{T}\mathbf{v}_1, \mathbf{v}_2\rangle & \langle \boldsymbol{T}\mathbf{v}_2, \mathbf{v}_2\rangle\end{bmatrix} = \frac{1}{2}\begin{bmatrix}-1 & \sqrt{3}\\ -\sqrt{3} & -1\end{bmatrix}.$$

4.2.8. From Example 3 on pages 239–240, the set $\left\{\frac{1}{2\pi}, \frac{1}{\sqrt{\pi}}\sin(n\theta), \frac{1}{\sqrt{\pi}}\cos(n\theta) : n \in \mathbb{N}\right\}$ is orthonormal in $C_{2\pi}([0,2\pi])$, and so

$$\begin{aligned}
&\left\|\sum_{k=1}^{n} a_k \sin(k\theta) + b_0 + \sum_{\ell=1}^{m} b_\ell \cos(\ell\theta)\right\| \\
&\qquad = \sqrt{\sum_{k=1}^{n} |a_k|^2 \left\|\sin(k\theta)\right\|^2 + |b_0|^2 \left\|1\right\|^2 + \sum_{\ell=1}^{m} |b_\ell|^2 \left\|\cos(\ell\theta)\right\|^2} \\
&\qquad = \sqrt{\pi \sum_{k=1}^{n} |a_k|^2 + |b_0|^2 (2\pi) + \pi \sum_{\ell=1}^{m} |b_\ell|^2},
\end{aligned}$$

where the first equality is by the Pythagorean theorem and the second is by the calculated norms of the functions in question.

4.2.9. Firstly, $\boldsymbol{D}[1] = 0$, so the first column of $[\boldsymbol{D}]_{\mathcal{B}}$ is all zeroes. Next, $\boldsymbol{D}[2\sqrt{3}x - \sqrt{3}] = 2\sqrt{3}$, so the second column is $\begin{bmatrix} 2\sqrt{3} \\ 0 \\ 0 \end{bmatrix}$. Similarly, $\boldsymbol{D}[6\sqrt{5}x^2 - 6\sqrt{5}x + \sqrt{5}] = 12\sqrt{5}x - 6\sqrt{5} = 2\sqrt{15}(2\sqrt{3}x - \sqrt{3})$, which means the third column is $\begin{bmatrix} 0 \\ 2\sqrt{15} \\ 0 \end{bmatrix}$, and so $[\boldsymbol{D}]_{\mathcal{B}} = \begin{bmatrix} 0 & 2\sqrt{3} & 0 \\ 0 & 0 & 2\sqrt{15} \\ 0 & 0 & 0 \end{bmatrix}$.

4.2.10. (a) By definition,

$$\left\|3 - 2x + x^2\right\|^2 = \int_0^1 (3-2x+x^2)^2 \, dx = \int_0^1 (9-12x+10x^2-4x^3+x^4) \, dx = \frac{83}{15}.$$

(b) The basis $(1, x, x^2)$ is not orthonormal (or even orthogonal), so Theorem 4.10 does not apply here.

4.2.11. (a)

- $\mathbf{f}_1 = \frac{1}{\|\mathbf{v}_1\|}\mathbf{v}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$
- $\widetilde{\mathbf{f}}_2 = \mathbf{v}_2 - \langle \mathbf{v}_2, \mathbf{f}_1 \rangle \, \mathbf{f}_1 = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} - \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$
- $\mathbf{f}_2 = \frac{1}{\|\widetilde{\mathbf{f}}_2\|}\widetilde{\mathbf{f}}_2 = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$
- $\widetilde{\mathbf{f}}_3 = \mathbf{v}_3 - \langle \mathbf{v}_3, \mathbf{f}_1 \rangle \, \mathbf{f}_1 - \langle \mathbf{v}_3, \mathbf{f}_2 \rangle = \begin{bmatrix} 0 \\ -1 \\ 2 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} - \frac{1}{3} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 2 \\ -4 \\ 2 \end{bmatrix}$

- $\mathbf{f}_3 = \frac{1}{\|\widetilde{\mathbf{f}}_3\|}\widetilde{\mathbf{f}}_3 = \frac{1}{\sqrt{6}}\begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$

(b) By the same method, we obtain $\left(\frac{1}{\sqrt{5}}\begin{bmatrix} 0 \\ -1 \\ 2 \end{bmatrix}, \frac{1}{\sqrt{30}}\begin{bmatrix} 5 \\ -2 \\ -1 \end{bmatrix}, \frac{1}{\sqrt{6}}\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}\right)$.

(c) By the same method, we obtain $\left(\frac{1}{\sqrt{3}}\begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \frac{1}{\sqrt{15}}\begin{bmatrix} 3 \\ -2 \\ 1 \\ 1 \end{bmatrix}, \frac{1}{\sqrt{35}}\begin{bmatrix} 3 \\ 3 \\ -4 \\ 1 \end{bmatrix}, \frac{1}{\sqrt{7}}\begin{bmatrix} 1 \\ 1 \\ 1 \\ -2 \end{bmatrix}\right)$

4.2.12. (a)

- $\mathbf{f}_1 = \frac{1}{\|\mathbf{v}_1\|}\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$
- $\widetilde{\mathbf{f}}_2 = \mathbf{v}_2 - \langle \mathbf{v}_2, \mathbf{f}_1 \rangle \mathbf{f}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} - \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$
- $\mathbf{f}_2 = \frac{1}{\|\widetilde{\mathbf{f}}_2\|}\widetilde{\mathbf{f}}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$
- $\widetilde{\mathbf{f}}_3 = \mathbf{v}_3 - \langle \mathbf{v}_3, \mathbf{f}_1 \rangle \mathbf{f}_1 - \langle \mathbf{v}_3, \mathbf{f}_2 \rangle = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$
- $\mathbf{f}_3 = \frac{1}{\|\widetilde{\mathbf{f}}_3\|}\widetilde{\mathbf{f}}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$

(b)

- $\mathbf{f}_1 = \frac{1}{\|\mathbf{v}_1\|}\mathbf{v}_1 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ i \\ 0 \end{bmatrix}$
- $\widetilde{\mathbf{f}}_2 = \mathbf{v}_2 - \langle \mathbf{v}_2, \mathbf{f}_1 \rangle \mathbf{f}_1 = \begin{bmatrix} 0 \\ 1 \\ i \end{bmatrix} + \frac{i}{2}\begin{bmatrix} 1 \\ i \\ 0 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} i \\ 1 \\ 2i \end{bmatrix}$
- $\mathbf{f}_2 = \frac{1}{\|\widetilde{\mathbf{f}}_2\|}\widetilde{\mathbf{f}}_2 = \frac{1}{\sqrt{6}}\begin{bmatrix} i \\ 1 \\ 2i \end{bmatrix}$
- $\widetilde{\mathbf{f}}_3 = \mathbf{v}_3 - \langle \mathbf{v}_3, \mathbf{f}_1 \rangle \mathbf{f}_1 - \langle \mathbf{v}_3, \mathbf{f}_2 \rangle = \begin{bmatrix} i \\ 0 \\ 1 \end{bmatrix} - \frac{i}{2}\begin{bmatrix} 1 \\ i \\ 0 \end{bmatrix} - \frac{1-2i}{6}\begin{bmatrix} i \\ 1 \\ 2i \end{bmatrix} = \frac{1}{3}\begin{bmatrix} -1+i \\ 1+i \\ 1-i \end{bmatrix}$
- $\mathbf{f}_3 = \frac{1}{\|\widetilde{\mathbf{f}}_3\|}\widetilde{\mathbf{f}}_3 = \frac{1}{\sqrt{6}}\begin{bmatrix} -1+i \\ 1+i \\ 1-i \end{bmatrix}$

(c) By the same method, we obtain $\left(\frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 0 \\ 0 \\ -1 \end{bmatrix}, \frac{1}{\sqrt{6}} \begin{bmatrix} -1 \\ 2 \\ 0 \\ -1 \end{bmatrix}, \frac{1}{2\sqrt{3}} \begin{bmatrix} -1 \\ -1 \\ 3 \\ -1 \end{bmatrix} \right)$.

4.2.13. (a) We can successfully carry out the Gram–Schmidt process to obtain

$$\left(\frac{1}{\sqrt{14}} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \frac{1}{5\sqrt{42}} \begin{bmatrix} 17 \\ 20 \\ -19 \end{bmatrix}, \frac{1}{5\sqrt{3}} \begin{bmatrix} 7 \\ -5 \\ 1 \end{bmatrix} \right),$$

so the list is linearly independent.

(b) We can successfully carry out the Gram–Schmidt process to obtain

$$\left(\frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}, \frac{1}{\sqrt{6}} \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix}, \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \right),$$

so the list is linearly independent.

(c) The Gram–Schmidt process produces the first two orthonormal vectors $\mathbf{f}_1 = \frac{1}{\sqrt{6}} \begin{bmatrix} -2 \\ -1 \\ 1 \end{bmatrix}$ and $\mathbf{f}_2 = \frac{1}{\sqrt{3}} \begin{bmatrix} -1 \\ 1 \\ -1 \end{bmatrix}$. At the next step we find

$$\widetilde{\mathbf{f}}_3 = \begin{bmatrix} 1 \\ 2 \\ -2 \end{bmatrix} + \begin{bmatrix} -2 \\ -1 \\ 1 \end{bmatrix} - \begin{bmatrix} -1 \\ 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix},$$

so it is impossible to complete the process. Therefore the original list must be linearly dependent.

4.2.14. (a) The Gram–Schmidt process produces the first two orthonormal vectors $\mathbf{f}_1 = \frac{1}{\sqrt{14}} \begin{bmatrix} 2 \\ 1 \\ -3 \end{bmatrix}$ and $\mathbf{f}_2 = \frac{1}{\sqrt{42}} \begin{bmatrix} -4 \\ 5 \\ -1 \end{bmatrix}$. At the next step we find

$$\widetilde{\mathbf{f}}_3 = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix} - \frac{2}{7} \begin{bmatrix} 2 \\ 1 \\ -3 \end{bmatrix} - \frac{1}{7} \begin{bmatrix} -4 \\ 5 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix},$$

so it is impossible to complete the process. Therefore the original list must be linearly dependent.

(b) We can successfully carry out the Gram–Schmidt process to obtain

$$\left(\frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}, \frac{1}{\sqrt{6}} \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix}, \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \right),$$

so the list is linearly independent.

(c) The Gram–Schmidt process produces the first two orthonormal vectors $\mathbf{f}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$ and $\mathbf{f}_2 = \frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}$. At the next step we find

$$\widetilde{\mathbf{f}}_3 = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix},$$

so it is impossible to complete the process. Therefore the original list must be linearly dependent.

4.2.15. Write $v_k(x) = x^{k-1}$ for $k = 1, 2, 3$.

(a)
- $\|v_1\| = \sqrt{\int_0^1 1 \, dx} = 1$.
- $e_1(x) = 1$.
- $\langle v_2, e_1 \rangle = \int_0^1 x \, dx = \frac{1}{2}$.
- $\tilde{e}_2(x) = v_2(x) - \langle v_2, e_1 \rangle \, e_1(x) = x - \frac{1}{2}$.
- $\|\tilde{e}_2\| = \sqrt{\int_0^1 \left(x - \frac{1}{2}\right)^2 \, dx} = \frac{1}{2\sqrt{3}}$.
- $e_2(x) = \frac{1}{\|\tilde{e}_2\|} \tilde{e}_2(x) = 2\sqrt{3}x - \sqrt{3}$.
- $\langle v_3, e_1 \rangle = \int_0^1 x^2 \, dx = \frac{1}{3}$.
- $\langle v_3, e_2 \rangle = \int_0^1 (2\sqrt{3}x - \sqrt{3}) x \, dx = \frac{1}{2\sqrt{3}}$.
- $\tilde{e}_3(x) = v_3(x) - \langle v_3, e_1 \rangle \, e_1(x) - \langle v_3, e_2 \rangle \, (x) = x^2 - x + \frac{1}{6}$.
- $\|\tilde{e}_3\| = \sqrt{\int_0^1 \left(x^2 - x + \frac{1}{6}\right)^2 \, dx} = \frac{1}{6\sqrt{5}}$.
- $e_3(x) = \frac{1}{\|\tilde{e}_3\|} \tilde{e}_3(x) = 6\sqrt{5}x^2 - 6\sqrt{5}x + \sqrt{5}$.
- $\langle v_4, e_1 \rangle = \int_0^1 x^3 \, dx = \frac{1}{4}$.
- $\langle v_4, e_2 \rangle = \int_0^1 x^3 (2\sqrt{3}x - \sqrt{3}) \, dx = \frac{3\sqrt{3}}{20}$.
- $\langle v_4, e_3 \rangle = \int_0^1 x^3 (6\sqrt{5}x^2 - 6\sqrt{5}x + \sqrt{5}) \, dx = \frac{1}{4\sqrt{5}}$.
- $\tilde{e}_4(x) = v_4(x) - \langle v_4, e_1 \rangle \, e_1(x) - \langle v_4, e_2 \rangle \, e_2(x) - \langle v_4, e_3 \rangle \, e_3(x) = x^3 - \frac{1}{4} - \frac{3}{2}x + \frac{3}{4} - \frac{3}{2}x^2 + \frac{3}{2}x + \frac{1}{4} = x^3 - \frac{3}{2}x^2 + \frac{3}{5}x - \frac{1}{20}$.
- $\|\tilde{e}_4\| = \sqrt{\int_{-1}^1 (x^3 - \frac{3}{5}x)^2 \, dx} = \frac{1}{20\sqrt{7}}$.
- $e_4(x) = \frac{1}{\|\tilde{e}_4\|} e_4(x) = 20\sqrt{7}x^3 - 30\sqrt{7}x^2 + 12\sqrt{7}x - \sqrt{7}$.

(b)
- $\|v_1\| = \sqrt{\int_{-1}^{1} 1\, dx} = \sqrt{2}$.
- $e_1(x) = \frac{1}{\sqrt{2}}$.
- $\langle v_2, e_1 \rangle = \frac{1}{\sqrt{2}} \int_{-1}^{1} x\, dx = 0$.
- $\tilde{e}_2(x) = v_2(x) - \langle v_2, e_1 \rangle\, e_1(x) = v_2(x) = x$.
- $\|\tilde{e}_2\| = \sqrt{\int_{-1}^{1} x^2\, dx} = \sqrt{\frac{2}{3}}$.
- $e_2(x) = \frac{1}{\|\tilde{e}_2\|} \tilde{e}_2(x) = \sqrt{\frac{3}{2}} x$.
- $\langle v_3, e_1 \rangle = \frac{1}{\sqrt{2}} \int_{-1}^{1} x^2\, dx = \frac{\sqrt{2}}{3}$.
- $\langle v_3, e_2 \rangle = \sqrt{\frac{3}{2}} \int_{-1}^{1} x^3\, dx = 0$.
- $\tilde{e}_3(x) = v_3(x) - \langle v_3, e_1 \rangle\, e_1(x) - \langle v_3, e_2 \rangle\, (x) = x^2 - \frac{1}{3}$.
- $\|\tilde{e}_3\| = \sqrt{\int_{-1}^{1} (x^2 - \frac{1}{3})^2\, dx} = \frac{2}{3}\sqrt{\frac{2}{5}}$.
- $e_3(x) = \frac{1}{\|\tilde{e}_3\|} \tilde{e}_3(x) = \frac{3\sqrt{5}}{2\sqrt{2}} x^2 - \frac{\sqrt{5}}{2\sqrt{2}}$.
- $\langle v_4, e_1 \rangle = \langle v_4, e_3 \rangle = 0$ (this only involves integrating odd powers of x).
- $\langle v_4, e_2 \rangle = \sqrt{\frac{3}{2}} \int_{-1}^{1} x^4\, dx = \frac{\sqrt{6}}{5}$.
- $\tilde{e}_4(x) = v_4(x) - \langle v_4, e_2 \rangle\, e_2(x) = x^3 - \frac{3}{5} x$.
- $\|\tilde{e}_4\| = \sqrt{\int_{-1}^{1} (x^3 - \frac{3}{5} x)^2\, dx} = \frac{2\sqrt{2}}{5\sqrt{7}}$.
- $e_4(x) = \frac{1}{\|\tilde{e}_4\|} e_4(x) = \frac{5\sqrt{7}}{2\sqrt{2}} x^3 - \frac{3\sqrt{7}}{2\sqrt{2}}$.

(c)
- $\|v_1\| = \sqrt{\int_{-1}^{1} x^2\, dx} = \sqrt{\frac{2}{3}}$.
- $e_1(x) = \sqrt{\frac{3}{2}}$.
- $\langle v_2, e_1 \rangle = \sqrt{\frac{3}{2}} \int_{-1}^{1} x^3\, dx = 0$.
- $\tilde{e}_2(x) = v_2(x) - \langle v_2, e_1 \rangle\, e_1(x) = v_2(x) = x$.
- $\|\tilde{e}_2\| = \sqrt{\int_{-1}^{1} x^4\, dx} = \sqrt{\frac{2}{5}}$.
- $e_2(x) = \frac{1}{\|\tilde{e}_2\|} \tilde{e}_2(x) = \sqrt{\frac{5}{2}} x$.
- $\langle v_3, e_1 \rangle = \sqrt{\frac{3}{2}} \int_{-1}^{1} x^4\, dx = \frac{\sqrt{6}}{5}$.
- $\langle v_3, e_2 \rangle = \sqrt{\frac{5}{2}} \int_{-1}^{1} x^5\, dx = 0$.
- $\tilde{e}_3(x) = v_3(x) - \langle v_3, e_1 \rangle\, e_1(x) - \langle v_3, e_2 \rangle\, (x) = x^2 - \frac{3}{5}$.
- $\|\tilde{e}_3\| = \sqrt{\int_{-1}^{1} \left(x^2 - \frac{3}{5}\right)^2 x^2\, dx} = \frac{2\sqrt{2}}{5\sqrt{7}}$.
- $e_3(x) = \frac{1}{\|\tilde{e}_3\|} \tilde{e}_3(x) = \frac{5\sqrt{7}}{2\sqrt{2}} \left(x^2 - \frac{3}{5}\right)$.
- $\langle v_4, e_1 \rangle = \langle v_4, e_3 \rangle = 0$ (this only involves integrating odd powers of x).
- $\langle v_4, e_2 \rangle = \sqrt{\frac{5}{2}} \int_{-1}^{1} x^6\, dx = \frac{\sqrt{10}}{7}$.
- $\tilde{e}_4(x) = v_4(x) - \langle v_4, e_2 \rangle\, e_2(x) = x^3 - \frac{5}{7} x$.

- $\|\tilde{e}_4\| = \sqrt{\int_{-1}^{1}\left(x^3 - \frac{5}{7}x\right)^2 x^2\,dx} = \frac{2\sqrt{2}}{21}$.
- $e_4(x) = \frac{1}{\|\tilde{e}_4\|}e_4(x) = \frac{21}{2\sqrt{2}}x^3 - \frac{15}{2\sqrt{2}}$.

4.2.16. **Solution 1:** Start by picking any two noncollinear vectors in U, for example

$$(\mathbf{v}_1, \mathbf{v}_2) = \left(\begin{bmatrix}2\\1\\0\end{bmatrix}, \begin{bmatrix}3\\0\\-1\end{bmatrix}\right),$$

and a third vector not in U, say $\mathbf{v}_3 = \mathbf{e}_1$. Since $\mathbf{e}_1 \notin U$, the list $(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)$ is a basis of $\mathbb{R}^3$, so we can apply the Gram–Schmidt process to this basis. For the choice of vectors above we obtain

$$(\mathbf{f}_1, \mathbf{f}_2) = \left(\frac{1}{\sqrt{5}}\begin{bmatrix}2\\1\\0\end{bmatrix}, \frac{1}{\sqrt{70}}\begin{bmatrix}3\\-6\\-5\end{bmatrix}, \frac{1}{\sqrt{14}}\begin{bmatrix}1\\-2\\3\end{bmatrix}\right).$$

Since U is two-dimensional, $\langle \mathbf{f}_1, \mathbf{f}_2\rangle = \langle \mathbf{v}_1, \mathbf{v}_2\rangle = U$, as desired.

Solution 2: As above, start by picking a basis of U and performing the Gram–Schmidt process to obtain an orthonormal basis $(\mathbf{f}_1, \mathbf{f}_2)$ of U. Then since

$$U = \left\{\mathbf{x} \in \mathbb{R}^3 \,\middle|\, \left\langle \mathbf{x}, \begin{bmatrix}1\\-2\\3\end{bmatrix}\right\rangle = 0\right\},$$

$\mathbf{v} = \begin{bmatrix}1\\-2\\3\end{bmatrix}$ is orthogonal to both the basis vectors $\mathbf{f}_1$ and $\mathbf{f}_2$ of U. Therefore $\mathbf{f}_3 = \frac{1}{\|\mathbf{v}\|}\mathbf{v} = \frac{1}{\sqrt{14}}\mathbf{v}$ is a unit vector orthogonal to $\mathbf{f}_1$ and $\mathbf{f}_2$, so $(\mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_3)$ is an orthonormal list in $\mathbb{R}^3$, hence an orthonormal basis.

4.2.17. We perform the Gram–Schmidt process, beginning with the standard basis on $\mathbb{R}^3$ (which is *not* orthonormal with respect to this non-standard inner product). Note that the norm corresponding to this inner product is given by $\|\mathbf{v}\|_{\mathbf{A}} = \sqrt{\langle \mathbf{v}, \mathbf{v}\rangle_{\mathbf{A}}} = \|\mathbf{A}\mathbf{v}\|$.

- $\mathbf{f}_1 = \frac{1}{\|\mathbf{A}\mathbf{e}_1\|}\mathbf{e}_1 = \frac{1}{\sqrt{2}}\mathbf{e}_1$.
- $\tilde{\mathbf{f}}_2 = \mathbf{e}_2 - \langle \mathbf{A}\mathbf{e}_2, \mathbf{A}\mathbf{f}_1\rangle \mathbf{f}_1 = \mathbf{e}_2 - \dfrac{1}{\sqrt{2}}\mathbf{f}_2 = \begin{bmatrix}-1/2\\1\\0\end{bmatrix}$.
- $\mathbf{f}_2 = \dfrac{1}{\left\|\mathbf{A}\tilde{\mathbf{f}}_2\right\|}\tilde{\mathbf{f}}_2 = \dfrac{1}{\sqrt{3/2}}\tilde{\mathbf{f}}_2 = \dfrac{1}{\sqrt{6}}\begin{bmatrix}-1\\2\\0\end{bmatrix}$
- $\tilde{\mathbf{f}}_3 = \mathbf{e}_3 - \langle \mathbf{A}\mathbf{e}_3, \mathbf{A}\mathbf{f}_1\rangle \mathbf{f}_1 - \langle \mathbf{A}\mathbf{e}_3, \mathbf{A}\mathbf{f}_2\rangle \mathbf{f}_2 = \mathbf{e}_3 - \dfrac{1}{\sqrt{2}}\mathbf{f}_1 - \dfrac{1}{\sqrt{6}}\mathbf{f}_2 = \begin{bmatrix}-1/3\\-1/3\\1\end{bmatrix}$.

- $\mathbf{f}_3 = \dfrac{1}{\left\|\mathbf{A}\tilde{\mathbf{f}}_3\right\|}\tilde{\mathbf{f}}_3 = \dfrac{1}{\sqrt{4/3}}\tilde{\mathbf{f}}_3 = \dfrac{1}{\sqrt{12}}\begin{bmatrix}-1\\-1\\3\end{bmatrix}.$

4.2.18. (a) Recall that $\mathbf{A}\mathbf{e}_j = \mathbf{a}_j$, the jth column of $\mathbf{A}$, and so

$$\langle \mathbf{e}_j, \mathbf{e}_k\rangle_{\mathbf{A}} = \langle \mathbf{A}\mathbf{e}_j, \mathbf{A}\mathbf{e}_k\rangle = \langle \mathbf{a}_j, \mathbf{a}_k\rangle .$$

(b) The standard basis $(\mathbf{e}_1, \dots, \mathbf{e}_n)$ is an orthonormal basis for $\mathbb{C}^n$ with respect to $\langle\cdot,\cdot\rangle_{\mathbf{A}}$ if and only if $\langle \mathbf{e}_j, \mathbf{e}_k\rangle_{\mathbf{A}} = 1$ when $j = k$ and is zero otherwise, which by the previous part is true if and only if $(\mathbf{a}_1, \dots, \mathbf{a}_n)$ is an orthonormal list (hence basis).

4.2.19. For each $n \in \mathbb{N}$, the list $(\sin(x), \sin(2x), \dots, \sin(nx))$ is an orthogonal, hence linearly independent, list of vectors in $C_{2\pi}(\mathbb{R})$, so $C_{2\pi}(\mathbb{R})$ is infinite-dimensional by Theorem 3.18.

4.2.20. Let $\mathcal{B} = (v_1, \dots, v_n)$. By Lemma 3.68, the matrix $[\boldsymbol{T}]_{\mathcal{B}}$ is upper triangular if and only if $\boldsymbol{T}v_j \in \langle v_1, \dots, v_j\rangle$ for all $v_j \in \mathcal{B}$. If $\mathcal{C} = (e_1, \dots, e_n)$ is the orthonormal basis obtained from $\mathcal{B}$ via the Gram–Schmidt process, then $\langle v_1 \dots, v_j\rangle = \langle e_1, \dots, e_j\rangle$ for each j. Since $e_j \in \langle v_1, \dots, v_j\rangle$ and $\boldsymbol{T}$ is linear, $\boldsymbol{T}e_j \in \langle \boldsymbol{T}v_1, \dots, \boldsymbol{T}v_j\rangle \subseteq \langle v_1, \dots, v_j\rangle = \langle e_1, \dots, e_j\rangle$. Then by Lemma 3.68, $[\boldsymbol{T}]_{\mathcal{C}}$ is upper triangular.

4.2.21. By definition,

$$\tilde{e}_j = v_j - \sum_{k=1}^{j-1} \langle v_j, e_k\rangle e_k,$$

so

$$\langle v_j, e_j\rangle = \left\langle \tilde{e}_j + \sum_{k=1}^{j-1} \langle v_j, e_k\rangle e_k, e_j \right\rangle = \langle \tilde{e}_j, e_j\rangle + \sum_{k=1}^{j-1} \langle v_j, e_k\rangle \langle e_k, e_j\rangle .$$

Since $(e_1, \dots, e_n)$ is orthonormal and $\tilde{e}_j = \|\tilde{e}_j\| e_j$, it follows that

$$\langle v_j, e_j\rangle = \|v_j\| > 0.$$

4.2.22. Let $\mathbf{a}_j$ denote the jth column of $\mathbf{A}$, and let $(\mathbf{f}_1, \dots, \mathbf{f}_n)$ denote the orthonormal basis constructed from $(\mathbf{a}_1, \dots, \mathbf{a}_n)$ via the Gram–Schmidt process. We first prove by induction on k that for each $k = 1, \dots, n$, $\mathbf{f}_k = \omega_k \mathbf{e}_k$ for some $\omega_k \in \mathbb{C}$.

First, $\mathbf{f}_1 = \frac{1}{\|\mathbf{a}_1\|}\mathbf{a}_1 = \frac{a_{11}}{|a_{11}|}\mathbf{e}_1$ since $\mathbf{A}$ is upper triangular.

Now suppose that $\mathbf{f}_j = \omega_j \mathbf{e}_j$ for $j = 1, \dots, k-1$. We know $\langle \mathbf{f}_1, \dots, \mathbf{f}_k\rangle = \langle \mathbf{a}_1, \dots, \mathbf{a}_k\rangle \subseteq \langle \mathbf{e}_1, \dots, \mathbf{e}_k\rangle$, where the latter containment holds because $\mathbf{A}$ is upper triangular. In particular, $\mathbf{f}_k \in \langle \mathbf{e}_1, \dots, \mathbf{e}_k\rangle$, so

$$\mathbf{f}_k = \sum_{j=1}^{k} \langle \mathbf{f}_k, \mathbf{e_j}\rangle \mathbf{e}_j.$$

Since $(\mathbf{f}_1, \dots, \mathbf{f}_n)$ is orthonormal, the induction hypothesis implies that $\langle \mathbf{f}_k, \mathbf{e}_j\rangle = 0$ for $j \neq k$, and so $\mathbf{f}_k = \langle \mathbf{f}_k, \mathbf{e}_k\rangle \mathbf{e}_k$.

Finally, $|\omega_k| = \frac{\|\mathbf{f}_k\|}{\|\mathbf{e}_k\|} = 1$.

4.3 Orthogonal projections and optimization

4.3.1. (a) Let $\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 0 & -1 \\ -1 & 1 \\ 2 & 0 \end{bmatrix}$. Then $\mathbf{A}^*\mathbf{A} = \begin{bmatrix} 6 & 1 \\ 1 & 6 \end{bmatrix}$ and $(\mathbf{A}^*\mathbf{A})^{-1} = \frac{1}{35}\begin{bmatrix} 6 & -1 \\ -1 & 6 \end{bmatrix}$. By Proposition 4.18, the matrix of the orthogonal projection is

$$\mathbf{A}(\mathbf{A}^*\mathbf{A})^{-1}\mathbf{A}^* = \frac{1}{35}\begin{bmatrix} 1 & 2 \\ 0 & -1 \\ -1 & 1 \\ 2 & 0 \end{bmatrix}\begin{bmatrix} 6 & -1 \\ -1 & 6 \end{bmatrix}\begin{bmatrix} 1 & 0 & -1 & 2 \\ 2 & -1 & 1 & 0 \end{bmatrix} = \frac{1}{35}\begin{bmatrix} 26 & -11 & 7 & 8 \\ -11 & 6 & -7 & 2 \\ 7 & -7 & 14 & -14 \\ 8 & 2 & -14 & 24 \end{bmatrix}.$$

(b) By the same method, we obtain $\frac{1}{23}\begin{bmatrix} 18 & -2-i & 6-7i \\ -2+i & 22 & 1-4i \\ 6+7i & 1+4i & 6 \end{bmatrix}$

(c) By the same method, we obtain $\frac{1}{266}\begin{bmatrix} 66 & 74 & 82 & 26 & 18 \\ 74 & 87 & 100 & 9 & -4 \\ 82 & 100 & 118 & -8 & -26 \\ 26 & 9 & -8 & 111 & 128 \\ 18 & -4 & -26 & 128 & 150 \end{bmatrix}$

(d) Since there is only one vector, the first step of the Gram–Schmidt process implies that $\frac{1}{\sqrt{n}}\begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$ is an orthonormal basis of the space, so by Proposition 4.17, the matrix is

$$\frac{1}{n}\begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}\begin{bmatrix} 1 & \cdots & 1 \end{bmatrix} = \frac{1}{n}\begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix}.$$

4.3.2. (a) Let $\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix}$. Then $\mathbf{A}^*\mathbf{A} = \begin{bmatrix} 4 & 10 \\ 10 & 30 \end{bmatrix}$ and $(\mathbf{A}^*\mathbf{A})^{-1} = \frac{1}{10}\begin{bmatrix} 15 & -5 \\ -5 & 2 \end{bmatrix}$. By Proposition 4.18, the matrix of the orthogonal projection is

$$\mathbf{A}(\mathbf{A}^*\mathbf{A})^{-1}\mathbf{A}^* = \frac{1}{10}\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix}\begin{bmatrix} 15 & -5 \\ -5 & 2 \end{bmatrix}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \end{bmatrix} = \frac{1}{10}\begin{bmatrix} 7 & 4 & 1 & -2 \\ 4 & 3 & 2 & 1 \\ 1 & 2 & 3 & 4 \\ -2 & 1 & 4 & 7 \end{bmatrix}.$$

(b) By the same method, we obtain $\frac{1}{3}\begin{bmatrix} 2 & -i & 1 \\ i & 2 & -i \\ 1 & i & 2 \end{bmatrix}$.

(c) The given vectors are clearly orthogonal, so $\left(\frac{1}{\sqrt{3}}\begin{bmatrix}1\\0\\1\\0\\1\end{bmatrix}, \frac{1}{\sqrt{2}}\begin{bmatrix}0\\1\\0\\1\\0\end{bmatrix}\right)$ is an orthonormal basis of the space. Then by Proposition 4.17, the matrix is

$$\frac{1}{3}\begin{bmatrix}1\\0\\1\\0\\1\end{bmatrix}\begin{bmatrix}1 & 0 & 1 & 0 & 1\end{bmatrix} + \frac{1}{2}\begin{bmatrix}0\\1\\0\\1\\0\end{bmatrix}\begin{bmatrix}0 & 1 & 0 & 1 & 0\end{bmatrix} = \frac{1}{6}\begin{bmatrix}2 & 0 & 2 & 0 & 2\\0 & 3 & 0 & 3 & 0\\2 & 0 & 2 & 0 & 2\\0 & 3 & 0 & 3 & 0\\2 & 0 & 2 & 0 & 2\end{bmatrix}.$$

(d) Since there is only one vector, the first step of the Gram–Schmidt process implies that $\frac{1}{8\sqrt{2}}\begin{bmatrix}5\\7\\7\\2\\1\end{bmatrix}$ is an orthonormal basis of the space, so by Proposition 4.17, the matrix is

$$\frac{1}{128}\begin{bmatrix}5\\7\\7\\2\\1\end{bmatrix}\begin{bmatrix}5 & 7 & 7 & 2 & 1\end{bmatrix} = \frac{1}{128}\begin{bmatrix}25 & 35 & 35 & 10 & 5\\35 & 49 & 49 & 14 & 7\\35 & 49 & 49 & 14 & 7\\10 & 14 & 14 & 4 & 2\\5 & 7 & 7 & 2 & 1\end{bmatrix}.$$

4.3.3. (a) The space $U = \left\langle \begin{bmatrix}1\\2\\3\end{bmatrix} \right\rangle$ has orthonormal basis $\frac{1}{\sqrt{14}}\begin{bmatrix}1\\2\\3\end{bmatrix}$, so by Proposition 4.17,

$$[\boldsymbol{P}_U]_{\mathcal{E}} = \frac{1}{14}\begin{bmatrix}1\\2\\3\end{bmatrix}\begin{bmatrix}1\\2\\3\end{bmatrix} = \frac{1}{14}\begin{bmatrix}1 & 2 & 3\\2 & 4 & 6\\3 & 6 & 9\end{bmatrix}.$$

Then by part 8 of Theorem 4.16,

$$[\boldsymbol{P}_{U^\perp}]_{\mathcal{E}} = \mathbf{I}_n - [\boldsymbol{P}_U]_{\mathcal{E}} = \frac{1}{14}\begin{bmatrix}13 & -2 & -3\\-2 & 10 & -6\\-3 & -6 & 5\end{bmatrix}.$$

(b) The given space is $\left\langle \begin{bmatrix}3\\-1\\-5\end{bmatrix} \right\rangle^\perp$. By the same method as in part (a), the matrix of the orthogonal projection is

$$\frac{1}{35}\begin{bmatrix}26 & 3 & 15\\3 & 34 & -5\\15 & -5 & 10\end{bmatrix}.$$

(c) The given space is $\left\langle \begin{bmatrix} 2 \\ 1+i \\ -i \end{bmatrix} \right\rangle^{\perp}$. By the same method as in part (a), the matrix of the orthogonal projection is

$$\frac{1}{7}\begin{bmatrix} 3 & -2+2i & 2i \\ -2-2i & 5 & -1+i \\ -2i & -1-i & 6 \end{bmatrix}.$$

(d) The matrix has RREF $\begin{bmatrix} 1 & 0 & -\frac{1}{2} & \frac{3}{2} \\ 0 & 1 & \frac{7}{4} & -\frac{3}{4} \end{bmatrix}$. Therefore elements of the kernel can be written as

$$\begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix} = \frac{z}{4}\begin{bmatrix} 2 \\ -7 \\ 4 \\ 0 \end{bmatrix} + \frac{w}{4}\begin{bmatrix} -6 \\ 3 \\ 0 \\ 4 \end{bmatrix},$$

so the kernel has basis $\left(\begin{bmatrix} 2 \\ -7 \\ 4 \\ 0 \end{bmatrix}, \begin{bmatrix} -6 \\ 3 \\ 0 \\ 4 \end{bmatrix} \right)$, and by Proposition 4.18, the matrix of the orthogonal projection onto the kernel is

$$\frac{1}{195}\begin{bmatrix} 121 & -32 & -19 & -87 \\ -32 & 139 & -82 & -6 \\ -19 & -82 & 61 & 33 \\ -87 & -6 & 33 & 69 \end{bmatrix}.$$

4.3.4. (a) The space $U = \left\langle \begin{bmatrix} 3 \\ 1 \\ 4 \end{bmatrix} \right\rangle$ has orthonormal basis $\frac{1}{\sqrt{26}}\begin{bmatrix} 3 \\ 1 \\ 4 \end{bmatrix}$, so by Proposition 4.17,

$$[\boldsymbol{P}_U]_{\mathcal{E}} = \frac{1}{26}\begin{bmatrix} 3 \\ 1 \\ 4 \end{bmatrix}\begin{bmatrix} 3 & 1 & 4 \end{bmatrix} = \frac{1}{26}\begin{bmatrix} 9 & 3 & 12 \\ 3 & 1 & 4 \\ 12 & 4 & 16 \end{bmatrix}.$$

Then by part 8 of Theorem 4.16,

$$[\boldsymbol{P}_{U^\perp}]_{\mathcal{E}} = \mathbf{I}_3 - [\boldsymbol{P}_U]_{\mathcal{E}} = \frac{1}{26}\begin{bmatrix} 17 & -3 & -12 \\ -3 & 25 & -4 \\ -12 & -4 & 10 \end{bmatrix}.$$

(b) The given space is $\left\langle \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix} \right\rangle^{\perp}$. By the same method as in part (a), the matrix of the orthogonal projection is

$$\frac{1}{14}\begin{bmatrix} 5 & -6 & -3 \\ -6 & 10 & -2 \\ -3 & -2 & 13 \end{bmatrix}.$$

(c) The given space is $\left\langle \begin{bmatrix} 1 \\ -i \\ i \end{bmatrix} \right\rangle^{\perp}$. By the same method as in part (a), the matrix of the orthogonal projection is

$$\frac{1}{3}\begin{bmatrix} 2 & -i & i \\ i & 2 & 1 \\ -i & 1 & 2 \end{bmatrix}.$$

(d) The matrix has RREF $\begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$, so the column space has basis $\left(\begin{bmatrix} 1 \\ -1 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 0 \\ 3 \\ 2 \\ 1 \end{bmatrix} \right)$, and by Proposition 4.18, the matrix of the orthogonal projection onto the kernel is

$$\frac{1}{83}\begin{bmatrix} 14 & -11 & 2 & 29 \\ -11 & 62 & 34 & -5 \\ 2 & 34 & 24 & 16 \\ 29 & -5 & 16 & 66 \end{bmatrix}.$$

4.3.5. By Theorem 4.19, in each case the desired vector is $\mathbf{Px}$, where $\mathbf{P}$ is the matrix of the orthogonal projection, found in one of the earlier exercises.

(a) $\mathbf{P}$ was found in Exercise 4.3.1(a); $\mathbf{Px} = \frac{1}{35}\begin{bmatrix} 57 \\ -12 \\ -21 \\ 66 \end{bmatrix}$.

(b) $\mathbf{P}$ was found in Exercise 4.3.1(b); $\mathbf{Px} = \frac{1}{23}\begin{bmatrix} 61 - 9i \\ -5 + 21i \\ 20 + 22i \end{bmatrix}$.

(c) $\mathbf{P}$ was found in Exercise 4.3.3(b); $\mathbf{Px} = \frac{1}{35}\begin{bmatrix} 44 \\ 32 \\ 20 \end{bmatrix}$.

(d) $\mathbf{P}$ was found in Exercise 4.3.3(d); $\mathbf{Px} = \frac{1}{195}\begin{bmatrix} -467 \\ 109 \\ 83 \\ 339 \end{bmatrix}$.

4.3.6. By Theorem 4.19, in each case the desired vector is $\mathbf{Px}$, where $\mathbf{P}$ is the matrix of the orthogonal projection, found in one of the earlier exercises.

(a) $\mathbf{P}$ was found in Exercise 4.3.2(a); $\mathbf{Px} = \begin{bmatrix} 4 \\ 3 \\ 2 \\ 1 \end{bmatrix}$.

(b) $\mathbf{P}$ was found in Exercise 4.3.2(b); $\mathbf{Px} = \begin{bmatrix} -1 \\ -i \\ 0 \end{bmatrix}$.

(c) $\mathbf{P}$ was found in Exercise 4.3.4(c); $\mathbf{Px} = \frac{1}{3}\begin{bmatrix} 4-2i \\ 1+2i \\ -1-2i \end{bmatrix}$.

(d) $\mathbf{P}$ was found in Exercise 4.3.4(d); $\mathbf{Px} = \frac{1}{83}\begin{bmatrix} 34 \\ 80 \\ 76 \\ 106 \end{bmatrix}$.

4.3.7. As in the example on pages 259–260, we will find the orthogonal projection of $\mathbf{y} = \begin{bmatrix} 1 \\ 2 \\ 5 \\ 4 \\ 8 \end{bmatrix}$ onto $U = \left\langle \begin{bmatrix} -2 \\ -1 \\ 0 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \right\rangle$. By Proposition 4.18,

$$\boldsymbol{P}_U\mathbf{y} = \begin{bmatrix} -2 & 1 \\ -1 & 1 \\ 0 & 1 \\ 1 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 10 & 0 \\ 0 & 5 \end{bmatrix}^{-1} \begin{bmatrix} -2 & -1 & 0 & 1 & 2 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 5 \\ 4 \\ 8 \end{bmatrix} = \begin{bmatrix} -2 & 1 \\ -1 & 1 \\ 0 & 1 \\ 1 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} \frac{8}{5} \\ 4 \end{bmatrix}.$$

The best fitting line, i.e., the one corresponding to this orthogonal projection, is $y = \frac{8}{5}x + 4$.

4.3.8. As in the example on pages 259–260, we will find the orthogonal projection of $\mathbf{y} = \begin{bmatrix} 5 \\ 3 \\ 2 \\ 2 \\ 3 \end{bmatrix}$ onto $U = \left\langle \begin{bmatrix} -2 \\ -1 \\ 0 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \right\rangle$. By Proposition 4.18,

$$\boldsymbol{P}_U\mathbf{y} = \begin{bmatrix} -2 & 1 \\ -1 & 1 \\ 0 & 1 \\ 1 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 10 & 0 \\ 0 & 5 \end{bmatrix}^{-1} \begin{bmatrix} -2 & -1 & 0 & 1 & 2 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 5 \\ 3 \\ 2 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} -2 & 1 \\ -1 & 1 \\ 0 & 1 \\ 1 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} -\frac{1}{2} \\ 3 \end{bmatrix}.$$

The best fitting line, i.e., the one corresponding to this orthogonal projection, is $y = -\frac{x}{2} + 3$.

4.3.9. Let L be the line $y = mx$. Then $L = \left\langle \begin{bmatrix} 1 \\ m \end{bmatrix} \right\rangle$, so the unit vector $\frac{1}{\sqrt{1+m^2}} \begin{bmatrix} 1 \\ m \end{bmatrix}$ is an orthonormal basis of L. By Proposition 4.17, the matrix of the orthogonal projection P_L is

$$\frac{1}{1+m^2} \begin{bmatrix} 1 \\ m \end{bmatrix} \begin{bmatrix} 1 \\ m \end{bmatrix} = \frac{1}{1+m^2} \begin{bmatrix} 1 & m \\ m & m^2 \end{bmatrix}.$$

Then by part 2 of Theorem 4.19, the closest point on L to (a, b) is

$$P_L \begin{bmatrix} a \\ b \end{bmatrix} = \frac{1}{1+m^2} \begin{bmatrix} 1 & m \\ m & m^2 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \frac{1}{1+m^2} \begin{bmatrix} a+mb \\ ma+mb^2 \end{bmatrix}.$$

4.3.10. By part 2 of Theorem 4.19, the desired polynomial is the projection onto $\mathcal{P}_2(\mathbb{R})$ of the function $|x|$, with respect to the inner product

$$\langle f, g \rangle = \int_{-1}^{1} f(x)g(x)\, dx$$

on $C([-1, 1])$.

By Exercise 4.2.15(b),

$$(e_1(x), e_2(x), e_3(x)) = \left(\frac{1}{\sqrt{2}}, \sqrt{\frac{3}{2}}x, \frac{3\sqrt{5}}{2\sqrt{2}}x^2 - \frac{\sqrt{5}}{2\sqrt{2}} \right)$$

is an orthonormal basis for $\mathcal{P}_2(\mathbb{R})$ with respect to this inner product. We compute

- $\langle |x|, e_1(x) \rangle = \int_{-1}^{1} |x| \frac{1}{\sqrt{2}}\, dx = 2 \int_{0}^{1} \frac{x}{\sqrt{2}}\, dx = \frac{1}{\sqrt{2}}$.
- $\langle |x|, e_2(x) \rangle = \int_{-1}^{1} |x| \sqrt{\frac{3}{2}}x\, dx = 0$ by symmetry.
- $\langle |x|, e_3(x) \rangle = \int_{-1}^{1} |x| \left(\frac{3\sqrt{5}}{2\sqrt{2}}x^2 - \frac{\sqrt{5}}{2\sqrt{2}} \right) dx = \frac{3\sqrt{5}}{4\sqrt{2}} - \frac{\sqrt{5}}{2\sqrt{2}} = \frac{\sqrt{5}}{2\sqrt{2}}$.

Then by part 2 of Theorem 4.16, the desired polynomial is

$$\frac{1}{\sqrt{2}} \frac{1}{\sqrt{2}} + 0 + \frac{\sqrt{5}}{2\sqrt{2}} \left(\frac{3\sqrt{5}}{2\sqrt{2}}x^2 - \frac{\sqrt{5}}{2\sqrt{2}} \right) = \frac{15}{8}x^2 - \frac{1}{8}.$$

4.3.11. By part 2 of Theorem 4.19, the desired polynomial is the projection onto $\mathcal{P}_1(\mathbb{R})$ of the function e^x, with respect to the inner product

$$\langle f, g \rangle = \int_{-1}^{1} f(x)g(x)\, dx$$

on $C([-1, 1])$.

By Exercise 4.2.15(a), $(e_1(x), e_2(x)) = (1, 2\sqrt{3}x - \sqrt{3})$ is an orthonormal basis of $\mathcal{P}_1(\mathbb{R})$ with respect to the given inner product.

We compute

- $\langle e^x, e_1(x)\rangle = \displaystyle\int_0^1 e^x 1\, dx = e - 1.$
- $\langle e^x, e_2(x)\rangle = \displaystyle\int_0^1 e^x(2\sqrt{3}x - \sqrt{3})\, dx = \sqrt{3}(3 - e).$

Then by part 2 of Theorem 4.16, the desired polynomial is

$$p(x) = \langle e^x, 1\rangle\, 1 + \left\langle e^x, 2\sqrt{3}x - \sqrt{3}\right\rangle (2\sqrt{3}x - \sqrt{3}) = 6(e-3)x + 4e - 10.$$

4.3.12. By part 2 of Theorem 4.19, the desired polynomial is the projection onto $\mathcal{P}_2(\mathbb{R})$ of the function e^x, with respect to the inner product

$$\langle f, g\rangle = \int_{-1}^1 f(x)g(x)\, dx$$

on $C([-1,1])$.

By Exercise 4.2.15(b),

$$(e_1(x), e_2(x), e_3(x)) = \left(\frac{1}{\sqrt{2}}, \sqrt{\frac{3}{2}}x, \frac{3\sqrt{5}}{2\sqrt{2}}x^2 - \frac{\sqrt{5}}{2\sqrt{2}}\right)$$

is an orthonormal basis of $\mathcal{P}_2(\mathbb{R})$ with respect to the given inner product.

We compute

- $\langle e^x, e_1(x)\rangle = \displaystyle\int_{-1}^1 e^x \frac{1}{\sqrt{2}}\, dx = \frac{1}{\sqrt{2}}(e - e^{-1}).$
- $\langle e^x, e_2(x)\rangle = \displaystyle\int_{-1}^1 e^x \sqrt{\frac{3}{2}}x\, dx = \frac{\sqrt{6}}{e}.$
- $\langle e^x, e_3(x)\rangle = \displaystyle\int_{-1}^1 e^x \left(\frac{3\sqrt{5}}{2\sqrt{2}}x^2 - \frac{\sqrt{5}}{2\sqrt{2}}\right)\, dx = \sqrt{\frac{5}{2}}(e - 7e^{-1}).$

Then by part 2 of Theorem 4.16, the desired polynomial is

$$\begin{aligned} p(x) &= \langle e^x, e_1(x)\rangle\, e_1(x) + \langle e^x, e_2(x)\rangle\, e_x(x) + \langle e^x, e_3(x)\rangle\, e_3(x) \\ &= \frac{15}{4}(e - 7e^{-1})x^2 + \frac{3}{e}x - \frac{3}{4}e + \frac{33}{4e}. \end{aligned}$$

4.3.13. By part 2 of Theorem 4.19, the desired function is the projection of the function $f(x) = x$ onto the space U of such functions $g(x)$ with respect to the inner product

$$\langle f, g\rangle = \int_{-1}^1 f(x)g(x)\, dx$$

on $C([-1,1])$.

By Example 3 on pages 239–240,

$$(e_1(x), e_2(x), e_3(x), e_4(x), e_5(x)) = \left(\frac{1}{\sqrt{2\pi}}, \frac{1}{\sqrt{\pi}}\sin(x), \frac{1}{\sqrt{\pi}}\sin(2x), \frac{1}{\sqrt{\pi}}\cos(x), \frac{1}{\sqrt{\pi}}\cos(2x)\right)$$

is an orthonormal basis of U with respect to the given inner product.

We compute

- $\langle f(x), e_1(x)\rangle = \int_0^{2\pi} x\frac{1}{\sqrt{2\pi}}\,dx = \sqrt{2}\pi^{3/2}$.
- $\langle f(x), e_2(x)\rangle = \int_0^{2\pi} x\frac{1}{\sqrt{\pi}}\sin(x)\,dx = -2\sqrt{\pi}$.
- $\langle f(x), e_3(x)\rangle = \int_0^{2\pi} x\frac{1}{\sqrt{\pi}}\sin(2x)\,dx = -\sqrt{\pi}$.
- $\langle f(x), e_4(x)\rangle = \int_0^{2\pi} x\frac{1}{\sqrt{\pi}}\cos(x)\,dx = 0$.
- $\langle f(x), e_5(x)\rangle = \int_0^{2\pi} x\,dx = 0$.

Then by part 2 of Theorem 4.16, the desired function is

$$\begin{aligned} g(x) &= \langle x, e_1(x)\rangle\, e_1(x) + \langle x, e_2(x)\rangle\, e_2(x) + \langle x, e_3(x)\rangle\, e_3(x) + \langle x, e_4(x)\rangle\, e_4(x) + \langle x, e_5(x)\rangle\, e_5(x) \\ &= \pi - 2\sin(x) - \sin(2x). \end{aligned}$$

4.3.14. (a) First observe that if $\mathbf{0}$ is the zero matrix, then $\mathbf{0} = \mathbf{0}^{\mathrm{T}} = -\mathbf{0}^{\mathrm{T}}$, so both sets contain zero. Suppose that $\mathbf{A}, \mathbf{B} \in V$ and $c \in \mathbb{F}$. Then $\mathbf{A}^{\mathrm{T}} = \mathbf{A}$ and $\mathbf{B}^{\mathrm{T}} = \mathbf{B}$, and so $(c\mathbf{A} + \mathbf{B})^{\mathrm{T}} = c\mathbf{A}^{\mathrm{T}} + \mathbf{B}^{\mathrm{T}} = c\mathbf{A} + \mathbf{B}$. That is, $c\mathbf{A} + \mathbf{B} \in V$, and so V is a subspace. The proof for W is similar.

(b) Let $\mathbf{A} \in W$ and let $\mathbf{B} \in V$. Then

$$\langle \mathbf{A}, \mathbf{B}\rangle_F = \operatorname{tr}(\mathbf{A}\mathbf{B}^{\mathrm{T}}) = \operatorname{tr}(\mathbf{A}\mathbf{B}),$$

since $\mathbf{B}^{\mathrm{T}} = \mathbf{B}$. But by the symmetry of the inner product,

$$\langle \mathbf{A}, \mathbf{B}\rangle_F = \langle \mathbf{B}, \mathbf{A}\rangle_F = \operatorname{tr}(\mathbf{B}\mathbf{A}^{\mathrm{T}}) = -\operatorname{tr}(\mathbf{B}\mathbf{A}) = -\operatorname{tr}(\mathbf{A}\mathbf{B}),$$

where the third equality follows because $\mathbf{A} \in W$. Since $\langle \mathbf{A}, \mathbf{B}\rangle_F = -\langle \mathbf{A}, \mathbf{B}\rangle_F$, $\langle \mathbf{A}, \mathbf{B}\rangle_F = 0$. This means that $W \subseteq V^{\perp}$ and that $V \subseteq W^{\perp}$. Because the space of $n \times n$ matrices is finite-dimensional, $(W^{\perp})^{\perp} = W$, and by the quiz, it then follows from the second inclusion that $W \subseteq V^{\perp}$, and so $W = V^{\perp}$.

(c) We can certainly write $\mathbf{A} = \operatorname{Re}\mathbf{A} + i\operatorname{Im}\mathbf{A} = \frac{1}{2}(\mathbf{A} + \mathbf{A}^{\mathrm{T}}) + \frac{1}{2}(\mathbf{A} - \mathbf{A}^{\mathrm{T}})$. We can easily check that

$$(\operatorname{Re}\mathbf{A})^{\mathrm{T}} = \frac{1}{2}(\mathbf{A}^{\mathrm{T}} + \mathbf{A}) = \operatorname{Re}\mathbf{A},$$

and

$$(i\operatorname{Im}\mathbf{A})^{\mathrm{T}} = \frac{1}{2}(\mathbf{A}^{\mathrm{T}} - \mathbf{A}) = -i\operatorname{Im}\mathbf{A},$$

so that $\operatorname{Re}\mathbf{A} \in V$ and $i\operatorname{Im}\mathbf{A} \in W = V^{\perp}$, so it follows that $P_V\mathbf{A} = \operatorname{Re}\mathbf{A}$, and similarly, $P_W\mathbf{A} = i\operatorname{Im}\mathbf{A}$.

4.3.15. Let $v \in U_2^\perp$. Then $\langle v, u\rangle = 0$ for every $u \in U_2$, so in particular $\langle v, u\rangle = 0$ for every $u \in U_1$.

4.3.16. Let $(u_1, \dots, u_k)$ be an orthonormal basis of U and let $(w_1, \dots, w_\ell)$ be an orthonormal basis of $U^\perp$. Then $(u_1, \dots, u_k, w_1, \dots, w_\ell)$ is an orthonormal list, since every w_j is orthogonal to every u_i ($w_j \in U^\perp$ and $u_i \in U$). Moreover, since each $v \in V$ can be written as $v = u + w$ with $u \in U$ and $w \in U^\perp$, it follows that $(u_1, \dots, u_k, w_1, \dots, w_\ell)$ spans V, and is therefore an orthonormal basis of V. In particular, $\dim V = k + \ell = \dim U + \dim U^\perp$.

4.3.17. By the linearity properties of the inner product,

$$\langle v, w\rangle = \sum_{j=1}^{m}\sum_{k=1}^{m} \langle v_j, w_k\rangle .$$

Now because $v_j \in U_j$ and $w_k \in U_k$, $\langle v_j, w_k\rangle = 0$ unless $j = k$, so

$$\langle v, w\rangle = \sum_{j=1}^{m} \langle v_j, w_j\rangle .$$

In particular, when $v = w$, we get

$$\|v\|^2 = \langle v, v\rangle = \sum_{j=1}^{m} \langle v_j, v_j\rangle = \sum_{j=1}^{m} \|v_j\|^2 .$$

4.3.18. Suppose that $u_1 + \cdots + u_m = v_1 + \cdots + v_n$, where $u_j, v_j \in U_j$ for each j. Then for each j,

$$u_j = P_{U_j}(u_1 + \cdots + u_m) = P_{U_j}(v_1 + \cdots + v_m) = v_j.$$

4.3.19. For each i, let $\mathcal{B}_i = (e^i_1, \dots, e^i_{n_i})$ be an orthonormal basis of U_i (so $n_i = \dim U_i$), and let $\mathcal{B} = (e^1_1, \dots, e^i_{n_1}, \dots, e^1_1, \dots, e^i_{n_m})$.

Now $\left\langle e^i_j, e^i_k \right\rangle = 0$ unless $j = k$, since $\mathcal{B}_i$ is orthonormal, and $\left\langle e^{i_1}_j, e^{i_2}_k \right\rangle = 0$ whenever $i_1 \neq i_2$, since every vector in U_{i_1} is orthogonal to every vector in U_{i_2}. Therefore $\mathcal{B}$ is orthogonal, hence linearly independent.

Moreover, if $v \in V$ then $v = u_1 + \cdots + u_m$ for some $u_i \in U_i$, and each u_i is a linear combination of the vectors in $\mathcal{B}_i$. Therefore v is a linear combination of the vectors in $\mathcal{B}$. So $\mathcal{B}$ spans V, and therefore is a basis of V. It follows that $\dim V = n_1 + \cdots + n_m$.

4.3.20. For each i, let $\mathcal{B}_i = (e^i_1, \dots, e^i_{n_i})$ be an orthonormal basis of U_i (so $n_i = \dim U_i$). As in Exercise 4.3.19, $\mathcal{B} = (e^1_1, \dots, e^i_{n_1}, \dots, e^1_1, \dots, e^i_{n_m})$ is then an orthonormal basis for U. The claim now follows immediately from part 2 of Theorem 4.16.

4.3.21. This follows immediately from part 5. of Theorem 4.16.

4.3.22. (a) The jth column of $[\boldsymbol{P}_U]_{\mathcal{B}}$ is $[\boldsymbol{P}_U e_j]_{\mathcal{B}}$. If $j \le m$, then $e_j \in U$, and so $\boldsymbol{P}_U e_j = e_j$, and $[\boldsymbol{P}_U e_j]_{\mathcal{B}} = \mathbf{e}_j$. If $j > m$, then e_j is orthogonal to every basis element of U, so by the formula given in part 2 of the theorem, $\boldsymbol{P}_U e_j = 0$. That is, the first m columns have a 1 in the diagonal entry and zeroes elsewhere, and the remaining columns are all zero.

(b) By the definition of orthogonal projection, $P_U(v) \in U$, and so $\operatorname{range} P_U \subseteq U$. If $u \in U$, then $u = u + 0$ gives the unique way of writing u as a sum of a vector (u) in U and a vector (0) in $U^\perp$, so $P_U(u) = u$. This further implies that $U \subseteq \operatorname{range} P_U$.

(c) Let $w \in U^\perp$. Then $w = 0 + w$ gives the unique way of writing w as a sum of a vector (0) in U and a vector (w) in $U^\perp$. Then by the definition of orthogonal projection, $P_U(w) = 0$. Therefore $U^\perp \subseteq \ker P_U$.

Now suppose that $v \in \ker P_U$. Then by part 3 of Theorem 4.16, $v = v - P_U v \in U^\perp$. Therefore $\ker P_U \subseteq U^\perp$.

(d) If V is finite-dimensional, then any $v \in V$ can be written $v = u + w$ with $v \in U = (U^\perp)^\perp$ and $w \in U^\perp$, and then

$$\boldsymbol{P}_{U^\perp} v = w = v - u = (\boldsymbol{I} - \boldsymbol{P}_U)v.$$

(e) By part (b), for any $v \in V$, $\boldsymbol{P}_U v \in U$, and then again by part (b), this means that $\boldsymbol{P}_U(\boldsymbol{P}_U v) = \boldsymbol{P}_U v$.

4.3.23. Let $\mathcal{B} = (e_1, \dots, e_m)$ be an orthonormal basis of U. By part 2 of Theorem 4.16, for any $c \in \mathbb{F}$ and $v, w \in V$,

$$\boldsymbol{P}_U(cv + w) = \sum_{j=1}^m \langle cv + w, e_j \rangle e_j = c \sum_{j=1}^m \langle v, e_j \rangle e_j + \sum_{j=1}^m \langle w, e_j \rangle e_j = c\boldsymbol{P}_U(v) + \boldsymbol{P}_U(w).$$

4.4 Normed spaces

4.4.1. (a) In $\mathbb{R}^n$, $\|\mathbf{e}_1 + \mathbf{e}_2\|_1^2 + \|\mathbf{e}_1 - \mathbf{e}_2\|_1^2 = 2^2 + 2^2 = 8$, but $2\|\mathbf{e}_1\|_1^2 + 2\|\mathbf{e}_2\|_1^2 = 2 + 2 = 4$, so by the parallelogram identity (Proposition 4.20), the ℓ^1 norm is not the norm associated to any inner product.

(b) In $\mathbb{R}^n$, $\|\mathbf{e}_1 + \mathbf{e}_2\|_\infty^2 + \|\mathbf{e}_1 - \mathbf{e}_2\|_\infty^2 = 1^2 + 1^2 = 2$, but $2\|\mathbf{e}_1\|_\infty^2 + 2\|\mathbf{e}_2\|_\infty^2 = 2 + 2 = 4$, so by the parallelogram identity (Proposition 4.20), the ℓ^∞ norm is not the norm associated to any inner product.

4.4.2. Define $\mathbf{A}, \mathbf{B} \in \mathrm{M}_{m,n}(\mathbb{F})$ by $a_{11} = 1$ and $a_{jk} = 0$ for all other entries, and $b_{22} = 1$ and $b_{jk} = 0$ for all other entries. Then it is easy to check that

$$\|\mathbf{A}\|_{op} = \|\mathbf{B}\|_{op} = \|\mathbf{A} + \mathbf{B}\|_{op} = \|\mathbf{A} - \mathbf{B}\|_{op} = 1.$$

From this it follows that the parallelogram identity (Proposition 4.20) does not hold for the operator norm, and therefore the operator norm is not the norm associated to any inner product.

4.4.3. By definition,

$$\|\boldsymbol{T}\|_{op} = \max_{\|v\|=1} \|\boldsymbol{T}(v)\|,$$

and so for any $v \in V$,

$$\|\boldsymbol{T}(v)\| = \left\|\boldsymbol{T}\left(\|v\| \frac{v}{\|v\|}\right)\right\| = \left\|\|v\| \boldsymbol{T}\left(\frac{v}{\|v\|}\right)\right\| = \|v\| \left\|\boldsymbol{T}\left(\frac{v}{\|v\|}\right)\right\| \le \|v\| \|\boldsymbol{T}\|_{op},$$

since $\frac{v}{\|v\|}$ is a unit vector. That is, $C = \|\boldsymbol{T}\|$ has the property that $\|\boldsymbol{T}v\| \leq C\|v\|$ for all $v \in V$.

Suppose now that C' is another constant with this property. Then in particular if $u \in V$ is a unit vector,

$$\|\boldsymbol{T}u\| \leq C'\|u\| = C',$$

and so

$$\|\boldsymbol{T}\|_{op} = \max_{\|v\|=1} \|\boldsymbol{T}v\| \leq C'.$$

That is, $\|T\|_{op}$ is the smallest C such that $\|\boldsymbol{T}v\| \leq C\|v\|$ for all $v \in V$.

4.4.4. Let $\lambda \in \mathbb{C}$ be an eigenvalue of $\mathbf{A}$ and let $\mathbf{v}$ be a corresponding eigenvector. Since $\mathbf{v} \neq \mathbf{0}$, we can normalize to get a unit eigenvector $\mathbf{u} = \frac{\mathbf{v}}{\|\mathbf{v}\|}$. Then

$$\|\mathbf{A}\|_{op} = \max_{\|\mathbf{x}\|=1} \|\mathbf{A}\mathbf{x}\| \geq \|\mathbf{A}\mathbf{u}\| = \|\lambda\mathbf{u}\| = |\lambda|.$$

4.4.5. Suppose that $\mathbf{b}$ is the true value, and $\mathbf{e}$ is the vector of errors, so that you are actually solving the system $\mathbf{A}\mathbf{y} = \mathbf{b} + \mathbf{e}$ by computing $\mathbf{y} = \mathbf{A}^{-1}(\mathbf{b} + \mathbf{e})$. Then the error in the computed value is

$$\mathbf{y} - \mathbf{x} = \mathbf{A}^{-1}(\mathbf{b} + \mathbf{e}) - \mathbf{A}^{-1}\mathbf{b} = \mathbf{A}^{-1}\mathbf{e}.$$

Then the size of this error, in terms of its (standard) norm, is

$$\left\|\mathbf{A}^{-1}\mathbf{e}\right\| \leq \left\|\mathbf{A}^{-1}\right\|_{op} \|e\| = \left\|\mathbf{A}^{-1}\right\|_{op} \sqrt{\sum_{j=1}^{n} |e_j|^2} \leq \left\|\mathbf{A}^{-1}\right\|_{op} \epsilon\sqrt{m}.$$

4.4.6. Write $\mathbf{D} = \mathbf{diag}(d_1, \ldots, d_n)$. Suppose that $|d_k| = \|(d_1, \ldots, d_n)\|_\infty$. Then for any $\mathbf{v} \in \mathbb{F}^n$ with $\|\mathbf{v}\| = 1$,

$$\mathbf{D}\mathbf{v} = (d_1 v_1, \ldots, d_n v_n),$$

and so

$$\|\mathbf{D}\mathbf{v}\| = \sqrt{\sum_{j=1}^{n} |d_j v_j|^2} \leq |d_k| \sqrt{\sum_{j=1}^{n} |v_j|^2} = |d_k|.$$

Therefore $\|\mathbf{D}\| \leq |d_k| = \|(d_1, \ldots, d_n)\|_\infty$.

On the other hand, $\mathbf{D}\mathbf{e}_k = d_k\mathbf{e}_k$, so $\|\mathbf{D}\mathbf{e}_k\| = |d_k| \|\mathbf{e}_k\| = |d_k|$. Therefore $\|\mathbf{D}\| \geq |d_k|$.

4.4.7. Since $\|\mathbf{e}_j\| = 1$, $\|\mathbf{A}\|_{op} \geq \|\mathbf{A}\mathbf{e}_j\| = \|\mathbf{e}_j\|$.

4.4.8. If $\mathbf{a}_1, \ldots, \mathbf{a}_n$ denote the columns of $\mathbf{A}$, then $\|\mathbf{A}\|_{HS}^2 = \sum_{j=1}^{n} \|\mathbf{a}_j\|^2$ by Exercise 4.1.4. Moreover, if $\mathbf{b}_1, \ldots, \mathbf{b}_p$ are the columns of $\mathbf{B}$, then the columns of $\mathbf{AB}$ are $\mathbf{A}\mathbf{b}_1, \ldots, \mathbf{A}\mathbf{b}_p$. Therefore

$$\|\mathbf{AB}\|_F^2 = \sum_{j=1}^{p} \|\mathbf{A}\mathbf{b}_j\|^2 \leq \sum_{j=1}^{n} \|\mathbf{A}\|_{op}^2 \|\mathbf{b}_j\|^2 = \|\mathbf{A}\|_{op}^2 \sum_{j=1}^{n} \|\mathbf{b}_j\|^2 = \|\mathbf{A}\|_{op}^2 \|\mathbf{B}\|_F^2,$$

where we have used the inequality (4.11) in the middle step. Taking square roots of both sides shows that $\|\mathbf{AB}\|_F \leq \|\mathbf{A}\|_{op} \|\mathbf{B}\|_F$.

4.4.9. Let $u \in U$ be any unit vector. Then

$$\|\boldsymbol{T}\boldsymbol{S}u\| \le \|\boldsymbol{T}\|_{op} \|\boldsymbol{S}u\| \le \|\boldsymbol{T}\|_{op} \|\boldsymbol{S}\|_{op},$$

where the first inequality follows from (4.11). Therefore $\|\boldsymbol{T}\boldsymbol{S}\|_{op} \le \|\boldsymbol{T}\|_{op} \|\boldsymbol{S}\|_{op}$.

4.4.10. (a) For any $\mathbf{x} \in \mathrm{M}_n(\mathbb{C})$,

$$\|\mathbf{x}\| = \left\|\mathbf{A}^{-1}\mathbf{A}\mathbf{x}\right\| \le \left\|\mathbf{A}^{-1}\right\|_{op} \|\mathbf{A}\mathbf{x}\|$$

by the inequality (4.11).

(b) Suppose that $\mathbf{x} \in \ker \mathbf{B}$ and $\mathbf{x} \neq \mathbf{0}$. Then

$$\|\mathbf{A}\mathbf{x}\| = \|(\mathbf{A} - \mathbf{B})\mathbf{x}\| \le \|\mathbf{A} - \mathbf{B}\|_{op} \|\mathbf{x}\| < \left\|\mathbf{A}^{-1}\right\|_{op}^{-1},$$

where the equality holds since $\mathbf{B}\mathbf{x} = \mathbf{0}$, the first inequality holds by (4.11), and the second inequality is by the assumption in the problem and the fact that $\|\mathbf{x}\| > 0$. It follows that

$$\|\mathbf{A}\|_{op}^{-1} \|\mathbf{A}\mathbf{x}\| < \|\mathbf{x}\|.$$

This contradicts part (a), so there can be no such $\mathbf{x}$. That is, $\ker \mathbf{B} = \{\mathbf{0}\}$, and so by Corollary 3.36, $\mathbf{B}$ is invertible.

4.4.11. (a) Let $\mathbf{u} \in \mathbb{C}^n$ be a unit vector, and let $\mathbf{v} = \mathbf{A}^{-1}\mathbf{u}$. Then by (4.11),

$$1 = \|\mathbf{u}\| = \|\mathbf{A}\mathbf{v}\| \le \|\mathbf{A}\|_{op} \|\mathbf{v}\| = \|\mathbf{A}\|_{op} \left\|\mathbf{A}^{-1}\mathbf{u}\right\| \le \|\mathbf{A}\|_{op} \left\|\mathbf{A}^{-1}\right\|_{op} = \kappa(\mathbf{A}).$$

Alternatively, this follows from Exercise 4.4.9 and the fact that $\|\mathbf{I}_n\|_{op} = 1$.

(b) Inequality (4.11) implies that $\|\mathbf{b}\| = \|\mathbf{A}\mathbf{x}\| \le \|\mathbf{A}\|_{op} \|\mathbf{x}\|$ and $\left\|\mathbf{A}^{-1}\mathbf{h}\right\| \le \left\|\mathbf{A}^{-1}\right\|_{op} \|\mathbf{h}\|$. Therefore

$$\frac{\left\|\mathbf{A}^{-1}\mathbf{h}\right\|}{\|\mathbf{x}\|} \le \frac{\left\|\mathbf{A}^{-1}\right\|_{op} \|\mathbf{h}\|}{\|\mathbf{b}\| / \|\mathbf{A}\|_{op}} = \kappa(\mathbf{A}) \frac{\|\mathbf{h}\|}{\|\mathbf{b}\|}.$$

4.4.12. Write

$$\|\mathbf{A}\|_{\mathbb{R}} = \max_{\substack{\mathbf{v} \in \mathbb{R}^n \\ \|\mathbf{v}\| = 1}} \|\mathbf{A}\mathbf{v}\| \qquad \text{and} \qquad \|\mathbf{A}\|_{\mathbb{C}} = \max_{\substack{\mathbf{v} \in \mathbb{C}^n \\ \|\mathbf{v}\| = 1}} \|\mathbf{A}\mathbf{v}\|.$$

Then $\|\mathbf{A}\|_{\mathbb{R}} \le \|\mathbf{A}\|_{\mathbb{C}}$ immediately.

Suppose that $\mathbf{u} \in \mathbb{C}^n$ is such that $\|\mathbf{u}\| = 1$ and $\|\mathbf{A}\mathbf{u}\| = \|\mathbf{A}\|_{\mathbb{C}}$. Write $\mathbf{u} = \mathbf{x} + i\mathbf{y}$ for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. Then

$$\|\mathbf{u}\|^2 = \sum_{j=1}^{n} |u_j|^2 = \sum_{j=1}^{n} (x_j^2 + y_j^2) = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2$$

and similarly

$$\|\mathbf{A}\mathbf{u}\|^2 = \|\mathbf{A}\mathbf{x} + i\mathbf{A}\mathbf{y}\|^2 = \|\mathbf{A}\mathbf{x}\|^2 + \|\mathbf{A}\mathbf{y}\|^2$$

since $\mathbf{A}\mathbf{x}, \mathbf{A}\mathbf{y} \in \mathbb{R}^m$. Furthermore,

$$\|\mathbf{A}\mathbf{x}\| \le \|\mathbf{A}\|_{\mathbb{R}} \|\mathbf{x}\| \qquad \text{and} \qquad \|\mathbf{A}\mathbf{y}\| \le \|\mathbf{A}\|_{\mathbb{R}} \|\mathbf{y}\|$$

by (4.11), and therefore

$$\|\mathbf{Au}\| = \sqrt{\|\mathbf{Ax}\|^2 + \|\mathbf{Ay}\|^2} \leq \|\mathbf{A}\|_{\mathbb{R}} \sqrt{\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2} = \|\mathbf{A}\|_{\mathbb{R}} \|\mathbf{u}\|^2 = \|\mathbf{A}\|_{\mathbb{R}},$$

which implies that $\|\mathbf{A}\|_{\mathbb{C}} \leq \|\mathbf{A}\|_{\mathbb{R}}$.

4.4.13. By Quick Exercise #16 in Section 3.4, if $\mathbf{A} \in \mathrm{M}_{m,n}(\mathbb{R})$ has rank 1, then $\mathbf{A} = \mathbf{v}\mathbf{w}^{\mathrm{T}}$ for some nonzero vectors $\mathbf{v} \in \mathbb{R}^m$ and $\mathbf{w} \in \mathbb{R}^n$.

By Proposition 3.60,

$$\|\mathbf{A}\|_F^2 = \operatorname{tr} \mathbf{A}\mathbf{A}^{\mathrm{T}} = \operatorname{tr} \mathbf{v}\mathbf{w}^{\mathrm{T}}\mathbf{w}\mathbf{v}^{\mathrm{T}} = \operatorname{tr} \mathbf{v}^{\mathrm{T}}\mathbf{v}\mathbf{w}^{\mathrm{T}}\mathbf{w}.$$

Recall that $\mathbf{v}^{\mathrm{T}}\mathbf{v} = \|\mathbf{v}\|^2$, so the 1×1 matrix $\mathbf{v}^{\mathrm{T}}\mathbf{v}\mathbf{w}^{\mathrm{T}}\mathbf{w} = \|\mathbf{v}\|^2 \|\mathbf{w}\|^2$, and thus $\|\mathbf{A}\|_F = \|\mathbf{v}\| \|\mathbf{w}\|$.

Alternatively, $a_{ij} = v_i w_j$, so

$$\|\mathbf{A}\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n v_i^2 w_j^2} = \sqrt{\sum_{i=1}^m v_i^2 \sum_{j=1}^n w_j^2} = \|\mathbf{v}\| \|\mathbf{w}\|.$$

Now for any $\mathbf{x} \in \mathbb{R}^n$,

$$\|\mathbf{Ax}\| = \left\|\mathbf{v}\mathbf{w}^{\mathrm{T}}\mathbf{x}\right\| = \|\mathbf{v}\langle \mathbf{x}, \mathbf{w}\rangle\| = |\langle \mathbf{x}, \mathbf{w}\rangle| \|\mathbf{v}\| \leq \|\mathbf{x}\| \|\mathbf{w}\| \|\mathbf{v}\|$$

by the Cauchy–Schwarz inequality. Therefore

$$\|\mathbf{A}\| \leq \|\mathbf{v}\| \|\mathbf{w}\| = \|\mathbf{A}\|_F.$$

Finally, if $\mathbf{x} = \frac{1}{\|\mathbf{w}\|}\mathbf{w}$, then $\|\mathbf{x}\| = 1$ and

$$\|\mathbf{Ax}\| = \frac{1}{\|\mathbf{w}\|} \|\mathbf{Aw}\| = \frac{1}{\|\mathbf{w}\|} \left\|\mathbf{v}\mathbf{w}^{\mathrm{T}}\mathbf{w}\right\| = \frac{1}{\|\mathbf{w}\|} \|\mathbf{w}\|^2 \|\mathbf{v}\| = \|\mathbf{v}\| \|\mathbf{w}\|.$$

Therefore $\|\mathbf{A}\|_{op} \geq \|\mathbf{v}\| \|\mathbf{w}\| = \|\mathbf{A}\|_F$.

4.4.14. Let $\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix}$. Since $\mathbf{B}$ is upper triangular, we can read off that it has distinct eigenvalues 1 and 2, so it is diagonalizable, hence similar to $\mathbf{A}$. On the other hand, $\|\mathbf{A}\|_{op} = \|(1,2)\|_\infty = 2$ by Exercise 4.4.6, but

$$\|\mathbf{B}\|_{op} \geq \|\mathbf{b}_2\| = \left\| \begin{bmatrix} 1 \\ 2 \end{bmatrix} \right\| = \sqrt{5}$$

by Exercise 4.4.7, so $\|\mathbf{B}\|_{op} \neq \|\mathbf{A}\|_{op}$.

4.4.15. Let $\mathbf{x} = (x_1, \ldots, x_n)$. Then $\|\mathbf{x}\|_\infty = \sqrt{\max_{1 \leq j \leq n} |x_j|^2} \leq \sqrt{|x_1|^2 + \cdots + |x_n|^2} = \|\mathbf{x}\|_2$. On the other hand, $\sqrt{|x_1|^2 + \cdots + |x_n|^2} \leq \sqrt{n\,|x_j|^2}$, where $\|\mathbf{x}\|_\infty = |x_j|$.

4.4.16. For any $\mathbf{x} \in \mathbb{C}^n$,

$$\|\mathbf{x}\|_2 = \|x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n\|_2 \le \sum_{j=1}^n \|x_j\mathbf{e}_j\|_2 = \sum_{j=1}^n |x_j| \, \|\mathbf{e}_j\|_2 = \sum_{j=1}^n |x_j| = \|\mathbf{x}\|_1 \,,$$

where the inequality is just the triangle inequality for $\|\cdot\|_2$.

For the second inequality, let $\mathbf{1}$ denote the vector $(1, \dots, 1)$, and let $\mathbf{y}$ be the vector with $y_j = |x_j|$. Then by the Cauchy–Schwarz inequality,

$$\|\mathbf{x}\|_1 = \sum_{j=1}^n |x_j| = \langle \mathbf{y}, \mathbf{1} \rangle \le \|\mathbf{y}\|_2 \, \|\mathbf{1}\|_2 = \sqrt{n} \, \|\mathbf{x}\|_2 \,,$$

since $\|\mathbf{y}\|_2 = \|\mathbf{x}\|_2$ and $\|\mathbf{1}\|_2 = \sqrt{n}$.

4.4.17. Let $\mathbf{a}_j$ denote the jth column of $\mathbf{A}$. Then for any $\mathbf{x} \in \mathbb{C}^n$ with $\|\mathbf{x}\| = 1$,

$$\|\mathbf{A}\mathbf{x}\| \le \left\| \sum_{j=1}^n x_j \mathbf{a}_j \right\| \le \sum_{j=1}^n \|x_j\mathbf{a}_j\| = \sum_{j=1}^n |x_j| \, \|\mathbf{a}_j\| \le \sqrt{\sum_{j=1}^n |x_j|^2} \sqrt{\sum_{j=1}^n \|\mathbf{a}_j\|^2} = \|\mathbf{A}\|_F \,,$$

where the first inequality follows from the triangle inequality and the second inequality follows from the Cauchy–Schwarz inequality. Therefore $\|\mathbf{A}\|_{op} \le \|\mathbf{A}\|_F$.

For the other estimate,

$$\|\mathbf{A}\|_F = \sqrt{\sum_{j=1}^n \|\mathbf{a}_j\|^2} \le \sqrt{\sum_{j=1}^n \|\mathbf{A}\|_{op}^2} = \sqrt{n} \, \|\mathbf{A}\|_{op}$$

by Exercise 4.4.7.

4.4.18. (a) By the Cauchy–Schwarz inequality (see also Exercise 4.1.16),

$$\|f\|_1 = \int_0^1 |f(x)| \; dx = \langle |f(x)| \, , 1 \rangle \le \sqrt{\int_0^1 |f(x)|^2 \; dx} \sqrt{\int_0^1 1^2 \; dx} = \|f\| \,.$$

By basic properties of integrals,

$$\|f\| = \sqrt{\int_0^1 |f(x)| \; dx} \le \sqrt{\int_0^1 \|f\|_\infty \; dx} = \|f\|_\infty \,.$$

(b) For the given functions,

$$\|f_n\| = \sqrt{\int_0^{1/n} (1 - nx)^2 \; dx} = \sqrt{\int_0^{1/n} (1 - 2nx + n^2x^2) \; dx} = \frac{1}{\sqrt{3n}},$$

and

$$\|f_n\|_1 = \int_0^{1/n} (1 - nx) \; dx = \frac{1}{2n}$$

(the latter integral is the area of a triangle with height 1 and width $\frac{1}{n}$). Therefore

$$\|f_n\| = \frac{2n}{\sqrt{3n}} \|f_n\|_1 = \frac{2}{\sqrt{3}}\sqrt{n}\, \|f_n\|_1 .$$

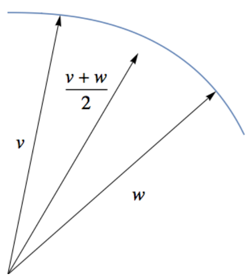
If there were such a constant $C > 0$, we would therefore have that $C \geq \frac{2}{\sqrt{3}}\sqrt{n}$ for every $n \in \mathbb{N}$, which is impossible.

(c) For the given functions, $\|f_n\|_\infty = 1$, so

$$\|f_n\|_\infty = \sqrt{3n}\, \|f_n\| .$$

If there were such a constant $C > 0$, we would therefore have that $C \geq \sqrt{3n}$ for every $n \in \mathbb{N}$, which is impossible.

4.4.19. (a)



(b) If $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ are both unit vectors, then by the parallelogram identity (Proposition 4.20), $\|\mathbf{v}+\mathbf{w}\|^2 + \|\mathbf{v}-\mathbf{w}\|^2 = 4$. Therefore

$$\left\|\frac{1}{2}(\mathbf{v}+\mathbf{w})\right\| = \sqrt{\frac{1}{4}\|\mathbf{v}+\mathbf{w}\|^2} = \sqrt{1 - \frac{1}{4}\|\mathbf{v}-\mathbf{w}\|^2}.$$

If $\mathbf{v} \neq \mathbf{w}$, then $\|\mathbf{v}-\mathbf{w}\| > 0$, and so $\left\|\frac{1}{2}(\mathbf{v}+\mathbf{w})\right\| < 1$.

(c) $\|\mathbf{e}_1\|_1 = \|\mathbf{e}_2\|_1 = 1$ but $\left\|\frac{1}{2}(\mathbf{e}_1+\mathbf{e}_2)\right\|_1 = 1$.

4.4.20. By the definition of the norm in an inner product space and the additivity properties of the inner product,

$$\begin{aligned}\|v+w\|^2 + \|v-w\|^2 &= \langle v+w, v+w\rangle + \langle v-w, v-w\rangle \\ &= \langle v,v\rangle + \langle v,w\rangle + \langle w,v\rangle + \langle w,w\rangle + \langle v,v\rangle - \langle v,w\rangle - \langle w,v\rangle + \langle w,w\rangle \\ &= 2\langle v,v\rangle + 2\langle w,w\rangle = 2\|v\|^2 + 2\|w\|^2 .\end{aligned}$$

4.5 Isometries

4.5.1. (a) Write $\boldsymbol{R}_U v = v + 2(\boldsymbol{P}_U v - v)$, and observe that $\boldsymbol{P}_U v = v + (\boldsymbol{P} v - v)$. This shows that $\boldsymbol{R}_U v$ gives the result of starting from the end of v, then going in the direction of the projection $\boldsymbol{P}_U v$, but twice as far. So you end up at the reflection of v in the subspace U.

(b) Let $v \in V$. By part 3 of Theorem 4.16, $\boldsymbol{P}_U v - v \in U^\perp$, so it is orthogonal to $\boldsymbol{P}_U v$, so by Theorem 4.4,

$$\begin{aligned} \|\boldsymbol{R}_U v\|^2 &= \|\boldsymbol{P}_U v + (\boldsymbol{P}_U v - v)\|^2 \\ &= \|\boldsymbol{P}_U v\|^2 + \|\boldsymbol{P}_U v - v\|^2 \\ &= \|\boldsymbol{P}_U v - (\boldsymbol{P}_U v - v)\|^2 \\ &= \|v\|^2 . \end{aligned}$$

4.5.2. If $\mathbf{x} \in \mathbb{C}^n$ is a unit vector, then

$$(\mathbf{I}_n - 2\mathbf{x}\mathbf{x}^*)(\mathbf{I}_n - 2\mathbf{x}\mathbf{x}^*)^* = (\mathbf{I}_n - 2\mathbf{x}\mathbf{x}^*)(\mathbf{I}_n - 2\mathbf{x}\mathbf{x}^*) = \mathbf{I}_n - 4\mathbf{x}\mathbf{x}^* + 4\mathbf{x}\mathbf{x}^*\mathbf{x}\mathbf{x}^*.$$

Since $\mathbf{x}^*\mathbf{x} = \langle \mathbf{x}, \mathbf{x} \rangle = 1$, this simplifies to $\mathbf{I}_n$. Thus $\mathbf{I}_n - \mathbf{x}\mathbf{x}^*$ is unitary.

4.5.3. Write $\mathbf{f}_k$ for the kth column of $\mathbf{F}$. Then

$$\langle \mathbf{f}_k, \mathbf{f}_\ell \rangle = \sum_{j=1}^n f_{jk}\overline{f_{j\ell}} = \frac{1}{n}\sum_{j=1}^n \omega^{jk}\omega^{-j\ell} = \frac{1}{n}\sum_{j=1}^n \big(\omega^{(k-\ell)}\big)^j.$$

If $k = \ell$, then each term of the sum is 1, so $\|\mathbf{f}_k\|^2 = 1$. If $k \neq \ell$, then from the above,

$$\begin{aligned} \langle \mathbf{f}_k, \mathbf{f}_\ell \rangle &= \frac{\omega^{(k-\ell)}}{n}\sum_{j=0}^{n-1}\big(\omega^{(k-\ell)}\big)^j = \frac{\omega^{(k-\ell)}}{n}\left(\frac{1-\big(\omega^{(k-\ell)}\big)^n}{1-\omega^{(k-\ell)}}\right) \\ &= \frac{\omega^{(k-\ell)}}{n}\left(\frac{1-\big(\omega^{n}\big)^{(k-\ell)}}{1-\omega^{(k-\ell)}}\right) = 0 \end{aligned}$$

since $\omega^n = 1$. Thus the columns of $\mathbf{F}$ are orthonormal, so $\mathbf{F}$ is a unitary matrix.

4.5.4. (a) Since $\boldsymbol{R}\left(\begin{bmatrix}1\\0\end{bmatrix}\right) = \begin{bmatrix}1\\0\end{bmatrix}$ and $\boldsymbol{R}\left(\begin{bmatrix}1\\1\end{bmatrix}\right) = \begin{bmatrix}1\\-1\end{bmatrix} = 2\begin{bmatrix}1\\0\end{bmatrix} - \begin{bmatrix}1\\1\end{bmatrix}$, the matrix $\mathbf{A}$ of $\boldsymbol{R}$ with respect to the given basis is

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

(b) Since

$$\mathbf{A}^T = \begin{bmatrix}1 & 0\\2 & -1\end{bmatrix} \neq \mathbf{A},$$

the matrix $\mathbf{A}$ is not orthogonal (even though $\mathbf{A}$ is an isometry).

(c) This does not contradict Corollary 4.30, since the given basis is not orthonormal.

4.5.5. **Solution 1:** The line making an angle of $\frac{\theta}{2}$ radians with the positive x-axis is spanned by the unit vector $\mathbf{v}_1 = \begin{bmatrix}\cos(\theta/2)\\ \sin(\theta/2)\end{bmatrix}$. The unit vector $\mathbf{v}_2 = \begin{bmatrix}-\sin(\theta/2)\\ \cos(\theta/2)\end{bmatrix}$ is orthogonal to it. Therefore $\mathcal{B} = (\mathbf{v}_1, \mathbf{v}_2)$ is an orthonormal basis of $\mathbb{R}^2$. If $\boldsymbol{R}$ is the reflection across the line spanned by $\mathbf{v}_1$, then $\boldsymbol{R}\mathbf{v}_1 = \mathbf{v}_1$ and $\boldsymbol{R}\mathbf{v}_2 = -\mathbf{v}_2$, so $[\boldsymbol{R}]_{\mathcal{B}} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$. Then by the change of basis formula,

$$\begin{aligned}
[\boldsymbol{R}]_{\mathcal{E}} &= [\boldsymbol{I}]_{\mathcal{B},\mathcal{E}}\,[\boldsymbol{R}]_{\mathcal{B}}\,[\boldsymbol{I}]_{\mathcal{E},\mathcal{B}} \\
&= \begin{bmatrix} \cos(\theta/2) & -\sin(\theta/2) \\ \sin(\theta/2) & \cos(\theta/2) \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} \cos(\theta/2) & -\sin(\theta/2) \\ \sin(\theta/2) & \cos(\theta/2) \end{bmatrix}^{-1} \\
&= \begin{bmatrix} \cos(\theta/2) & -\sin(\theta/2) \\ \sin(\theta/2) & \cos(\theta/2) \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} \cos(\theta/2) & \sin(\theta/2) \\ -\sin(\theta/2) & \cos(\theta/2) \end{bmatrix} \\
&= \begin{bmatrix} \cos^2(\theta/2) - \sin^2(\theta/2) & 2\cos(\theta/2)\sin(\theta/2) \\ 2\cos(\theta/2)\sin(\theta/2) & \sin^2(\theta/2) - \cos^2(\theta/2) \end{bmatrix} = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ \sin(\theta) & -\cos(\theta) \end{bmatrix},
\end{aligned}$$

where the last equality follows from standard trigonometric identities.

Solution 2: The matrix of the orthogonal projection onto the span of the unit vector $\mathbf{v}_1$ is $\mathbf{v}_1\mathbf{v}_1^*$. By Exercise 4.5.1, the matrix of the reflection across that span is therefore

$$\begin{aligned}
2\begin{bmatrix} \cos(\theta/2) \\ \sin(\theta/2) \end{bmatrix} \begin{bmatrix} \cos(\theta/2) & \sin(\theta/2) \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} &= \begin{bmatrix} 2\cos^2(\theta/2) - 1 & 2\cos(\theta/2)\sin(\theta/2) \\ 2\cos(\theta/2)\sin(\theta/2) & \sin^2(\theta/2) - 1 \end{bmatrix} \\
&= \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ \sin(\theta) & -\cos(\theta) \end{bmatrix},
\end{aligned}$$

where the last equality follows from standard trigonometric identities.

4.5.6. First, $(\boldsymbol{T}(cf+g))(x) = (cf+g)(x+t) = cf(x+t) + g(x+t) = c\boldsymbol{T}f(x) + \boldsymbol{T}g(x) = (c\boldsymbol{T}f + \boldsymbol{T}g)(x)$, so $\boldsymbol{T}(cf+g) = c\boldsymbol{T}f + \boldsymbol{T}g$. Therefore $\boldsymbol{T}$ is linear.

Second,

$$\begin{aligned}
\|\boldsymbol{T}f\|^2 &= \int_0^{2\pi} |(\boldsymbol{T}f)(x)|^2\,dx \\
&= \int_0^{2\pi} |f(x+t)|^2\,dx \\
&= \int_t^{2\pi+t} |f(u)|^2\,du \\
&= \int_t^{2\pi} |f(u)|^2\,du + \int_{2\pi}^{2\pi+t} |f(u)|^2\,du \\
&= \int_t^{2\pi} |f(u)|^2\,du + \int_0^{t} |f(u)|^2\,du \\
&= \int_0^{2\pi} |f(u)|^2\,du \\
&= \|f\|^2,
\end{aligned}$$

where the fifth equality follows since f is 2π-periodic.

Finally, if $g \in C_{2\pi}(\mathbb{R})$, then $g = \boldsymbol{T}f$, where $f(x) = g(x-t)$. Thus $\boldsymbol{T}$ is surjective.

4.5.7. (a) We first apply the Gram–Schmidt process to the columns $\mathbf{a}_1, \mathbf{a}_2$ of $\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$:

- $\mathbf{q}_1 = \dfrac{1}{\|\mathbf{a}_1\|}\mathbf{a}_1 = \dfrac{1}{\sqrt{1^2+3^2}}\begin{bmatrix} 1 \\ 3 \end{bmatrix} = \dfrac{1}{\sqrt{10}}\begin{bmatrix} 1 \\ 3 \end{bmatrix}$
- $\tilde{\mathbf{q}}_2 = \mathbf{a}_2 - \langle \mathbf{a}_2, \mathbf{q}_1 \rangle \mathbf{q}_1 = \begin{bmatrix} 2 \\ 4 \end{bmatrix} - \dfrac{(1\cdot 2 + 3\cdot 4)}{\sqrt{10}} \dfrac{1}{\sqrt{10}}\begin{bmatrix} 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 3/5 \\ -1/5 \end{bmatrix}$
- $\mathbf{q}_2 = \dfrac{1}{\|\tilde{\mathbf{q}}_2\|}\tilde{\mathbf{q}}_2 = \dfrac{1}{\sqrt{(3/5)^2 + (-1/5)^2}}\begin{bmatrix} 3/5 \\ -1/5 \end{bmatrix} = \dfrac{1}{\sqrt{10}}\begin{bmatrix} 3 \\ -1 \end{bmatrix}.$

Thus $\mathbf{Q} = \dfrac{1}{\sqrt{10}}\begin{bmatrix} 1 & 3 \\ 3 & -1 \end{bmatrix}$ and $\mathbf{A} = \mathbf{Q}^{\mathrm{T}}\mathbf{A} = \dfrac{1}{\sqrt{10}}\begin{bmatrix} 10 & 14 \\ 0 & 2 \end{bmatrix} = \sqrt{\dfrac{2}{5}}\begin{bmatrix} 5 & 7 \\ 0 & 1 \end{bmatrix}$.

(b) Applying the Gram–Schmidt process to the basis $\left(\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}\right)$ of $\mathbb{R}^3$ yields

$$\mathbf{q}_1 = \frac{1}{\sqrt{2}}\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \qquad \mathbf{q}_2 = \frac{1}{\sqrt{6}}\begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix}, \qquad , \mathbf{q}_3 = \frac{1}{\sqrt{3}}\begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix},$$

so

$$\mathbf{Q} = \frac{1}{\sqrt{6}}\begin{bmatrix} 0 & 2 & \sqrt{2} \\ \sqrt{3} & -1 & \sqrt{2} \\ \sqrt{3} & 1 & -\sqrt{2} \end{bmatrix}$$

and

$$\mathbf{R} = \mathbf{Q}^{\mathrm{T}}\mathbf{A} = \frac{1}{\sqrt{6}}\begin{bmatrix} 0 & \sqrt{3} & \sqrt{3} \\ 2 & -1 & 1 \\ \sqrt{2} & \sqrt{2} & -\sqrt{2} \end{bmatrix}\begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} = \frac{1}{\sqrt{6}}\begin{bmatrix} 2\sqrt{3} & \sqrt{3} & \sqrt{3} \\ 0 & 3 & 1 \\ 0 & 0 & 2\sqrt{2} \end{bmatrix}.$$

(c) Applying the Gram–Schmidt process to the basis $\left(\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix}\right)$ of $\mathbb{R}^3$ yields

$$\mathbf{q}_1 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \qquad \mathbf{q}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \qquad , \mathbf{q}_3 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix},$$

so

$$\mathbf{Q} = \begin{bmatrix} 1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 0 & 1 & 0 \\ -1/\sqrt{2} & 0 & 1/\sqrt{2} \end{bmatrix}$$

and

$$\mathbf{R} = \mathbf{Q}^{\mathrm{T}}\mathbf{A} = \begin{bmatrix} \sqrt{2} & -\sqrt{2} & \sqrt{2} \\ 0 & 2 & -1 \\ 0 & 0 & \sqrt{2} \end{bmatrix}.$$

(d) In this case we can observe that the columns of $\mathbf{A}$ are already orthogonal and all have the same norm 2. Thus orthonormalizing produces $\mathbf{Q} = \frac{1}{2}\mathbf{A}$, and so $\mathbf{A} = 2\mathbf{Q} = \mathbf{Q}(2\mathbf{I}_4)$, so $\mathbf{R} = 2\mathbf{I}_4$.

4.5.8. (a) Performing the Gram–Schmidt algorithm on the columns of the given matrix: $\mathbf{q}_1 = \frac{1}{\sqrt{5}}\begin{bmatrix} 2 \\ -1 \end{bmatrix}$, and

$$\tilde{\mathbf{q}}_2 = \begin{bmatrix} 0 \\ 3 \end{bmatrix} - \frac{1}{5}\left\langle \begin{bmatrix} 0 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ -1 \end{bmatrix} \right\rangle \begin{bmatrix} 2 \\ -1 \end{bmatrix} = \frac{1}{5}\begin{bmatrix} 6 \\ 12 \end{bmatrix},$$

so

$$\mathbf{q}_2 = \frac{1}{\sqrt{5}}\begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

We thus have

$$\mathbf{Q} = \frac{1}{\sqrt{5}}\begin{bmatrix} 2 & 1 \\ -1 & 2 \end{bmatrix} \quad \text{and} \quad \mathbf{R} = \mathbf{Q}^*\begin{bmatrix} 2 & 0 \\ -1 & 3 \end{bmatrix} = \frac{1}{5}\begin{bmatrix} 5 & -3 \\ 0 & 6 \end{bmatrix}.$$

(b) Following the same procedure,

$$\mathbf{Q} = \begin{bmatrix} \frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{6}} & \frac{2}{\sqrt{30}} \\ 0 & \frac{1}{\sqrt{6}} & \frac{5}{\sqrt{30}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{30}} \end{bmatrix} \quad \text{and} \quad \mathbf{R} = \mathbf{Q}^*\begin{bmatrix} 1 & -2 & 0 \\ 0 & 1 & 2 \\ 2 & 1 & -1 \end{bmatrix} = \begin{bmatrix} \sqrt{5} & 0 & -\frac{2}{\sqrt{5}} \\ 0 & \sqrt{6} & \frac{1}{\sqrt{6}} \\ 0 & 0 & \frac{11}{\sqrt{30}} \end{bmatrix}.$$

(c) Following the same procedure,

$$\mathbf{Q} = \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} \end{bmatrix} \quad \text{and} \quad \mathbf{R} = \mathbf{Q}^*\begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & -1 & -1 \end{bmatrix} = \begin{bmatrix} \sqrt{3} & 0 & 0 \\ 0 & \sqrt{2} & \frac{1}{\sqrt{2}} \\ 0 & 0 & \frac{3}{\sqrt{6}} \end{bmatrix}.$$

(d) Since the given matrix $\mathbf{A}$ is already upper triangular, we can just take $\mathbf{Q} = \mathbf{I}$ and $\mathbf{R} = \mathbf{A}$.

4.5.9. (a) The QR decomposition of $\mathbf{A} = \begin{bmatrix} -3 & 4 \\ 4 & -3 \end{bmatrix}$ is given by

$$\mathbf{Q} = \frac{1}{5}\begin{bmatrix} -3 & 4 \\ 4 & 3 \end{bmatrix} \quad \text{and} \quad \mathbf{R} = \frac{1}{5}\begin{bmatrix} 25 & -24 \\ 0 & 7 \end{bmatrix}.$$

The system $\mathbf{A}\mathbf{x} = \mathbf{b}$ is equivalent to $\mathbf{R}\mathbf{x} = \mathbf{Q}^*\mathbf{b}$, so we want to solve the system

$$\frac{1}{5}\begin{bmatrix} 25 & -24 \\ 0 & 7 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \frac{1}{5}\begin{bmatrix} -3 & 4 \\ 4 & 3 \end{bmatrix}\begin{bmatrix} -2 \\ 5 \end{bmatrix} = \frac{1}{5}\begin{bmatrix} 26 \\ 7 \end{bmatrix},$$

or equivalently

$$\begin{aligned} 25x - 24y &= 26 \\ 7y &= 7. \end{aligned}$$

We solve the second equation for y first ($y = 1$), then back-substitute to get $25x - 24 = 26$, so $25x = 50$ and $x = 2$.

(b) The QR decomposition of $\mathbf{A} = \begin{bmatrix} 1 & -1 & 1 \\ 0 & 1 & -1 \\ 1 & 0 & 2 \end{bmatrix}$ is given by

$$\mathbf{Q} = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{6} & -1/\sqrt{3} \\ 0 & \sqrt{2/3} & -1/\sqrt{3} \\ 1/\sqrt{2} & 1/\sqrt{6} & 1/\sqrt{3} \end{bmatrix} \qquad \text{and} \qquad \mathbf{R} = \begin{bmatrix} \sqrt{2} & -1/\sqrt{2} & 3/\sqrt{2} \\ 0 & \sqrt{3/2} & -1/\sqrt{6} \\ 0 & 0 & 2/\sqrt{3} \end{bmatrix}.$$

The system $\mathbf{Ax} = \mathbf{b}$ is equivalent to $\mathbf{Rx} = \mathbf{Q}^*\mathbf{b}$, so we want to solve the system

$$\begin{bmatrix} \sqrt{2} & -1/\sqrt{2} & 3/\sqrt{2} \\ 0 & \sqrt{3/2} & -1/\sqrt{6} \\ 0 & 0 & 2/\sqrt{3} \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} & 0 & 1/\sqrt{2} \\ -1/\sqrt{6} & \sqrt{2/3} & 1/\sqrt{6} \\ -1/\sqrt{3} & -1/\sqrt{3} & 1/\sqrt{3} \end{bmatrix} \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 3/\sqrt{2} \\ -3/\sqrt{6} \\ 0 \end{bmatrix},$$

or equivalently

$$\begin{aligned} 2x - y + 3z &= 3 \\ 3y - z &= -3 \\ z &= 0. \end{aligned}$$

Back-substituting $z = 0$ into the second equation gives $y = -1$, and back-substituting $y = -1$ and $z = 0$ into the first equation gives $2x + 1 = 3$, so $x = 1$.

4.5.10. (a) The QR decomposition of the coefficient matrix $\mathbf{A} = \begin{bmatrix} 3 & -1 \\ 1 & 5 \end{bmatrix}$ is given by

$$\mathbf{Q} = \frac{1}{\sqrt{10}} \begin{bmatrix} 3 & -1 \\ 1 & 3 \end{bmatrix} \qquad \text{and} \qquad \mathbf{R} = \frac{1}{\sqrt{10}} \begin{bmatrix} 10 & 2 \\ 0 & 16 \end{bmatrix}.$$

Solving $\mathbf{Rx} = \mathbf{Q}^*\mathbf{b}$ by back-substitution gives $x = \frac{29}{16}$ and $y = \frac{23}{16}$.

(b) Following the same procedure, the QR-decomposition of the coefficient matrix is given by

$$\mathbf{Q} = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ 0 & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \end{bmatrix} \qquad \text{and} \qquad \mathbf{R} = \begin{bmatrix} \sqrt{2} & -\frac{3}{\sqrt{2}} & 0 \\ 0 & \frac{3}{\sqrt{6}} & 0 \\ 0 & 0 & \sqrt{3} \end{bmatrix}.$$

Solving $\mathbf{Rx} = \mathbf{Q}^*\mathbf{b}$ by back-substitution gives $(x, y, z) = \left(3, \frac{2}{3}, -\frac{2}{3}\right)$.

4.5.11. Performing the Gram–Schmidt process on $\left(\begin{bmatrix} 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 3 \\ -4 \end{bmatrix} \right)$ yields the orthonormal basis $(\mathbf{e}_1, -\mathbf{e}_2)$, so $\mathbf{Q} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$, and then

$$\mathbf{R} = \mathbf{Q}^*\mathbf{A} = \begin{bmatrix} 2 & 3 \\ 0 & 4 \end{bmatrix}.$$

On the other hand, since $\mathbf{A}$ is already upper triangular, $\mathbf{A} = \mathbf{I}_2\mathbf{A}$ is also a QR decomposition.

4.5.12. (a) Suppose that $\boldsymbol{T}$ is an isometry and that λ is an eigenvalue of $\boldsymbol{T}$. If v is a corresponding eigenvector, then since $\boldsymbol{T}$ is an isometry,

$$\|v\| = \|\boldsymbol{T}v\| = \|\lambda v\| = |\lambda| \, \|v\| ,$$

and so $|\lambda| = 1$.

(b) If $\mathbf{U}$ is unitary, then the map given by multiplication by $\mathbf{U}$ is an isometry of $\mathbb{C}^n$, and so if λ is an eigenvalue of $\mathbf{U}$, it follows from the previous part that $|\lambda| = 1$.

4.5.13. All that is needed is to show that $\boldsymbol{T}$ is surjective. Since $\dim V = \dim W$, by Corollary 3.36 this is equivalent to showing $\boldsymbol{T}$ is injective. Suppose that $\boldsymbol{T}v = 0$. Then $\|\boldsymbol{T}v\| = \|v\| = 0$, and so $v = 0$. Thus $\boldsymbol{T}$ is injective and hence surjective.

4.5.14. (a) For every $\mathbf{x} \in \mathbb{C}^n$, $\|\mathbf{U}\mathbf{x}\| = \|\mathbf{x}\|$. Therefore

$$\|\mathbf{U}\| = \max_{\substack{\mathbf{x}\in\mathbb{C}^n \\ \|\mathbf{x}\|=1}} \|\mathbf{U}\mathbf{x}\| = \max_{\substack{\mathbf{x}\in\mathbb{C}^n \\ \|\mathbf{x}\|=1}} \|\mathbf{x}\| = 1.$$

(b) $\|\mathbf{U}\|_F^2 = \operatorname{tr} \mathbf{U}\mathbf{U}^* = \operatorname{tr} \mathbf{I}_n = n$, so $\|\mathbf{U}\|_F = \sqrt{n}$.

4.5.15. By definition, and since $\mathbf{U}$ acts as an isometry on $\mathbb{C}^m$,

$$\|\mathbf{U}\mathbf{A}\mathbf{V}\| = \max_{\substack{\mathbf{x}\in\mathbb{C}^n \\ \|\mathbf{x}\|=1}} \|\mathbf{U}\mathbf{A}\mathbf{V}\mathbf{x}\| = \max_{\substack{\mathbf{x}\in\mathbb{C}^n \\ \|\mathbf{x}\|=1}} \|\mathbf{A}\mathbf{V}\mathbf{x}\| .$$

Now substituting $\mathbf{y} = \mathbf{V}\mathbf{x}$, since $\mathbf{V}$ acts as an isometry on $\mathbb{C}^n$,

$$\|\mathbf{U}\mathbf{A}\mathbf{V}\| = \max_{\substack{\mathbf{y}\in\mathbb{C}^n \\ \|\mathbf{y}\|=1}} \|\mathbf{A}\mathbf{y}\| = \|\mathbf{A}\| .$$

Next, since $\mathbf{U}$ and $\mathbf{V}$ are unitary,

$$\begin{aligned} \|\mathbf{U}\mathbf{A}\mathbf{V}\|_F^2 &= \operatorname{tr}(\mathbf{U}\mathbf{A}\mathbf{V})(\mathbf{U}\mathbf{A}\mathbf{V}^*) = \operatorname{tr} \mathbf{U}\mathbf{A}\mathbf{V}\mathbf{V}^*\mathbf{A}^*\mathbf{U}^* = \operatorname{tr} \mathbf{U}\mathbf{A}\mathbf{A}^*\mathbf{U}^* \\ &= \operatorname{tr} \mathbf{U}^*\mathbf{U}\mathbf{A}\mathbf{A}^* = \operatorname{tr} \mathbf{A}\mathbf{A}^* = \|\mathbf{A}\|_F^2 , \end{aligned}$$

where we have also used Proposition 3.60.

4.5.16. Given any $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{x} = \sum_{j=1}^n x_j \mathbf{e}_j$, and so $\boldsymbol{T}_\pi \mathbf{x} = \sum_{j=1}^n x_j \mathbf{e}_{\pi(j)}$. Since π is bijective, this means that $\boldsymbol{T}_\pi \mathbf{x}$ has the same entries as $\mathbf{x}$, but in a different order. Since the ℓ^1 and ℓ^∞ norms don't depend on the order of the entries of a vector, $\|\boldsymbol{T}_\pi \mathbf{x}\|_1 = \|\mathbf{x}\|_1$ and $\|\boldsymbol{T}_\pi \mathbf{x}\|_\infty = \|\mathbf{x}\|_\infty$. Surjectivity can be shown directly, or follows from Exercise 4.5.13.

4.5.17. First observe that

$$\left\| \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \right\|_\infty = \max\{|x+y| , |x-y|\}.$$

If x, y have the same sign, then

$$\max\{|x+y|, |x-y|\} = |x+y| = \begin{cases} x+y, & x, y \geq 0; \\ -x-y, & x, y \leq 0, \end{cases}$$
$$= |x| + |y| = \left\| \begin{bmatrix} x \\ y \end{bmatrix} \right\|_1 .$$

If x and y have opposite signs, then

$$\max\{|x+y|, |x-y|\} = |x-y| = \begin{cases} x-y, & x \geq 0, y \leq 0; \\ -x+y, & x \leq 0, y \geq 0, \end{cases}$$
$$= |x| + |y| = \left\| \begin{bmatrix} x \\ y \end{bmatrix} \right\|_1 .$$

4.5.18. If $\kappa(\mathbf{A}) = 1$ and $\mathbf{x} \in \mathbb{C}^n$, then

$$\|\mathbf{x}\| = \left\|\mathbf{A}^{-1}\mathbf{A}\mathbf{x}\right\| \leq \left\|\mathbf{A}^{-1}\right\|_{op} \|\mathbf{A}\mathbf{x}\| \leq \left\|\mathbf{A}^{-1}\right\|_{op} \|\mathbf{A}\|_{op} \|\mathbf{x}\| = \|\mathbf{x}\| .$$

Since the first and last expressions in this string of inequalities are equal, we must have equality throughout. In particular, $\left\|\mathbf{A}^{-1}\right\|_{op} \|\mathbf{A}\mathbf{x}\| = \left\|\mathbf{A}^{-1}\right\|_{op} \|\mathbf{A}\|_{op} \|\mathbf{x}\|$ and so $\|\mathbf{A}\mathbf{x}\| = \|\mathbf{A}\|_{op} \|\mathbf{x}\|$. So if we define $\mathbf{B} = \frac{1}{\|\mathbf{A}\|_{op}}\mathbf{A}$, then $\|\mathbf{B}\mathbf{x}\| = \|\mathbf{x}\|$ for every $\mathbf{x} \in \mathbb{C}^n$. Therefore $\mathbf{B}$ represents an isometry, and is thus a unitary matrix.

4.5.19. By Exercise 4.5.15, $\|\mathbf{Q}\mathbf{R}\|_{op} = \|\mathbf{R}\|_{op}$ and $\left\|(\mathbf{Q}\mathbf{R})^{-1}\right\|_{op} = \left\|\mathbf{R}^{-1}\mathbf{Q}^{-1}\right\|_{op} = \left\|\mathbf{R}^{-1}\right\|_{op}$, since both $\mathbf{Q}$ and $\mathbf{Q}^{-1}$ are unitary. Therefore

$$\kappa(\mathbf{A}) = \|\mathbf{A}\|_{op} \left\|\mathbf{A}^{-1}\right\|_{op} = \|\mathbf{Q}\mathbf{R}\|_{op} \left\|(\mathbf{Q}\mathbf{R})^{-1}\right\|_{op} = \|\mathbf{R}\|_{op} \left\|\mathbf{R}^{-1}\right\|_{op} = \kappa(\mathbf{R}).$$

4.5.20. Since $\mathbf{R} = \mathbf{Q}^*\mathbf{A}$,

$$r_{jj} = \mathbf{q}_j^*\mathbf{a}_j = \langle \mathbf{a}_j, \mathbf{q}_j \rangle$$

In the proof of Theorem 4.32, $\mathbf{Q}$ is obtained by performing the Gram–Schmidt process on the columns of $\mathbf{A}$, so $r_{jj} > 0$ by Exercise 4.2.21.

4.5.21. $\|\mathbf{a}_j\| = \|\mathbf{A}\mathbf{e}_j\| = \left\|\mathbf{Q}\mathbf{R}\mathbf{e}_j\right\| = \|\mathbf{R}\mathbf{e_j}\| = \|\mathbf{r}_j\|$, where $\mathbf{r}_j$ is the jth column of $\mathbf{R}$, and $\|\mathbf{r}_j\| = \sqrt{\sum_{i=1}^{j} |r_{ij}|^2} \geq |r_{jj}|$.

4.5.22. Let $\mathbf{A} \in \mathrm{M}_n(\mathbb{C})$ be any nonzero matrix, and let $\mathbf{B} \in \mathrm{M}_{n,m}(\mathbb{C})$ be the matrix whose columns, in order, are those columns of $\mathbf{A}$ which do not lie in the span of the earlier columns. By the linear dependence lemma, the columns of $\mathbf{B}$ are linearly independent, so we may perform the Gram–Schmidt process on them to obtain a list of orthonormal vectors $(\mathbf{u}_1, \ldots, \mathbf{u}_m)$ in $\mathbb{C}^n$. We can extend this to an orthonormal basis $(\mathbf{q}_1, \ldots, \mathbf{q}_n)$ of $\mathbb{C}^n$, inserting the new vectors of this basis so that the original vectors $\mathbf{u}_1, \ldots, \mathbf{u}_m$ have the same positions in the new basis as the columns of $\mathbf{B}$ had in the original matrix $\mathbf{A}$.

Now let $\mathbf{Q} \in \mathrm{M}_n(\mathbb{C})$ be the matrix with jth column $\mathbf{q}_j$. Then $\mathbf{Q}$ is unitary, and we claim that $\mathbf{R} := \mathbf{Q}^*\mathbf{A}$ is upper triangular. It then follows that $\mathbf{A} = \mathbf{Q}\mathbf{R}$.

To prove the claim, as in the proof of Theorem 4.32, $r_{jk} = \mathbf{q}_j^* \mathbf{a}_k = \langle \mathbf{a}_k, \mathbf{q}_j \rangle$. If $\mathbf{a}_k$ is one of the columns in $\mathbf{B}$, say $\mathbf{b}_\ell$, then by the Gram–Schmidt process

$$\mathbf{a}_k = \mathbf{b}_\ell \in \langle \mathbf{u}_1, \ldots, \mathbf{u}_\ell \rangle \subseteq \langle \mathbf{q}_1, \ldots, \mathbf{q}_k \rangle .$$

If $\mathbf{a}_k$ is *not* one of the columns of $\mathbf{B}$, then it is in the span of the earlier columns of $\mathbf{A}$ which were included in $\mathbf{B}$, say $\mathbf{b}_1, \ldots, \mathbf{b}_\ell$, and so

$$\mathbf{a}_k \in \langle \mathbf{b}_1, \ldots, \mathbf{b}_\ell \rangle \in \langle \mathbf{u}_1, \ldots, \mathbf{u}_\ell \rangle \subseteq \langle \mathbf{q}_1, \ldots, \mathbf{q}_k \rangle ,$$

In either case, by the orthogonality of the $\mathbf{q}_j$, $r_{jk} = \langle \mathbf{a}_k, \mathbf{q}_j \rangle = 0$ if $j > k$.

4.5.23. If $\boldsymbol{T} : V \to W$ is an isometry and $\mathbb{F} = \mathbb{C}$, then

$$\begin{aligned}
\langle \boldsymbol{T}v_1, \boldsymbol{T}v_2 \rangle &= \frac{1}{4}\left(\|\boldsymbol{T}v_1 + \boldsymbol{T}v_2\|^2 - \|\boldsymbol{T}v_1 - \boldsymbol{T}v_2\|^2 + i\,\|\boldsymbol{T}v_1 + i\boldsymbol{T}v_2\|^2 - i\,\|\boldsymbol{T}v_1 - i\boldsymbol{T}v_2\|^2 \right) \\
&= \frac{1}{4}\left(\|\boldsymbol{T}(v_1 + v_2)\|^2 - \|\boldsymbol{T}(v_1 - v_2)\|^2 + i\,\|\boldsymbol{T}(v_1 + iv_2)\|^2 - i\,\|\boldsymbol{T}(v_1 - iv_2)\|^2 \right) \\
&= \frac{1}{4}\left(\|v_1 + v_2\|^2 - \|v_1 - v_2\|^2 + i\,\|v_1 + iv_2\|^2 - i\,\|v_1 - iv_2\|^2 \right) \\
&= \langle v_1, v_2 \rangle .
\end{aligned}$$

Chapter 5

Singular value decomposition and the spectral theorem

5.1 Singular value decomposition of linear maps

5.1.1. The vectors

$$(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4) = \left(\frac{1}{2}\begin{bmatrix}1\\1\\1\\1\end{bmatrix}, \frac{1}{2}\begin{bmatrix}1\\1\\-1\\-1\end{bmatrix}, \frac{1}{2}\begin{bmatrix}1\\-1\\1\\-1\end{bmatrix}, \frac{1}{2}\begin{bmatrix}1\\-1\\-1\\1\end{bmatrix} \right)$$

are orthonormal, hence they form an orthonormal basis of $\mathbb{R}^4$. We compute

$$\boldsymbol{T}\mathbf{v}_1 = \begin{bmatrix}1\\1\end{bmatrix}, \qquad \boldsymbol{T}\mathbf{v}_2 = \begin{bmatrix}1\\-1\end{bmatrix}, \qquad \boldsymbol{T}\mathbf{v}_3 = \boldsymbol{T}\mathbf{v}_4 = \mathbf{0}.$$

So we write $\mathbf{u}_1 = \frac{1}{\sqrt{2}}\begin{bmatrix}1\\1\end{bmatrix}$ and $\mathbf{u}_2 = \frac{1}{\sqrt{2}}\begin{bmatrix}1\\-1\end{bmatrix}$ (renormalizing the nonzero vectors above). Then $(\mathbf{u}_1, \mathbf{u}_2)$ is an orthonormal basis of $\mathbb{R}^2$ and $\boldsymbol{T}\mathbf{v}_1 = \sqrt{2}\mathbf{u}_1$, and $\boldsymbol{T}\mathbf{v}_2 = \sqrt{2}\mathbf{u}_2$.

Therefore $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4$ are right singular vectors of $\boldsymbol{T}$; $\mathbf{u}_1, \mathbf{u}_2$ are left singular vectors; and the singular values are $\sqrt{2}$, $\sqrt{2}$.

5.1.2. Observe that

$$\boldsymbol{T}\begin{bmatrix}1\\1\end{bmatrix} = \begin{bmatrix}1\\1\\0\end{bmatrix} \qquad \text{and} \qquad \boldsymbol{T}\begin{bmatrix}1\\-1\end{bmatrix} = \begin{bmatrix}1\\-1\\-2\end{bmatrix}$$

are orthogonal, and $\begin{bmatrix}1\\-1\\1\end{bmatrix}$ is orthogonal to both of these. Renormalizing, we have that

$$(\mathbf{v}_1, \mathbf{v}_2) = \left(\frac{1}{\sqrt{2}}\begin{bmatrix}1\\1\end{bmatrix}, \frac{1}{\sqrt{2}}\begin{bmatrix}1\\-1\end{bmatrix} \right)$$

is an orthonormal basis of $\mathbb{R}^2$,

$$(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3) = \left(\frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ -1 \\ -2 \end{bmatrix}, \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \right)$$

is an orthonormal basis of $\mathbb{R}^3$,

$$\boldsymbol{T}\mathbf{v}_1 = \mathbf{u}_1, \qquad \text{and} \qquad \boldsymbol{T}\mathbf{v}_2 = \sqrt{3}\mathbf{u}_2.$$

To have the singular values in decreasing order we reorder the bases slightly: the singular values are $\sqrt{3}$ and 1; the right singular vectors are $\mathbf{v}_2, \mathbf{v}_1$; and the left singular vectors are $\mathbf{u}_2, \mathbf{u}_1, \mathbf{v}_3$.

5.1.3. The range of $\boldsymbol{T}$ is the line $L = \left\langle \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\rangle$. So we start with orthonormal bases of L and of its orthogonal complement:

$$(\mathbf{u}_1, \mathbf{u}_2) = \left(\frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right).$$

If $\boldsymbol{R} \in \mathcal{L}(\mathbb{R}^2)$ is the *clockwise* rotation by $\pi/3$, then $\boldsymbol{T}\boldsymbol{R}\mathbf{u}_1 = \mathbf{u}_1$ and $\boldsymbol{T}\boldsymbol{R}\mathbf{u}_2 = \mathbf{0}$. So we let

$$\mathbf{v}_1 = \boldsymbol{R}\mathbf{u}_1 = \begin{bmatrix} \cos(-\pi/3) & -\sin(-\pi/3) \\ \sin(-\pi/3) & \cos(-\pi/3) \end{bmatrix} \mathbf{u}_1 = \begin{bmatrix} 1/2 & \sqrt{3}/2 \\ -\sqrt{3}/2 & 1/2 \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \frac{1}{2\sqrt{2}} \begin{bmatrix} 1+\sqrt{3} \\ 1-\sqrt{3} \end{bmatrix}$$

and

$$\mathbf{v}_2 = \boldsymbol{R}\mathbf{u}_2 = \begin{bmatrix} 1/2 & \sqrt{3}/2 \\ -\sqrt{3}/2 & 1/2 \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \frac{1}{2\sqrt{2}} \begin{bmatrix} 1-\sqrt{3} \\ -1-\sqrt{3} \end{bmatrix}$$

Since $\boldsymbol{R}$ is an isometry, $(\mathbf{v}_1, \mathbf{v}_2)$ is an orthonormal basis of $\mathbb{R}^2$, and as pointed out above, $\boldsymbol{T}\mathbf{v}_1 = \mathbf{u}_1$ and $\boldsymbol{T}\mathbf{v}_2 = \mathbf{0}$. So $\boldsymbol{T}$ has singular values 1 and 0, right singular vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ as above, and left singular vectors $\mathbf{u}_1$ and $\mathbf{u}_2$ as above.

5.1.4. The idea is to come up with an orthonormal basis of $\mathbb{R}^3$ whose vectors are sent to orthogonal vectors by $\boldsymbol{T}$. Firstly, the vector $\begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix}$ first gets rotated to $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, and then projected to $\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$, so we take this as one of our basis vectors. Similarly, $\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ is fixed by $\boldsymbol{T}$, so this seems a natural choice. Finally, we need a third vector perpendicular to the first two; taking $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ and applying $\boldsymbol{T}$ gives $\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$, which is perpendicular to the images of the first two. That is, we can just take $(\mathbf{e}_3, \mathbf{e}_1, -\mathbf{e}_2)$ as an orthonormal basis of the domain, and then the image of this basis under $\boldsymbol{T}$ is $(\mathbf{e}_3, \mathbf{e}_2, 0)$, and so the rank of $\boldsymbol{T}$ is 2 and $\sigma_1 = \sigma_2 = 1$.

5.1.5. The basis

$$\left(\frac{1}{\sqrt{2\pi}}, \frac{1}{\sqrt{\pi}}\sin(x), \frac{1}{\sqrt{\pi}}\sin(2x), \dots, \frac{1}{\sqrt{\pi}}\sin(nx), \frac{1}{\sqrt{\pi}}\cos(x), \frac{1}{\sqrt{\pi}}\cos(2x), \dots, \frac{1}{\sqrt{\pi}}\cos(nx)\right)$$

of V is orthonormal and $\boldsymbol{D}$ maps it to

$$\left(0, \frac{1}{\sqrt{\pi}}\cos(x), \frac{2}{\sqrt{\pi}}\cos(2x), \dots, \frac{n}{\sqrt{\pi}}\cos(nx), -\frac{1}{\sqrt{\pi}}\sin(x), -\frac{2}{\sqrt{\pi}}\sin(2x), \dots, -\frac{n}{\sqrt{\pi}}\sin(nx)\right).$$

The singular values are thus $\sigma_1 = \sigma_2 = n$, $\sigma_3 = \sigma_4 = n-1, \dots, \sigma_{2n-1} = \sigma_{2n} = 1, \sigma_{2n+1} = 0$. We can therefore take the left singular vectors to be

$$\left(\frac{1}{\sqrt{\pi}}\sin(nx), \frac{1}{\sqrt{\pi}}\cos(nx), \dots, \frac{1}{\sqrt{\pi}}\sin(x), \frac{1}{\sqrt{\pi}}\cos(x), \frac{1}{\sqrt{2\pi}}\right)$$

and the right singular vectors to be

$$\left(\frac{1}{\sqrt{\pi}}\sin(nx), -\frac{1}{\sqrt{\pi}}\cos(nx), \dots, \frac{1}{\sqrt{\pi}}\sin(x), -\frac{1}{\sqrt{\pi}}\cos(x), \frac{1}{\sqrt{2\pi}}\right).$$

5.1.6. By Theorem 4.16, there is an orthonormal basis $\mathcal{B}$ of V such that the matrix of the projection $\mathcal{B}$-coordinates is $\mathbf{diag}(1, \dots, 1, 0, \dots, 0)$, with the number of 1's equal to the dimension of the subspace onto which we are projecting. Choosing this basis as both $(e_1, \dots, e_n)$ and $(f_1, \dots, f_n)$ gives a singular value decomposition for $\boldsymbol{P}$; the singular values are all either zero or one.

5.1.7. Let $(e_1, \dots, e_n)$ be right singular vectors and $(f_1, \dots, f_n)$ be left singular vectors of $\boldsymbol{T}$. Then for any $v \in V$,

$$\boldsymbol{T}v = \boldsymbol{T}\left(\sum_{j=1}^{n} \langle v, e_j \rangle e_j\right) = \sum_{j=1}^{n} \langle v, e_j \rangle \boldsymbol{T}e_j = \sum_{j=1}^{n} \langle v, e_j \rangle \sigma_j f_j$$

and

$$\|v\|^2 = \sum_{j=1}^{n} |\langle v, e_j \rangle|^2$$

since $(e_1, \dots, e_n)$ is orthonormal, and then

$$\|\boldsymbol{T}v\|^2 = \left\|\sum_{j=1}^{n} \langle v, e_j \rangle \sigma_j f_j\right\|^2 = \sum_{j=1}^{n} |\langle v, e_j \rangle|^2 \sigma_j^2$$

since $(f_1, \dots, f_n)$ is orthonormal. Now if v is unit eigenvector of $\boldsymbol{T}$ with eigenvalue λ, then

$$|\lambda|^2 = \|\lambda v\|^2 = \|\boldsymbol{T}v\|^2 = \sum_{j=1}^{n} |\langle v, e_j \rangle|^2 \sigma_j^2 \leq \sigma_n^2 \sum_{j=1}^{n} |\langle v, e_j \rangle|^2 = \sigma_n^2$$

and

$$|\lambda|^2 = \|\lambda v\|^2 = \|\boldsymbol{T}v\|^2 = \sum_{j=1}^{n} |\langle v, e_j \rangle|^2 \sigma_j^2 \geq \sigma_1^2 \sum_{j=1}^{n} |\langle v, e_j \rangle|^2 = \sigma_1^2.$$

5.1.8. Suppose that $\boldsymbol{T}$ is invertible. Then $\boldsymbol{T}$ is an isomorphism of V and W, and so $\dim V = \dim W$. Moreover, since $\boldsymbol{T}$ is invertible, the rank of $\boldsymbol{T}$ is equal to $\dim V$, and so all of the σ_j are strictly positive.

Conversely, suppose that $\dim V = \dim W$ and that all the singular values of $\boldsymbol{T}$ are non-zero. If $(e_1, \ldots, e_n)$ and $(f_1, \ldots, f_n)$ are right- and left-singular vectors of $\boldsymbol{T}$, then we can define $\boldsymbol{S} : W \to V$ by taking $\boldsymbol{S} f_j = \frac{1}{\sigma_j} e_j$ and extending by linearity. Then $\boldsymbol{S}$ is an inverse to $\boldsymbol{T}$ since $\boldsymbol{ST} e_j = e_j$ for each j.

5.1.9. Since $\boldsymbol{T}$ is invertible, $\operatorname{rank} \boldsymbol{T} = n$. Let $(e_1, \ldots, e_n)$ be right singular vectors, $(f_1, \ldots, f_n)$ be left singular vectors, and $\sigma_1 \geq \cdots \geq \sigma_n > 0$ singular values of $\boldsymbol{T}$. Then $\boldsymbol{T} e_j = \sigma_j f_j$ for each j, so $\boldsymbol{T}^{-1} f_j = \sigma_j^{-1} e_j$ for each j. Therefore we can take $(f_n, \ldots, f_1)$ as right singular vectors of $\boldsymbol{T}^{-1}$, $(e_n, \ldots, e_1)$ as left singular vectors, and the singular values are $\sigma_n^{-1} \geq \cdots \geq \sigma_1^{-1} > 0$.

5.1.10. The solution to Exercise 5.1.9 shows that σ_n^{-1} is the largest singular value of $\boldsymbol{T}^{-1}$, and therefore $\left\|\boldsymbol{T}^{-1}\right\|_{op} = \sigma_n^{-1}$. Making the substitution $w = \boldsymbol{T} v$,

$$\left\|\boldsymbol{T}^{-1}\right\|_{op} = \max_{\|w\|=1} \left\|\boldsymbol{T}^{-1} w\right\| = \max_{w \neq 0} \frac{\left\|\boldsymbol{T}^{-1} w\right\|}{\|w\|} = \max_{v \neq 0} \frac{\|v\|}{\|\boldsymbol{T} v\|} = \max_{\|v\|=1} \frac{1}{\|\boldsymbol{T} v\|} = \left(\min_{\|v\|=1} \|\boldsymbol{T} v\|\right)^{-1}.$$

5.1.11. If all the singular values of $\boldsymbol{T}$ are 1, then $\boldsymbol{T}$ maps the orthonormal basis formed by its right singular vectors to the orthonormal basis formed by its left singular vectors. By Theorem 4.28, $\boldsymbol{T}$ is an isometry.

5.1.12. Let $(e_1, \ldots, e_n)$ and $(f_1, \ldots, f_m)$ be singular vectors of $\boldsymbol{T}$, so $\boldsymbol{T} e_1 = \sigma_1 f_1$ and $\boldsymbol{T} e_p = \sigma_p f_p$. For any $0 \leq t \leq 1$, let $v_t = t e_1 + \sqrt{1 - t^2} e_p$. Then

$$\|v_t\|^2 = t^2 + (1 - t^2) = 1,$$

$$\boldsymbol{T} v_t = t \sigma_1 f_1 + \sqrt{1 - t^2} \sigma_p f_p,$$

and

$$\|\boldsymbol{T} v_t\| = t^2 \sigma_1^2 + (1 - t^2) \sigma_p^2.$$

So as t goes from 0 to 1, $\|\boldsymbol{T} v_t\|$ goes continuously from σ_p to σ_1, so by the Intermediate Value Theorem from calculus, for each $s \in [\sigma_p, \sigma_1]$ there is a t such that $\|\boldsymbol{T} v_t\| = s$.

5.1.13. Fix any singular value decomposition of $\boldsymbol{T}$ with right singular vectors $(e_1, \ldots, e_n)$ and right singular vectors $(f_1, \ldots, f_m)$. We define:

- $\tau_1, \ldots, \tau_k$ are the distinct nonzero singular values of $\boldsymbol{T}$.
- For each $j = 1, \ldots, k$, V_j is the span of the right singular vectors that correspond to the singular value τ_j.
- V_0 is the span of the right singular vectors in $\ker \boldsymbol{T}$ (if there are any).
- For each $j = 1, \ldots, k$, $W_j = \boldsymbol{T}(V_j)$.
- $W_0 = (W_1 \oplus \cdots \oplus W_k)^{\perp}$.

- For each $j = 1, \ldots, k$, $\boldsymbol{T}_j : V_j \to W_j$ is defined by $\boldsymbol{T}_j v = \frac{1}{\tau_j} \boldsymbol{T}_j$.

Then:

- The spaces V_j are mutually orthogonal and span V because $(e_1, \ldots, e_n)$ is an orthonormal basis, so V is the orthogonal direct sum $V_0 \oplus V_1 \oplus \cdots \oplus V_k$.
- W_j is spanned by the left singular vectors corresponding to the singular value τ_j (or 0 if $j = 0$). It follows as above that the spaces W_j are mutually orthogonal, and so W is the orthogonal direct sum $W_0 \oplus W_1 \oplus \cdots \oplus W_k$.
- Each $\boldsymbol{T}_j$ is surjective by definition. If e_i is a right singular vector corresponding to τ_j, then $\boldsymbol{T}_j e_i = \frac{1}{\tau_j} \boldsymbol{T} e_i = f_i$. Therefore $\boldsymbol{T}_j$ maps the orthonormal basis of V_j consisting of the corresponding right singular vectors to the orthonormal basis of W_j consisting of the corresponding left singular vectors. So by Theorem 4.28, $\boldsymbol{T}_j$ is an isometry.
- Finally, if e_i is a right singular vector corresponding to the singular value τ_ℓ, then $\boldsymbol{P}_{V_\ell} e_i = e_i$ and $\boldsymbol{P}_{V_j} e_i = 0$ for $j \neq \ell$, so

$$\left(\sum_{j=1}^{k} \tau_j \boldsymbol{T}_j \boldsymbol{P}_{V_j} \right) e_i = \tau_\ell \boldsymbol{T}_\ell e_i = \boldsymbol{T} e_i.$$

  Similarly, if $e_i \in V_0$, then the expressions on each side of the above equation are 0. It follows from extension by linearity that

$$\sum_{j=1}^{k} \tau_j \boldsymbol{T}_j \boldsymbol{P}_{V_j} = \boldsymbol{T}.$$

5.1.14. Let $(f_1, \ldots, f_m)$ be corresponding left singular vectors of $\boldsymbol{T}$.

(a) If $v \in U_j$, then $v = \sum_{k=j}^{n} \langle v, e_k \rangle e_k$, so

$$\boldsymbol{T} v = \sum_{k=j}^{n} \langle v, e_k \rangle \boldsymbol{T} e_k = \sum_{k=j}^{n} \langle v, e_k \rangle \sigma_k f_k$$

and then

$$\|\boldsymbol{T} v\|^2 = \sum_{k=j}^{n} |\langle v, e_k \rangle|^2 \sigma_k^2 \leq \sigma_j^2 \sum_{k=j}^{n} |\langle v, e_k \rangle|^2 = \sigma_j^2 \|v\|^2 .$$

(b) If $v \in V_j$, then $v = \sum_{k=1}^{j} \langle v, e_k \rangle e_k$, so

$$\boldsymbol{T} v = \sum_{k=1}^{j} \langle v, e_k \rangle \boldsymbol{T} e_k = \sum_{k=1}^{j} \langle v, e_k \rangle \sigma_k f_k$$

and then

$$\|\boldsymbol{T} v\|^2 = \sum_{k=1}^{j} |\langle v, e_k \rangle|^2 \sigma_k^2 \geq \sigma_j^2 \sum_{k=1}^{j} |\langle v, e_k \rangle|^2 = \sigma_j^2 \|v\|^2 .$$

(c) $\dim V_j + \dim U = n + 1$, so by Lemma 3.22 there exists a nonzero vector $v \in V_j \cap U$. By part (b), any such vector satisfies $\|\boldsymbol{T}v\| \geq \sigma_j \|v\|$.

(d) By part (c), if U is any $(n-j+1)$-dimensional subspace of V, then $\max_{v \in U,\ \|v\|=1} \|\boldsymbol{T}v\| \geq \sigma_j$. Therefore

$$\min_{\dim U = n-j+1} \max_{v \in U,\ \|v\|=1} \|\boldsymbol{T}v\| \geq \sigma_j.$$

On the other hand, $\dim U_j = n - j + 1$, and by part (a), $\max_{v \in U_j,\ \|v\|=1} \|\boldsymbol{T}v\| \leq \sigma_j$. Therefore

$$\min_{\dim U = n-j+1} \max_{v \in U,\ \|v\|=1} \|\boldsymbol{T}v\| \leq \max_{v \in U_j,\ \|v\|=1} \|\boldsymbol{T}v\| \leq \sigma_j.$$

5.1.15. If $v \in V$, then $v = \sum_{j=1}^n \langle v, e_j \rangle e_j$, and so

$$\boldsymbol{T}v = \sum_{j=1}^n \langle v, e_j \rangle \boldsymbol{T}e_j = \sum_{j=1}^k \langle v, e_j \rangle \sigma_j f_j.$$

It follows immediately that $\operatorname{range} \boldsymbol{T} \subseteq \langle f_1, \ldots, f_k \rangle$. Moreover, since $f_j = \boldsymbol{T}(\sigma_j^{-1} e_j)$ for each $j = 1, \ldots, k$, it follows that $f_j \in \operatorname{range} \boldsymbol{T}$, so $\langle f_1, \ldots, f_k \rangle \subseteq \operatorname{range} \boldsymbol{T}$. Therefore $\operatorname{range} \boldsymbol{T} = \langle f_1, \ldots, f_k \rangle$, and since $(f_1, \ldots, f_m)$ is orthonormal, $(f_1, \ldots, f_k)$ is an orthonormal basis of $\operatorname{range} \boldsymbol{T}$, and therefore $\operatorname{rank} \boldsymbol{T} = k$.

5.2 Singular value decomposition of matrices

5.2.1. (a) Row-reducing, we find

$$\mathbf{A}^*\mathbf{A} - \lambda \mathbf{I}_2 = \begin{bmatrix} 5-\lambda & 4 \\ 4 & 5-\lambda \end{bmatrix} \to \begin{bmatrix} 4 & 5-\lambda \\ 5-\lambda & 4 \end{bmatrix} \to \begin{bmatrix} 4 & 5-\lambda \\ 0 & 4 - \frac{1}{4}(5-\lambda)^2 \end{bmatrix}$$

is singular iff $4 - \frac{1}{4}(5-\lambda)^2 = 0$, which is true iff $\lambda = 5 \pm 4 = 1, 9$. Therefore 1 and 9 are the eigenvalues of $\mathbf{A}^*\mathbf{A}$, so the singular values of $\mathbf{A}$ are 1 and 3.

(b) Here $\mathbf{A}^*\mathbf{A} = \begin{bmatrix} 20 & 16 \\ 16 & 20 \end{bmatrix}$. We found in part (a) that the eigenvalues of $\frac{1}{4}\mathbf{A}^*\mathbf{A}$ are 1 and 9, so the eigenvalues of $\mathbf{A}^*\mathbf{A}$ in this part are 4 and 36, and then the singular values of $\mathbf{A}$ are 2 and 6.

(c) Because it is smaller, in this case we work with $\mathbf{A}\mathbf{A}^*$.

$$\mathbf{A}\mathbf{A}^* - \lambda \mathbf{I}_2 = \begin{bmatrix} 6-\lambda & -2 \\ -2 & 6-\lambda \end{bmatrix} \rightsquigarrow \begin{bmatrix} -2 & 6-\lambda \\ 0 & -2 + \frac{1}{2}(6-\lambda)^2 \end{bmatrix}$$

is singular iff $-2 + \frac{1}{2}(6-\lambda)^2 = 0$, which is true iff $\lambda = 4, 8$. Therefore the singular values of $\mathbf{A}$ are 2 and $\sqrt{8} = 2\sqrt{2}$.

(d) In this case

$$\mathbf{A}\mathbf{A}^* = \begin{bmatrix} 6 & -2 & 0 \\ -2 & 6 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

where the upper-left 2×2 block is the same matrix we worked with in part (c). From this it follows that the singular values are the same as in part (c), with an additional 0.

(e) In this case $\mathbf{A}^*\mathbf{A} = \mathbf{diag}(1,4,1)$, which has eigenvalues 4, 1, and 1, so the singular values of $\mathbf{A}$ are 2, 1, and 1.

5.2.2. (a) $\mathbf{A}^*\mathbf{A} = 5\mathbf{I}_2$ has eigenvalues 5 and 5, so the singular values of $\mathbf{A}$ are $\sqrt{5}$ and $\sqrt{5}$.

(b) $\mathbf{A}\mathbf{A}^* = \mathbf{diag}(25,100)$ has eigenvalues 25 and 100, so the singular values of $\mathbf{A}$ are 5 and 10.

(c) $\mathbf{A}\mathbf{A}^* = \mathbf{diag}(1,2)$ has eigenvalues 1 and 2, so the singular values of $\mathbf{A}$ are 1 and $\sqrt{2}$.

(d) $\mathbf{A}\mathbf{A}^* = \mathbf{diag}(6,3)$ has eigenvalues 6 and 3, so the singular values of $\mathbf{A}$ are $\sqrt{6}$ and $\sqrt{3}$.

(e) $\mathbf{A}^*\mathbf{A} = \mathbf{diag}(9,1,4)$ has eigenvalues 9, 1, and 4, so the singular values of $\mathbf{A}$ are 3, 2, and 1.

5.2.3. Since $\mathbf{A}_z$ is triangular, its eigenvalues are the diagonal entries 1 and 2, regardless of the value of z. On the other hand, if σ_1 and σ_2 are the singular values of $\mathbf{A}$, then by Proposition 5.6,

$$\sigma_1^2 + \sigma_2^2 = \|\mathbf{A}_z\|^2 = 5 + |z|^2 .$$

Since the right side of this depends on z, the left side must as well.

5.2.4. Let $\mathcal{B}$ and $\mathcal{C}$ be the given orthonormal bases of V and W, so that $\mathbf{A} = [\boldsymbol{T}]_{\mathcal{B},\mathcal{C}}$, and let $\mathbf{A} = \sum_{j=1}^{p} \sigma_p \mathbf{u}_j \mathbf{v}_j^*$ be an SVD for $\mathbf{A}$. If necessary, add additional vectors so that $(\mathbf{u}_1, \ldots, \mathbf{u}_m)$ and $(\mathbf{v}_1, \ldots, \mathbf{v}_n)$ are orthonormal bases of $\mathbb{F}^m$ and $\mathbb{F}^n$, respectively. Then $\mathbf{A}\mathbf{v}_j = \sigma_j \mathbf{u}_j$ for $1 \le j \le p$ and $\mathbf{A}\mathbf{v}_j = \mathbf{0}$ for $j > p$.

Now define $e_j \in V$ to be the vector with $[e_j]_{\mathcal{B}} = \mathbf{v}_j$, and define $f_j \in V$ to be the vector with $[f_j]_{\mathcal{C}} = \mathbf{u}_j$. Then $(e_1, \ldots, e_n)$ and $(f_1, \ldots, f_m)$ are orthonormal bases of V and W respectively (see the remark after the statement of Theorem 4.10), and $\mathbf{A}\mathbf{v}_j = \sigma_j \mathbf{u}_j$ implies that $\boldsymbol{T} f_j = \sigma_j e_j$. Therefore the σ_j are the singular values of $\boldsymbol{T}$.

5.2.5. If $\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^*$ is an SVD for $\mathbf{A}$, then $\mathbf{A}^* = \mathbf{V}\boldsymbol{\Sigma}^{\mathrm{T}}\mathbf{U}^*$ is an SVD for $\mathbf{A}^*$, and for $1 \le j \le \min\{m,n\}$, $[\boldsymbol{\Sigma}^{\mathrm{T}}]_{jj} = \sigma_j$.

5.2.6. Let $\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^*$ be an SVD of $\mathbf{A}$. Then $\mathbf{A}^* = \mathbf{V}\boldsymbol{\Sigma}^*\mathbf{U}^*$, which is an SVD of $\mathbf{A}^*$, and the number of non-zero entries in Σ^* (which is necessarily the rank of $\mathbf{A}^*$) is the same as the number of non-zero entries in Σ.

5.2.7. Suppose $\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^*$ is an SVD for $\mathbf{A}$. Then $\mathbf{A}\mathbf{A}^* = \mathbf{U}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^*\mathbf{U}^*$ is similar to $\boldsymbol{\Sigma}\boldsymbol{\Sigma}^* = \mathbf{diag}(\sigma_1^2, \ldots, \sigma_r^2, 0, \ldots, 0)$, which has rank $r = \operatorname{rank} \mathbf{A}$. The proof for $\mathbf{A}^*\mathbf{A}$ is similar.

5.2.8. For each j, define

$$\omega_j = \begin{cases} \frac{\lambda_j}{|\lambda_j|} & \text{if } \lambda_j \ne 0, \\ 1 & \text{if } \lambda_j = 0, \end{cases}$$

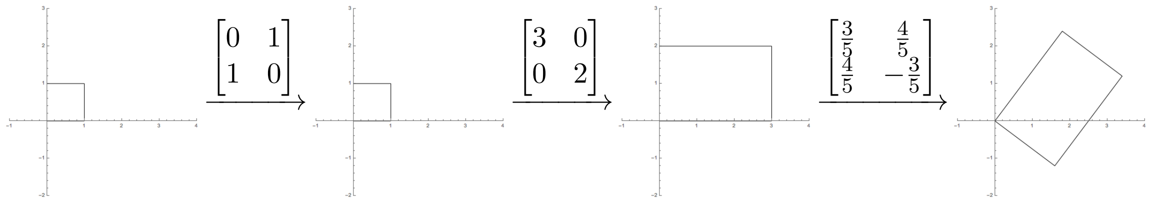
so that $\omega_j |\lambda_j| = \lambda_j$ for each j. Then

$$\mathbf{diag}(\omega_1, \ldots, \omega_n)$$

is unitary, so

$$\mathbf{A} = \mathbf{diag}(\omega_1, \ldots, \omega_n)\,\mathbf{diag}(|\lambda_1|, \ldots, |\lambda_n|)\mathbf{I}_n$$

is in SVD form except for the order. Since a reordering of coordinates is an isometry (see Example 4 on p. 280), $|\lambda_1|, \dots, |\lambda_n|$ are, up to order, the singular values of $\mathbf{A}$.

5.2.9. 

5.2.10. Let $\mathbf{A} = \sum_{j=1}^r \sigma_j \mathbf{u}_j \mathbf{v}_j^*$ be an SVD for $\mathbf{A}$. Then

$$\begin{aligned} |\operatorname{tr}(\mathbf{A})| &= \left| \sum_{j=1}^r \sigma_j \operatorname{tr}(\mathbf{u}_j \mathbf{v}_j^*) \right| \le \sum_{j=1}^r \sigma_j \left| \operatorname{tr}(\mathbf{u}_j \mathbf{v}_j^*) \right| = \sum_{j=1}^r \sigma_j \left| \operatorname{tr}(\mathbf{v}_j^* \mathbf{u}_j) \right| \\ &= \sum_{j=1}^r \sigma_j \left| \langle \mathbf{u}_j, \mathbf{v}_j \rangle \right| \le \sum_{j=1}^r \sigma_j, \end{aligned}$$

where the first inequality is by the triangle inequality, the next equality is because $\operatorname{tr}(\mathbf{AB}) = \operatorname{tr}(\mathbf{BA})$, the next uses that $\mathbf{v}_j^* \mathbf{u}_j = \langle \mathbf{u}_j, \mathbf{v}_j \rangle$, and the final inequality is by the Cauchy–Schwarz inequality and the fact that $\mathbf{u}_j$ and $\mathbf{v}_j$ are unit vectors.

5.2.11. $\kappa(\mathbf{A}) = \|\mathbf{A}\|_{op} \left\|\mathbf{A}^{-1}\right\|_{op}$; $\|\mathbf{A}\|_{op} = \sigma_1$ and $\left\|\mathbf{A}^{-1}\right\|_{op} = \sigma_n^{-1}$ by Exercise 5.1.9.

5.2.12. If $\mathbf{A} \in \mathrm{M}_{m,n}(\mathbb{R})$, then $\mathbf{A}$ has an SVD $\mathbf{A} = \sum_{j=1}^r \sigma_r \mathbf{u}_j \mathbf{v}_j^*$ with $\mathbf{u}_j \in \mathbb{R}^m$ and $\mathbf{v}_j \in \mathbb{R}^n$. Since $\sigma_j \in \mathbb{R}$, the best rank k approximation of $\mathbf{A}$

$$\mathbf{B} = \sum_{j=1}^k \sigma_j \mathbf{u}_j \mathbf{v}_j^* \in \mathrm{M}_{m,n}(\mathbb{R}).$$

5.2.13. Since $\mathbf{A}$ is invertible, it has rank n and therefore singular values $\sigma_1 \ge \cdots \ge \sigma_n > 0$. A matrix $\mathbf{B}$ is singular if and only if $\operatorname{rank} \mathbf{B} < n$. By Theorem 5.8, the minimum value of $\|\mathbf{A} - \mathbf{B}\|_{op}$ for $\mathbf{B}$ with rank $k < n$ is σ_{k+1}. Therefore the smallest value of $\|\mathbf{A} - \mathbf{B}\|_{op}$ for all matrices $\mathbf{B}$ with $\operatorname{rank} \mathbf{B} < n$ is σ_n. The statement for the Frobenius norm follows in the same way, using Theorem 5.9 instead of Theorem 5.8.

5.2.14. (a) Let $\sigma_1 \ge \cdots \ge \sigma_r$ be the non-zero singular values of $\mathbf{A}$. Then by Proposition 5.6 and the comment immediately above it,

$$\operatorname{srank} \mathbf{A} = \frac{\|\mathbf{A}\|_F^2}{\|\mathbf{A}\|_{op}^2} = \frac{\sum_{j=1}^r \sigma_j^2}{\sigma_1^2} \le \frac{\sum_{j=1}^r \sigma_1^2}{\sigma_1^2} = r,$$

since $\sigma_j^2 \le \sigma_1^2$ for each j.

(b) Note first that $k \ge \alpha \operatorname{srank} \mathbf{A} = \frac{\alpha \|\mathbf{A}\|_F^2}{\|\mathbf{A}\|_{op}^2}$ if and only if $\frac{1}{\|\mathbf{A}\|_{op}} \le \frac{1}{\|\mathbf{A}\|_F} \sqrt{\frac{k}{\alpha}}$. By Theorem 5.8, if $\mathbf{B}$ is the best rank k approximation of $\mathbf{A}$, then $\|\mathbf{A} - \mathbf{B}\|_{op} = \sigma_{k+1}$, and so

$$\frac{\|\mathbf{A} - \mathbf{B}\|_{op}}{\|\mathbf{A}\|_{op}} \le \frac{1}{\sqrt{\alpha}} \sqrt{\frac{k \sigma_{k+1}^2}{\sum_{j=1}^r \sigma_j^2}} \le \frac{1}{\sqrt{\alpha}},$$

since $k\sigma_{k+1}^2 \le \sum_{j=1}^{k} \sigma_j^2$.

5.2.15. (a) If $\mathbf{A}$ is invertible, then $m = n$ and $\operatorname{rank}\mathbf{A} = n$ so all the singular values of $\mathbf{A}$ are positive. Therefore $\mathbf{\Sigma}^\dagger = \mathbf{diag}(\sigma_1^{-1}, \dots, \sigma_n^{-1}) = \mathbf{\Sigma}^{-1}$, and so

$$\mathbf{A}^\dagger\mathbf{A} = \mathbf{V}\mathbf{\Sigma}^\dagger\mathbf{U}^*\mathbf{U}\mathbf{\Sigma}\mathbf{V}^* = \mathbf{V}\mathbf{Sigma}^\dagger\mathbf{\Sigma}\mathbf{V}^* = \mathbf{V}\mathbf{I}_n\mathbf{V}^* = \mathbf{I}_n.$$

(b) $\mathbf{\Sigma}\mathbf{\Sigma}^\dagger$ is the $m \times m$ matrix with 1 in the (j, j) entry for $1 \le j \le r$ and zeroes otherwise; $\mathbf{\Sigma}^\dagger\mathbf{\Sigma}$ is the $n \times n$ matrix with 1 in the (j, j) entry for $1 \le j \le r$ and zeroes otherwise.

(c) $\mathbf{A}\mathbf{A}^\dagger = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^*\mathbf{V}\mathbf{\Sigma}^\dagger\mathbf{U}^* = \mathbf{U}\mathbf{\Sigma}\mathbf{\Sigma}^\dagger\mathbf{U}^*$. By part (b), $\left(\mathbf{\Sigma}\mathbf{\Sigma}^\dagger\right)^* = \mathbf{\Sigma}\mathbf{\Sigma}^\dagger$, and therefore

$$\left(\mathbf{A}\mathbf{A}^\dagger\right)^* = \left(\mathbf{U}\mathbf{\Sigma}\mathbf{\Sigma}^\dagger\mathbf{U}^*\right)^* = \mathbf{U}\left(\mathbf{\Sigma}\mathbf{\Sigma}^\dagger\right)^*\mathbf{U}^* = \mathbf{U}\mathbf{\Sigma}\mathbf{\Sigma}^\dagger\mathbf{U}^* = \mathbf{A}\mathbf{A}^\dagger.$$

The other statement follows similarly.

(d) It follows from part (b) that $\mathbf{\Sigma}\mathbf{\Sigma}^\dagger\mathbf{\Sigma} = \mathbf{\Sigma}$. Therefore

$$\mathbf{A}\mathbf{A}^\dagger\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^*\mathbf{V}\mathbf{\Sigma}^\dagger\mathbf{U}^*\mathbf{U}\mathbf{\Sigma}\mathbf{V}^* = \mathbf{U}\mathbf{\Sigma}\mathbf{\Sigma}^\dagger\mathbf{Sigma}\mathbf{V}^* = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^* = \mathbf{A}.$$

The other statement follows similarly.

5.2.16. (a) If $\mathbf{A}\mathbf{A}^\dagger\mathbf{b} = \mathbf{b}$, then $\mathbf{x} = \mathbf{A}^\dagger\mathbf{b}$ is a solution of $\mathbf{A}\mathbf{x} = \mathbf{b}$.
If $\mathbf{A}\mathbf{x} = \mathbf{b}$ has any solution $\mathbf{x}$, then by part (d) of Exercise 5.2.15,

$$\mathbf{b} = \mathbf{A}\mathbf{x} = \left(\mathbf{A}\mathbf{A}^\dagger\mathbf{A}\right)\mathbf{x} = \left(\mathbf{A}\mathbf{A}^\dagger\right)(\mathbf{A}\mathbf{x}) = \mathbf{A}\mathbf{A}^\dagger\mathbf{b}.$$

(b) If $\mathbf{A}\mathbf{x} = \mathbf{b}$, then

$$\left\|\mathbf{A}^\dagger\mathbf{b}\right\| = \left\|\mathbf{A}^\dagger\mathbf{A}\mathbf{x}\right\| = \left\|\mathbf{U}\mathbf{\Sigma}\mathbf{\Sigma}^\dagger\mathbf{U}^*\mathbf{x}\right\| = \left\|\mathbf{\Sigma}\mathbf{\Sigma}^\dagger\mathbf{y}\right\|,$$

where $\mathbf{y} = \mathbf{U}^*\mathbf{x}$, since $\mathbf{U}$ is unitary. Part (b) of Exercise 5.2.15 shows that $\mathbf{\Sigma}\mathbf{\Sigma}^\dagger$ represents the orthogonal projection onto the first r coordinates, so

$$\left\|\mathbf{\Sigma}\mathbf{\Sigma}^\dagger\mathbf{y}\right\| \le \|\mathbf{y}\| = \|\mathbf{x}\|.$$

(c) Suppose that $\mathbf{x} \in C\big(\mathbf{I}_n - \mathbf{A}^\dagger\mathbf{A}\big)$. Then $\mathbf{x} = (\mathbf{I}_n - \mathbf{A}^\dagger\mathbf{A})\mathbf{y}$ for some $\mathbf{y}$, so

$$\mathbf{A}\mathbf{x} = \mathbf{A}\mathbf{y} - \mathbf{A}\mathbf{A}^\dagger\mathbf{A}\mathbf{y} = \mathbf{A}\mathbf{y} - \mathbf{A}\mathbf{y} = \mathbf{0}$$

by part (d) of Exercise 5.2.15. Therefore $\ker\mathbf{A} \subseteq C\big(\mathbf{I}_n - \mathbf{A}^\dagger\mathbf{A}\big)$.
By part (b) of Exercise 5.2.15 and the Rank–Nullity Theorem,

$$\operatorname{rank}(\mathbf{I}_n - \mathbf{A}^\dagger\mathbf{A}) = \operatorname{rank}\left(\mathbf{U}(\mathbf{I}_n - \Sigma^\dagger\mathbf{\Sigma})\mathbf{U}^*\right) = \operatorname{rank}\left(\mathbf{I}_n - \mathbf{\Sigma}^\dagger\mathbf{\Sigma}\right) = n - r = \operatorname{null}\mathbf{A}.$$

Therefore we must have $\ker\mathbf{A} = C\big(\mathbf{I}_n - \mathbf{A}^\dagger\mathbf{A}\big)$.
We know that if $\mathbf{A}\mathbf{x} = \mathbf{b}$ is consistent, then its set of solutions is

$$\{\mathbf{x} + \mathbf{y} \mid \mathbf{y} \in \ker\mathbf{A}\}$$

for any given solution $\mathbf{x}$. The given form now follows by using $\mathbf{x} = \mathbf{A}^\dagger\mathbf{b}$ and the fact proved above that

$$\ker\mathbf{A} = \left\{(\mathbf{I}_n - \mathbf{A}^\dagger\mathbf{A})\mathbf{w} \;\middle|\; \mathbf{w} \in \mathbb{C}^n\right\}.$$

5.2.17. (a) By Theorem 4.19, the desired point is the orthogonal projection of $\begin{bmatrix}1\\1\end{bmatrix}$ onto $\left\langle \begin{bmatrix}2\\1\end{bmatrix} \right\rangle$, which is

$$\frac{\left\langle \begin{bmatrix}1\\1\end{bmatrix}, \begin{bmatrix}2\\1\end{bmatrix} \right\rangle}{\left\| \begin{bmatrix}2\\1\end{bmatrix} \right\|^2} \begin{bmatrix}2\\1\end{bmatrix} = \frac{3}{5} \begin{bmatrix}2\\1\end{bmatrix}.$$

(b) Write a general point on the line $y = \frac{1}{2}x$ as $\left(t, \frac{t}{2}\right)$ for $t \in \mathbb{R}$. We consider separate cases:

- $t \leq 1$: $\left\|\left(t-1, \frac{t}{2}-1\right)\right\|_1 = (1-t) + \left(1 - \frac{t}{2}\right) = 2 - \frac{3}{2}t$. This is minimized when t is largest, that is, when $t = 1$. The distance in that case is $\frac{1}{2}$.
- $1 \leq t \leq 2$: $\left\|\left(t-1, \frac{t}{2}-1\right)\right\|_1 = (t-1) + \left(1 - \frac{t}{2}\right) = \frac{1}{2}t$. This is minimized when t is smallest, that is, when $t = 1$ as in the previous case.
- $2 \leq t$: $\left\|\left(t-1, \frac{t}{2}-1\right)\right\|_1 = (t-1) + \left(\frac{t}{2} - 1\right) = \frac{3}{2}t - 2$. This is minimized when t is smallest, that is, when $t = 2$. The distance in that case is 1.

It follows that the point on the line with smallest ℓ^1 distance to $(1,1)$ occurs when $t = 1$, so it is $(1, 1/2)$.

(c) Write a general point on the line $y = \frac{1}{2}x$ as $\left(t, \frac{t}{2}\right)$ for $t \in \mathbb{R}$. We consider separate cases:

- $t \leq 0$: $\left\|\left(t-1, \frac{t}{2}-1\right)\right\|_\infty = \max\left\{1-t, 1-\frac{t}{2}\right\} = 1-t$. This is minimized when t is largest, that is, when $t = 0$. The distance in that case is 1.
- $0 \leq t \leq 1$: $\left\|\left(t-1, \frac{t}{2}-1\right)\right\|_\infty = \max\left\{1-t, 1-\frac{t}{2}\right\} = 1-\frac{t}{2}$. This is minimized when t is largest, that is, when $t = 1$. The distance in that case is $1/2$.
- $1 \leq t \leq 4/3$: $\left\|\left(t-1, \frac{t}{2}-1\right)\right\|_1 = \max\left\{t-1, 1-\frac{t}{2}\right\} = 1-\frac{t}{2}$. This is minimized when t is largest, that is, when $t = 4/3$. The distance in that case is $1/3$.
- $4/3 \leq t \leq 2$: $\left\|\left(t-1, \frac{t}{2}-1\right)\right\|_1 = \max\left\{t-1, 1-\frac{t}{2}\right\} = t-1$. This is minimized when t is smallest, that is, when $t = 4/3$. The distance in that case is $1/3$.
- $2 \leq t$: $\left\|\left(t-1, \frac{t}{2}-1\right)\right\|_1 = \max\left\{t-1, \frac{t}{2}-1\right\} = t-1$. This is minimized when t is smallest, that is, when $t = 2$. The distance in that case is 1.

It follows that the point on the line with smallest ℓ^∞ distance to $(1,1)$ occurs when $t = 4/3$, so it is $(4/3, 2/3)$.

5.2.18. (a) For simplicity we write $V = V_W$. Then $\mathbf{A} = \boldsymbol{P}_V\mathbf{A} + \boldsymbol{P}_{V^\perp}\mathbf{A}$, so

$$\mathbf{A}\mathbf{x} = \left[\boldsymbol{P}_V\mathbf{A}\right]\mathbf{x} + \left[\boldsymbol{P}_{V^\perp}\mathbf{A}\right]\mathbf{x}.$$

We claim that $\left[\boldsymbol{P}_V\mathbf{A}\right]\mathbf{x} \in W$ and that $\left[\boldsymbol{P}_{V^\perp}\mathbf{A}\right]\mathbf{x} \in W^\perp$, which implies that $\left[\boldsymbol{P}_V\mathbf{A}\right]\mathbf{x} = \boldsymbol{P}_W(\mathbf{A}\mathbf{x})$.

By the definition of V, since $\boldsymbol{P}_V\mathbf{A} \in V$ we have that $C(\boldsymbol{P}_V\mathbf{A}) \subseteq W$, and therefore $\left[\boldsymbol{P}_V\mathbf{A}\right]\mathbf{x} \in W$.

Write $\boldsymbol{P}_{V^\perp}\mathbf{A} = \mathbf{C}$, and let $\mathbf{w} \in W$. Then

$$\langle \mathbf{w}, \mathbf{C}\mathbf{x} \rangle = \mathbf{C}\mathbf{x}^*\mathbf{w} = \mathbf{x}^*\mathbf{C}^*\mathbf{w} = \operatorname{tr}\left(\mathbf{x}^*\mathbf{C}^*\mathbf{w}\right) = \operatorname{tr}\left(\mathbf{w}\mathbf{x}^*\mathbf{C}^*\right) = \langle \mathbf{w}\mathbf{x}^*, \mathbf{C} \rangle_F.$$

Now $C(\mathbf{w}\mathbf{x}^*) \subseteq \langle \mathbf{w} \rangle \subseteq W$, so $\mathbf{w}\mathbf{x}^* \in V$. Since $\mathbf{C} \in V^\perp$, this implies that $\langle \mathbf{w}, \mathbf{C}\mathbf{x} \rangle = 0$, so $\mathbf{C}\mathbf{x} \in W^\perp$, which completes the proof of the claim.

Alternatively, this follows from the calculation of $\boldsymbol{P}_{V_W}\mathbf{A}$ on page 307.

(b) Let $W = \langle \mathbf{u}_1, \dots, \mathbf{u}_k \rangle$. Then the matrix of the orthogonal projection $\boldsymbol{P}_W$ is $\sum_{j=1}^k \mathbf{u}_j \mathbf{u}_j^*$, so for any $\mathbf{x} \in \mathbb{C}^n$,

$$\boldsymbol{P}_W(\mathbf{A}\mathbf{x}) = \sum_{\ell=1}^k \mathbf{u}_\ell \mathbf{u}_\ell^* \sum_{j=1}^p \sigma_j \mathbf{u}_j \mathbf{v}_j^* \mathbf{x} = \sum_{j=1}^p \sigma_j \sum_{\ell=1}^k \mathbf{u}_\ell \mathbf{u}_\ell^* \mathbf{u}_j \mathbf{v}_j^* \mathbf{x} = \sum_{j=1}^k \sigma_j \mathbf{u}_j \mathbf{v}_j^* \mathbf{x}$$

since the $\mathbf{u}_j$ are orthonormal. By part (a), this implies that

$$\big[\boldsymbol{P}_{V_W}\mathbf{A}\big]\mathbf{x} = \boldsymbol{P}_W(\mathbf{A}\mathbf{x}) = \sum_{j=1}^k \sigma_j \mathbf{u}_j \mathbf{v}_j^* \mathbf{x}$$

for every $\mathbf{x} \in \mathbb{C}^n$, and therefore $\boldsymbol{P}_{V_W}\mathbf{A} = \sum_{j=1}^k \sigma_j \mathbf{u}_j \mathbf{v}_j^*$.

5.2.19. Since the $\mathbf{u}_j$ and the $\mathbf{v}_j$ are orthonormal,

$$\begin{aligned} \|\mathbf{A}\|_F^2 = \left\| \sum_{j=1}^r \sigma_j \mathbf{u}_j \mathbf{v}_j^* \right\|_F^2 &= \operatorname{tr}\left(\sum_{j,k=1}^r \sigma_j \sigma_k \mathbf{u}_j \mathbf{v}_j^* \mathbf{v}_k \mathbf{u}_k^* \right) = \sum_{j,k=1}^r \sigma_j \sigma_k \operatorname{tr}\big(\mathbf{u}_j \mathbf{v}_j^* \mathbf{v}_k \mathbf{u}_k^*\big) \\ &= \sum_{j,k=1}^r \sigma_j \sigma_k \operatorname{tr}\big(\mathbf{u}_k^* \mathbf{u}_j \mathbf{v}_j^* \mathbf{v}_k\big) = \sum_{j,k=1}^r \sigma_j \sigma_k \langle \mathbf{u}_j, \mathbf{u}_k \rangle \langle \mathbf{v}_k, \mathbf{v}_j \rangle = \sum_{j=1}^r \sigma_j^2. \end{aligned}$$

5.2.20. Let $\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^*$ be an SVD for $\mathbf{A}$ as in Theorem 5.4, and let $\mathbf{u}_j$ and $\mathbf{v}_j$ denote the columns of $\mathbf{U}$ and $\mathbf{V}$ respectively. Then for each $k = 1, \dots, n$, $\mathbf{V}^*\mathbf{v}_k$ is the kth column of $\mathbf{V}^*\mathbf{V} = \mathbf{I}_n$, so $\mathbf{V}^*\mathbf{v}_k = \mathbf{e}_k$. Therefore

$$\mathbf{A}\mathbf{v}_k = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^*\mathbf{v}_k = \mathbf{U}\boldsymbol{\Sigma}\mathbf{e}_k = \sigma_k \mathbf{U}\mathbf{e}_k = \sigma_k \mathbf{e}_k$$

for $1 \le k \le r$, and similarly $\mathbf{A}\mathbf{v}_k = \mathbf{0}$ if $k > r$.

On the other hand,

$$\left(\sum_{j=1}^r \sigma_j \mathbf{u}_j \mathbf{v}_j^* \right) \mathbf{v}_k = \sigma_k \mathbf{u}_k$$

if $1 \le k \le r$ since the $\mathbf{v}_j$ are orthonormal, and the expression above is $\mathbf{0}$ if $k > r$. Therefore

$$\mathbf{A}\mathbf{v}_k = \left(\sum_{j=1}^r \sigma_j \mathbf{u}_j \mathbf{v}_j^* \right) \mathbf{v}_k$$

for each k. Since $(\mathbf{v}_1, \dots, \mathbf{v}_n)$ is a basis of $\mathbb{F}^n$, this implies that

$$\mathbf{A} = \sum_{j=1}^r \sigma_j \mathbf{u}_j \mathbf{v}_j^*.$$

5.2.21. By assumption, $\mathbf{A} - \mathbf{B} = \sum_{j=k+1}^{r} \sigma_j \mathbf{u}_j \mathbf{v}_j^*$. We can read off from this that σ_{k+1} is the largest singular value of $\mathbf{A} - \mathbf{B}$, and so

$$\|\mathbf{A} - \mathbf{B}\|_{op} = \sigma_{k+1}.$$

5.2.22. Since $C(\mathbf{0}) = \{\mathbf{0}\} \subseteq W$, $\mathbf{0} \in V_W$.

If $\mathbf{A}, \mathbf{B} \in V_W$, then every column of $\mathbf{A}$ and every column of $\mathbf{B}$ lies in W. Since W is a subspace, this implies that every column of $\mathbf{A} + \mathbf{B}$ lies in W, so $\mathbf{A} + \mathbf{B} \in V_W$.

If $\mathbf{A} \in V_W$ and $c \in \mathbb{F}$, then every column of $\mathbf{A}$ lies in W, and since W is a subspace, every column of $c\mathbf{A}$ lies in W. Therefore $c\mathbf{A} \in V_W$.

This proves that V_W is a subspace.

If $\mathbf{B} \in V_W$, then each column $\mathbf{b}_\ell \in W$ can be written as

$$\mathbf{b}_\ell = \sum_{j=1}^{k} \langle \mathbf{b}_\ell, \mathbf{w}_j \rangle \mathbf{w}_j,$$

and so

$$\mathbf{B} = \sum_{j=1}^{k} \sum_{\ell=1}^{n} \langle \mathbf{b}_\ell, \mathbf{w}_j \rangle \mathbf{W}_{j,\ell}.$$

Therefore the $\mathbf{W}_{j,\ell}$ span V_W. Furthermore, $\langle \mathbf{W}_{j_1,\ell_1}, \mathbf{W}_{j_2,\ell_2} \rangle_F = 0$ immediately if $\ell_1 \neq \ell_2$; if $\ell_1 = \ell_2$, then it is $\langle \mathbf{w}_{j_1}, \mathbf{w}_{j_2} \rangle$, which is 0 if $j_1 \neq j_2$ and is 1 if $j_1 = j_2$. Therefore the $\mathbf{W}_{j,\ell}$ are orthonormal, and hence form an orthonormal basis of V_W.

5.3 Adjoint maps

5.3.1. For all these matrices the singular values were found in the corresponding parts of Exercise 5.2.1.

(a) $\mathbf{A}^*\mathbf{A} = \begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix}$ has eigenvalues 9 and 1. Then

$$\mathrm{Eig}_9(\mathbf{A}^*\mathbf{A}) = \ker \begin{bmatrix} -4 & 4 \\ 4 & -4 \end{bmatrix} = \left\langle \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\rangle = \left\langle \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\rangle = \langle \mathbf{v}_1 \rangle$$

and

$$\mathrm{Eig}_9(\mathbf{A}^*\mathbf{A}) = \ker \begin{bmatrix} 4 & 4 \\ 4 & 4 \end{bmatrix} = \left\langle \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right\rangle = \left\langle \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right\rangle = \langle \mathbf{v}_2 \rangle .$$

Then

$$\mathbf{u}_1 = \frac{1}{3}\mathbf{A}\mathbf{v}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

and

$$\mathbf{u}_2 = \mathbf{A}\mathbf{v}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 1 \end{bmatrix}.$$

So finally we have the SVD $\mathbf{U} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$, $\mathbf{\Sigma} = \mathbf{diag}(3, 1)$, $\mathbf{V} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$.

(b) $\mathbf{A}^*\mathbf{A} = \begin{bmatrix} 20 & 16 \\ 16 & 20 \end{bmatrix}$ has eigenvalues 36 and 4. Then

$$\mathrm{Eig}_{36}(\mathbf{A}^*\mathbf{A}) = \ker \begin{bmatrix} -16 & 16 \\ 16 & -16 \end{bmatrix} = \left\langle \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\rangle = \left\langle \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\rangle = \langle \mathbf{v}_1 \rangle$$

and

$$\mathrm{Eig}_{4}(\mathbf{A}^*\mathbf{A}) = \ker \begin{bmatrix} 16 & 16 \\ 16 & 16 \end{bmatrix} = \left\langle \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right\rangle = \left\langle \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right\rangle = \langle \mathbf{v}_2 \rangle .$$

Then

$$\mathbf{u}_1 = \frac{1}{6}\mathbf{A}\mathbf{v}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} i \\ i \end{bmatrix}$$

and

$$\mathbf{u}_2 = \frac{1}{2}\mathbf{A}\mathbf{v}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

So finally we have the SVD $\mathbf{U} = \frac{1}{\sqrt{2}} \begin{bmatrix} i & 1 \\ i & -1 \end{bmatrix}$, $\mathbf{\Sigma} = \mathbf{diag}(6, 2)$, $\mathbf{V} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$.

(c) In this case we apply Algorithm 5.17 to $\mathbf{A}^*$ instead of $\mathbf{A}$, since $\mathbf{A}\mathbf{A}^*$ is smaller than $\mathbf{A}^*\mathbf{A}$. (See also Exercise 5.3.22.)

$\mathbf{A}\mathbf{A}^* = \begin{bmatrix} 6 & -2 \\ -2 & 6 \end{bmatrix}$ has eigenvalues 8 and 4. Then

$$\mathrm{Eig}_{8}(\mathbf{A}\mathbf{A}^*) = \ker \begin{bmatrix} -2 & -2 \\ -2 & -2 \end{bmatrix} = \left\langle \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right\rangle = \left\langle \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right\rangle = \langle \mathbf{v}_1 \rangle$$

and

$$\mathrm{Eig}_{4}(\mathbf{A}\mathbf{A}^*) = \ker \begin{bmatrix} 2 & -2 \\ -2 & 2 \end{bmatrix} = \left\langle \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\rangle = \left\langle \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\rangle = \langle \mathbf{v}_2 \rangle .$$

Then

$$\mathbf{u}_1 = \frac{1}{\sqrt{8}}\mathbf{A}^*\mathbf{v}_1 = \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix}$$

and

$$\mathbf{u}_2 = \frac{1}{2}\mathbf{A}^*\mathbf{v}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}.$$

We need a third vector orthonormal to both $\mathbf{u}_1$ and $\mathbf{u}_2$, so we take

$$\mathbf{u}_3 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}.$$

(We can obtain this by applying Gram–Schmidt to $(\mathbf{u}_1, \mathbf{u}_2, \mathbf{x})$, where $\mathbf{x} \in \mathbb{R}^3$ is any vector not in the span of $\mathbf{u}_1$ and $\mathbf{u}_2$, or just by inspection.)

This yields the SVD

$$\mathbf{A}^* = \frac{1}{\sqrt{2}}\begin{bmatrix} 0 & 1 & 1 \\ 0 & -1 & 1 \\ -\sqrt{2} & 0 & 0 \end{bmatrix}\begin{bmatrix} 2\sqrt{2} & 0 \\ 0 & 2 \\ 0 & 0 \end{bmatrix}\frac{1}{\sqrt{2}}\begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix}$$

for $\mathbf{A}^*$, and so

$$\mathbf{A} = \frac{1}{\sqrt{2}}\begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 2\sqrt{2} & 0 & 0 \\ 0 & 2 & 0 \end{bmatrix}\frac{1}{\sqrt{2}}\begin{bmatrix} 0 & 0 & -\sqrt{2} \\ 1 & -1 & 0 \\ 1 & 1 & 0 \end{bmatrix}.$$

(d) Here $\mathbf{A}^*\mathbf{A}$ is the same as for the matrix in part (c). Therefore the singular values and the $\mathbf{V}$ part of the SVD for $\mathbf{A}$ are the same as in that part. Then

$$\mathbf{u}_1 = \frac{1}{\sigma_1}\mathbf{A}\mathbf{v}_1 = \frac{1}{2\sqrt{2}}\mathbf{A}\begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix} = \frac{1}{\sqrt{2}}\begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix},$$

$$\mathbf{u}_2 = \frac{1}{\sigma_2}\mathbf{A}\mathbf{v}_2 = \frac{1}{2}\mathbf{A}\frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix},$$

and we add the orthonormal vector $\mathbf{u}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$.

So finally we have the SVD $\mathbf{U} = \frac{1}{\sqrt{2}}\begin{bmatrix} -1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & \sqrt{2} \end{bmatrix}$, $\mathbf{\Sigma} = \mathbf{diag}(2\sqrt{2}, 2, 0)$, $\mathbf{V} = \frac{1}{\sqrt{2}}\begin{bmatrix} 0 & 1 & 1 \\ 0 & -1 & 1 \\ -\sqrt{2} & 0 & 0 \end{bmatrix}$.

(e) $\mathbf{A}^*\mathbf{A} = \mathbf{diag}(1, 4, 1)$ has eigenvalues 4, 1, and 1,

$$\mathrm{Eig}_4(\mathbf{A}^*\mathbf{A}) = \langle \mathbf{e}_2 \rangle,$$

and

$$\mathrm{Eig}_1(\mathbf{A}^*\mathbf{A}) = \langle \mathbf{e}_1, \mathbf{e}_3 \rangle.$$

Therefore we can take $\mathbf{v}_1 = \mathbf{e}_2$, $\mathbf{v}_2 = \mathbf{e}_3$, and $\mathbf{v}_3 = \mathbf{e}_1$, so

$$\mathbf{u}_1 = \frac{1}{2}\mathbf{A}\mathbf{e}_2 = \mathbf{e}_2, \quad \mathbf{u}_2 = \mathbf{A}\mathbf{e}_3 = -\mathbf{e}_1, \quad \text{and} \quad \mathbf{u}_3 = \mathbf{A}\mathbf{e}_1 = \mathbf{e}_3.$$

So finally we have the SVD $\mathbf{U} = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, $\mathbf{\Sigma} = \mathbf{diag}(2, 1, 1)$, $\mathbf{V} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$.

5.3.2. For all these matrices the singular values were found in the corresponding parts of Exercise 5.2.2.

(a) $\mathbf{A}^*\mathbf{A} = 5\mathbf{I}_2$ has eigenvalue 5, with eigenspace $\mathbb{R}^2$, so we may take $\mathbf{v}_1 = \mathbf{e}_1$ and $\mathbf{v}_2 = \mathbf{e}_2$. Then

$$\mathbf{u}_1 = \frac{1}{\sqrt{5}}\mathbf{A}\mathbf{v}_1 = \frac{1}{\sqrt{5}}\begin{bmatrix} 1 \\ -2 \end{bmatrix}$$

and

$$\mathbf{u}_2 = \frac{1}{\sqrt{5}}\mathbf{A}\mathbf{v}_2 = \frac{1}{\sqrt{5}}\begin{bmatrix} 2 \\ 1 \end{bmatrix}.$$

So finally we have the SVD $\mathbf{U} = \frac{1}{\sqrt{5}}\mathbf{A}$, $\mathbf{\Sigma} = \sqrt{5}\mathbf{I}_2$, and $\mathbf{V} = \mathbf{I}_2$.

(b) Here $\mathbf{A}\mathbf{A}^*$ turns out to be simpler than $\mathbf{A}^*\mathbf{A}$, so we apply Algorithm 5.17 to $\mathbf{A}^*$. $\mathbf{A}\mathbf{A}^* = \mathbf{diag}(25, 100)$ has eigenvalues 25 and 100 with corresponding eigenvalues $\mathbf{e}_1$ and $\mathbf{e}_2$, so we can take $\mathbf{v}_1 = \mathbf{e}_2$ and $\mathbf{v}_2 = \mathbf{e}_1$. Then

$$\mathbf{u}_1 = \frac{1}{10}\mathbf{A}^*\mathbf{v}_1 = \frac{1}{5}\begin{bmatrix} 3 \\ 4 \end{bmatrix}$$

and

$$\mathbf{u}_2 = \frac{1}{5}\mathbf{A}^*\mathbf{v}_2 = \frac{1}{5}\begin{bmatrix} 4 \\ -3 \end{bmatrix}.$$

This yields the SVD

$$\mathbf{A}^* = \frac{1}{5}\begin{bmatrix} 3 & 4 \\ 4 & -3 \end{bmatrix}\begin{bmatrix} 10 & 0 \\ 0 & 5 \end{bmatrix}\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

for $\mathbf{A}^*$, and so

$$\mathbf{A} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 10 & 0 \\ 0 & 5 \end{bmatrix}\frac{1}{5}\begin{bmatrix} 3 & 4 \\ 4 & -3 \end{bmatrix}.$$

(c) $\mathbf{A}^*\mathbf{A} = \mathbf{diag}(1, 2, 0)$ has eigenvalues 2, 1, and 0. We can take $(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3) = (\mathbf{e}_2, \mathbf{e}_1, \mathbf{e}_3)$, and compute

$$\mathbf{u}_1 = \frac{1}{\sqrt{2}}\mathbf{A}\mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \qquad \text{and} \qquad \mathbf{u}_2 = \mathbf{A}\mathbf{e}_1 = \begin{bmatrix} i \\ 0 \end{bmatrix},$$

and so we obtain the SVD

$$\mathbf{A} = \begin{bmatrix} 0 & i \\ 1 & 0 \end{bmatrix}\begin{bmatrix} \sqrt{2} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

(d) We apply Algorithm 5.17 to $\mathbf{A}^*$. $\mathbf{A}\mathbf{A}^* = \mathbf{diag}(6, 3)$ has eigenvalues 6 and 3 with corresponding eigenvectors $\mathbf{e}_1$ and $\mathbf{e}_2$. Then

$$\mathbf{u}_1 = \frac{1}{\sqrt{6}}\mathbf{A}^*\mathbf{e}_1 = \frac{1}{\sqrt{6}}\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} \qquad \mathbf{u}_2 = \frac{1}{\sqrt{3}}\mathbf{A}^*\mathbf{e}_2 = \frac{1}{\sqrt{3}}\begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}.$$

We need a $\mathbf{u}_3$ to form an orthonormal basis of $\mathbb{R}^3$. It will have to be a vector $\begin{bmatrix} x \\ y \\ z \end{bmatrix}$ satisfying

$$x + 2y + z = 0 \qquad \text{and} \qquad x - y + z = 0.$$

Solving the system yields $y = 0$ and $x = -z$; since we need a unit vector, we can take $\mathbf{u}_3 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$.

So we have the SVD

$$\mathbf{A}^* = \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} \\ \frac{2}{\sqrt{6}} & -\frac{1}{\sqrt{3}} & 0 \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \sqrt{6} & 0 \\ 0 & \sqrt{3} \\ 0 & 0 \end{bmatrix} \mathbf{I}_2,$$

for $\mathbf{A}^*$, and therefore

$$\mathbf{A} = \mathbf{I}_2 \begin{bmatrix} \sqrt{6} & 0 & 0 \\ 0 & \sqrt{3} & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \end{bmatrix}.$$

(e) $\mathbf{A}^*\mathbf{A} = \mathbf{diag}(9, 1, 4)$ has eigenvalues 9, 4, and 1 with corresponding eigenvectors $\mathbf{v}_1 = \mathbf{e}_1$, $\mathbf{v}_2 = \mathbf{e}_3$, and $\mathbf{v}_3 = \mathbf{e}_2$. Then

$$\mathbf{u}_1 = \frac{1}{3}\mathbf{A}\mathbf{e}_1 = \mathbf{e}_3, \quad \mathbf{u}_2 = \frac{1}{2}\mathbf{A}\mathbf{e}_3 = \mathbf{e}_2, \quad \text{and} \quad \mathbf{u}_3 = \mathbf{A}\mathbf{e}_2 = \mathbf{e}_1.$$

So finally we have the SVD $\mathbf{U} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$, $\mathbf{\Sigma} = \mathbf{diag}(3, 2, 1)$, $\mathbf{V} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$.

5.3.3. We start by finding an SVD for $\mathbf{A}$: $\mathbf{A}^*\mathbf{A} = 5 \begin{bmatrix} 1 & -2 \\ -2 & 4 \end{bmatrix}$. Clearly $C(\mathbf{A}^*\mathbf{A})$ is spanned by $\mathbf{v} = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$, so if $\mathbf{A}^*\mathbf{A}$ has any nonzero eigenvalue, this must be a corresponding eigenvector. We check

$$\mathbf{A}^*\mathbf{A}\mathbf{v} = 25\mathbf{v},$$

so 25 is an eigenvalue, with corresponding unit eigenvector $\mathbf{v}_1 = \frac{1}{\sqrt{5}}\mathbf{v}$. Since $\operatorname{rank}(\mathbf{A}^*\mathbf{A}) = 1 < 2$, $\mathbf{A}^*\mathbf{A}$ is singular, hence has 0 as an eigenvalue, and we can check that

$$\mathbf{v}_2 = \frac{1}{\sqrt{5}} \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

is a corresponding unit eigenvector. Then

$$\mathbf{u}_1 = \frac{1}{5}\mathbf{A}\mathbf{v}_1 = \frac{1}{\sqrt{5}} \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

and we can take $\mathbf{u}_2$ to be the orthogonal unit vector $\mathbf{u}_2 = \frac{1}{\sqrt{5}} \begin{bmatrix} 2 \\ -1 \end{bmatrix}$. This yields the SVD

$$\mathbf{A} = \frac{1}{\sqrt{5}} \begin{bmatrix} 1 & 2 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} 5 & 0 \\ 0 & 0 \end{bmatrix} \frac{1}{\sqrt{5}} \begin{bmatrix} 1 & -2 \\ 2 & 1 \end{bmatrix}$$

from which we find

$$\mathbf{A}^\dagger = \frac{1}{\sqrt{5}}\begin{bmatrix}1 & 2\\ -2 & 1\end{bmatrix}\begin{bmatrix}1/5 & 0\\ 0 & 0\end{bmatrix}\frac{1}{\sqrt{5}}\begin{bmatrix}1 & 2\\ 2 & -1\end{bmatrix} = \frac{1}{25}\begin{bmatrix}1 & 2\\ -2 & -4\end{bmatrix}.$$

5.3.4. From Exercise 5.3.1(c) we have the SVD

$$\mathbf{A} = \frac{1}{\sqrt{2}}\begin{bmatrix}-1 & 1\\ 1 & 1\end{bmatrix}\begin{bmatrix}2\sqrt{2} & 0 & 0\\ 0 & 2 & 0\end{bmatrix}\frac{1}{\sqrt{2}}\begin{bmatrix}0 & 0 & -\sqrt{2}\\ 1 & -1 & 0\\ 1 & 1 & 0\end{bmatrix}$$

from which we find

$$\mathbf{A}^\dagger = \frac{1}{\sqrt{2}}\begin{bmatrix}0 & 1 & 1\\ 0 & -1 & 1\\ -\sqrt{2} & 0 & 0\end{bmatrix}\begin{bmatrix}\frac{1}{2\sqrt{2}} & 0\\ 0 & \frac{1}{2}\\ 0 & 0\end{bmatrix}\frac{1}{\sqrt{2}}\begin{bmatrix}-1 & 1\\ 1 & 1\end{bmatrix} = \frac{1}{4}\begin{bmatrix}1 & 1\\ -1 & -1\\ 1 & -1\end{bmatrix}.$$

5.3.5. Since $\boldsymbol{R}^2 = \boldsymbol{I}$, $\langle \mathbf{v}, \boldsymbol{R}\mathbf{w}\rangle = \left\langle \boldsymbol{R}^2\mathbf{v}, \boldsymbol{R}\mathbf{w}\right\rangle$. Then since $\boldsymbol{R}$ is an isometry, $\left\langle \boldsymbol{R}^2\mathbf{v}, \boldsymbol{R}\mathbf{w}\right\rangle = \langle \boldsymbol{R}\mathbf{v}, \mathbf{w}\rangle$ by Theorem 4.26. Therefore $\boldsymbol{R}$ possesses the defining property of $\boldsymbol{R}^*$, so $\boldsymbol{R} = \boldsymbol{R}^*$.

Alternatively, write $\boldsymbol{R} = 2\boldsymbol{P}_L - \boldsymbol{I}$ (see Exercise 4.5.1). Since $\boldsymbol{P}_L$ is self-adjoint, it follows that

$$\boldsymbol{R}^* = (2\boldsymbol{P}_L - \boldsymbol{I})^* = 2\boldsymbol{P}_L^* - \boldsymbol{I}^* = 2\boldsymbol{P}_L - \boldsymbol{I} = \boldsymbol{R}.$$

5.3.6. For $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$,

$$\begin{aligned}\langle \boldsymbol{T}\mathbf{x}, \mathbf{y}\rangle_{\mathbf{A}} &= \langle \mathbf{B}\mathbf{x}, \mathbf{y}\rangle_{\mathbf{A}} = \langle \mathbf{A}\mathbf{B}\mathbf{x}, \mathbf{A}\mathbf{y}\rangle = \langle \mathbf{x}, \mathbf{B}^*\mathbf{A}^*\mathbf{A}\mathbf{y}\rangle = \left\langle \mathbf{A}^{-1}\mathbf{A}\mathbf{x}, \mathbf{B}^*\mathbf{A}^*\mathbf{A}\mathbf{y}\right\rangle\\ &= \left\langle \mathbf{A}\mathbf{x}, (\mathbf{A}^{-1})^*\mathbf{B}^*\mathbf{A}^*\mathbf{A}\mathbf{y}\right\rangle = \left\langle \mathbf{A}\mathbf{x}, \mathbf{A}\mathbf{A}^{-1}(\mathbf{A}^{-1})^*\mathbf{B}^*\mathbf{A}^*\mathbf{A}\mathbf{y}\right\rangle\\ &= \left\langle \mathbf{x}, \mathbf{A}^{-1}(\mathbf{A}^{-1})^*\mathbf{B}^*\mathbf{A}^*\mathbf{A}\mathbf{y}\right\rangle_{\mathbf{A}}.\end{aligned}$$

This means that $\mathbf{A}^{-1}(\mathbf{A}^{-1})^*\mathbf{B}^*\mathbf{A}^*\mathbf{A} = \mathbf{A}^{-1}(\mathbf{A}\mathbf{B}\mathbf{A}^{-1})^*\mathbf{A}$ acts as an adjoint to $\boldsymbol{T}$ with respect to $\langle\cdot,\cdot\rangle_{\mathbf{A}}$.

5.3.7. (a) Integrating by parts,

$$\begin{aligned}\langle \boldsymbol{D}f, g\rangle &= \int_0^{2\pi} f'(x)g(x)\,dx\\ &= f(x)g(x)]_0^{2\pi} - \int_0^{2\pi} f(x)g'(x)\,dx\\ &= \int_0^{2\pi} f(x)(-g'(x))\,dx\\ &= \langle f, -\boldsymbol{D}g\rangle,\end{aligned}$$

where the third equality follows because f and g are 2π-periodic. Therefore $\boldsymbol{D}^* = -\boldsymbol{D}$.

(b) Suppose that $\lambda \in \mathbb{R}$ is an eigenvalue of $\boldsymbol{D}$ with corresponding eigenvector $f \in C^\infty_{2\pi}(\mathbb{R})$. Then

$$\langle \boldsymbol{D}f, f\rangle = \langle \lambda f, f\rangle = \lambda \|f\|^2,$$

and also

$$\langle \boldsymbol{D}f, f\rangle = \langle f, \boldsymbol{D}^* f\rangle = \langle f, -\boldsymbol{D}f\rangle = \langle f, -\lambda f\rangle = -\lambda \|f\|^2.$$

Therefore $\lambda = -\lambda$, so $\lambda = 0$.

(c) This can be proved by integrating by parts as in part (a) twice, or from part (a):

$$(\boldsymbol{D}^2)^* = (\boldsymbol{D}^*)^2 = (-\boldsymbol{D})^2 = \boldsymbol{D}^2.$$

5.3.8. (a) By integration by parts,

$$\langle Dp, q\rangle = \int_{-\infty}^{\infty} p'(x)q(x)e^{-x^2/2}\,dx = -\int_{-\infty}^{\infty} p(x)\left[q'(x) - xq(x)\right]e^{-x^2/2}\,dx$$

(note that the boundary terms vanish because $e^{-x^2/2}$ goes to 0 as $x \to \pm\infty$ faster than any polynomial grows). This means that

$$[\boldsymbol{D}^* q](x) = xq(x) - q'(x),$$

with respect to the given inner product.

(b) The jth column of the matrix of $\boldsymbol{D}$ is the coordinate representation of $\boldsymbol{D}[x^{j-1}] = (j-1)x^{j-2}$ (when $j \geq 2$) and the first column is all zeroes. That is,

$$[\boldsymbol{D}]_{(1,\dots,x^n),(1,\dots,x^{n-1})} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & n-1 \end{bmatrix}.$$

(c) The jth column of $\boldsymbol{D}^*$ is the coordinate representation of $x^j - (j-1)x^{j-2}$ for $j \geq 2$, and the first column is the coordinate representation of x, so

$$[\boldsymbol{D}^*]_{(1,\dots,x^{n-1}),(1,\dots,x^n)} = \begin{bmatrix} 0 & -1 & 0 & \cdots & 0 \\ 1 & 0 & -2 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}.$$

(d) The given bases are not orthonormal with respect to the inner product we're working with.

5.3.9. If $u \in U$ and $v \in V$,

$$\langle \boldsymbol{J}u, v\rangle = \langle u, v\rangle = \langle u, \boldsymbol{P}_U v + \boldsymbol{P}_{U^\perp} v\rangle = \langle u, \boldsymbol{P}_U v\rangle + \langle u, \boldsymbol{P}_{U^\perp} v\rangle = \langle u, \boldsymbol{P}_U v\rangle,$$

where the last inequality follows since $\boldsymbol{P}_{U^\perp} v \in U^\perp$ and $u \in U$. Therefore $\boldsymbol{J}^* = \boldsymbol{P}_U$.

5.3.10. (a) Fix any orthonormal basis $\mathcal{B}$ of V, and let $\mathbf{A} = [\boldsymbol{T}]_{\mathcal{B}}$. Then $\mathbf{A}^* = [\boldsymbol{T}^*]_{\mathcal{B}}$, and so

$$\operatorname{tr} \boldsymbol{T}^* = \operatorname{tr} \mathbf{A}^* = \overline{(\operatorname{tr} \mathbf{A})} = \overline{(\operatorname{tr} \boldsymbol{T})}.$$

(b) If $\boldsymbol{T}$ is self-adjoint, then by part (a),

$$\operatorname{tr} \boldsymbol{T} = \operatorname{tr} \boldsymbol{T}^* = \overline{(\operatorname{tr} \boldsymbol{T})},$$

which implies that $\operatorname{tr} \boldsymbol{T} \in \mathbb{R}$.

5.3.11. $(\operatorname{Re} \mathbf{A})^* = \frac{1}{2}(\mathbf{A} + \mathbf{A}^*)^* = \frac{1}{2}(\mathbf{A}^* + \mathbf{A}) = \operatorname{Re} \mathbf{A}$ and $(\operatorname{Im} \mathbf{A})^* = -\frac{1}{2i}(\mathbf{A} - \mathbf{A}^*)^* = -\frac{1}{2i}(\mathbf{A}^* - \mathbf{A}) = \operatorname{Im} \mathbf{A}$.

5.3.12. If $\boldsymbol{T}$ is self-adjoint, then by Proposition 5.16, $\ker \boldsymbol{T} = \ker \boldsymbol{T}^* = (\operatorname{range} \boldsymbol{T})^\perp$. Therefore every $v \in V$ can be uniquely written as $v = \boldsymbol{T}v' + w$, where $\boldsymbol{T}w = 0$, and $\langle \boldsymbol{T}v', w\rangle = 0$. Then

$$\boldsymbol{T}v = \boldsymbol{T}(\boldsymbol{T}v' + w) = \boldsymbol{T}^2 v' + \boldsymbol{T}w = \boldsymbol{T}v',$$

using the facts that $\boldsymbol{T}^2 = \boldsymbol{T}$ and that $\boldsymbol{T}w = 0$ in the last step. But by the representation of v above, $\boldsymbol{T}v'$ is exactly the orthogonal projection of v onto the range of $\boldsymbol{T}$; i.e., $\boldsymbol{T}v = \boldsymbol{P}_{\operatorname{range} \boldsymbol{T}} v$ for all $v \in V$.

5.3.13. If $\boldsymbol{T}^*\boldsymbol{T} = \mathbf{0}$, then for each $v \in V$, $0 = \langle \boldsymbol{T}^*\boldsymbol{T}v, v\rangle = \|\boldsymbol{T}v\|^2$, and therefore $\boldsymbol{T} = \mathbf{0}$. The other direction is trivial.

5.3.14. **Solution 1:** Fix orthonormal bases of V and W and let $\mathbf{A}$ be the matrix of $\boldsymbol{T}$ with respect to those bases. By Exercises 5.2.4 and 5.2.5, $\boldsymbol{T}$ has the same singular values as $\mathbf{A}$, which are the singular values of $\mathbf{A}^*$, which are the singular values of $\boldsymbol{T}^*$. In particular, the operator norms coincide.

Solution 2: For any $w \in W$,

$$\begin{aligned} \|\boldsymbol{T}w\|^2 = \langle \boldsymbol{T}w, \boldsymbol{T}w\rangle = \langle w, \boldsymbol{T}^*(\boldsymbol{T}w)\rangle &\le \|w\| \, \|\boldsymbol{T}^*(\boldsymbol{T}w)\| \le \|w\| \, \|\boldsymbol{T}^*\|_{op} \, \|\boldsymbol{T}w\| \\ &\le \|w\| \, \|\boldsymbol{T}^*\|_{op} \, \|\boldsymbol{T}\|_{op} \, \|w\| , \end{aligned}$$

where the first inequality follows from the Cauchy–Schwarz inequality, and the others follow from the definition of the operator norm. Therefore, if $\|w\| = 1$, then $\|\boldsymbol{T}w\| \le \|\boldsymbol{T}^*\|_{op}$. This implies that $\|\boldsymbol{T}\|_{op} \le \|\boldsymbol{T}^*\|_{op}$.

Applying this to $\boldsymbol{T}^*$, it follows that also

$$\|\boldsymbol{T}^*\|_{op} \le \|(\boldsymbol{T}^*)^*\|_{op} = \|\boldsymbol{T}\|_{op} .$$

5.3.15. For any $v \in V$,

$$\|\boldsymbol{T}^*\boldsymbol{T}v\| \le \|\boldsymbol{T}^*\|_{op} \, \|\boldsymbol{T}v\| \le \|opT^*\|_{op} \, \|\boldsymbol{T}\|_{op} \, \|v\| = \|\boldsymbol{T}\|_{op}^2 \, \|v\|$$

by Exercise 5.3.14, and so $\|\boldsymbol{T}^*\boldsymbol{T}\| \le \|\boldsymbol{T}\|_{op}^2$.

Now suppose that $(e_1, \dots, e_n)$ and $(f_1, \dots, f_m)$ are left and right singular vectors of $\boldsymbol{T}$, and $\sigma_1 \geq \cdots \geq \sigma_p$ are the singular values. Then for each j,

$$\langle \boldsymbol{T}^* f_1, e_j \rangle = \langle f_1, \boldsymbol{T} e_j \rangle = \sigma_j \langle f_1, f_j \rangle = \begin{cases} \sigma_1 & \text{if } j = 1, \\ 0 & \text{otherwise,} \end{cases}$$

since the f_j are orthonormal. Therefore $\langle \boldsymbol{T}^* f_1, e_j \rangle = \langle \sigma_1 e_1, e_j \rangle$. Since the e_j form a basis of V, this implies that $\boldsymbol{T}^* f_1 = \sigma_1 e_1$. Therefore

$$\|\boldsymbol{T}^* \boldsymbol{T} e_1\| = \sigma_1 \|\boldsymbol{T}^* f_1\| = \sigma_1^2 \|e_1\| = \sigma_1^2 = \|\boldsymbol{T}\|_{op}^2,$$

which implies that $\|\boldsymbol{T}^* \boldsymbol{T}\|_{op} \geq \|\boldsymbol{T}\|_{op}^2$. Thus $\|\boldsymbol{T}^* \boldsymbol{T}\|_{op} = \|\boldsymbol{T}\|_{op}^2$

The other equality in the problem follows by applying this to $\boldsymbol{T}^*$ and using Exercise 5.3.14.

5.3.16. Let $V = \{\boldsymbol{T} \in \mathcal{L}(V) \mid \boldsymbol{T}^* = \boldsymbol{T}\}$. Then:

- $\boldsymbol{0}^* = \boldsymbol{0}$, so $\boldsymbol{0} \in V$.
- If $\boldsymbol{S}, \boldsymbol{T} \in V$, then $(\boldsymbol{S} + \boldsymbol{T})^* = \boldsymbol{S}^* + \boldsymbol{T}^* = \boldsymbol{S} + \boldsymbol{T}$, so $\boldsymbol{S} + \boldsymbol{T} \in V$.
- If $\boldsymbol{T} \in V$ and $a \in \mathbb{R}$, then $(a\boldsymbol{T})^* = a\boldsymbol{T}^* = a\boldsymbol{T}$, so $a\boldsymbol{T} \in V$.

5.3.17. Fix orthonormal bases on V and W. These bases define an isomorphism $\boldsymbol{C} : \mathcal{L}(V, W) \to \mathrm{M}_{m,n}(\mathbb{F})$ such that

$$\langle \boldsymbol{S}, \boldsymbol{T} \rangle_F = \operatorname{tr}(\boldsymbol{S}\boldsymbol{T}^*) = \operatorname{tr} \boldsymbol{C}(S)\boldsymbol{C}(\boldsymbol{T})^* = \langle \boldsymbol{C}(S), \boldsymbol{C}(\boldsymbol{T}) \rangle_F,$$

where the second equality uses Theorems 3.47 and 5.13. Since we know that the Frobenius inner product is indeed an inner product on $\mathrm{M}_{m,n}(\mathbb{F})$, it follows from Exercise 4.1.9 that the expression defined here gives an inner product on $\mathcal{L}(V, W)$.

5.3.18. Suppose that v is an eigenvector of $\boldsymbol{T}$ with eigenvalue λ. Then

$$\lambda \|v\|^2 = \langle \lambda v, v \rangle = \langle \boldsymbol{T} v, v \rangle = \langle v, \boldsymbol{T}^* v \rangle = -\langle v, \boldsymbol{T} v \rangle = -\langle v, \lambda v \rangle = -\overline{\lambda} \|v\|^2.$$

Since $v \neq 0$, we can divide both sides by $\|v\|^2$ to get $\lambda = -\overline{\lambda}$, which is equivalent to the statement $\lambda = ia$ for some real number a (i.e., λ is purely imaginary).

5.3.19. Let v_1 and v_2 be eigenvectors corresponding to distinct eigenvalues λ_1 and λ_2. Then

$$\lambda_1 \langle v_1, v_2 \rangle = \langle \lambda_1 v_1, v_2 \rangle = \langle \boldsymbol{T} v_1, v_2 \rangle = -\langle v_1, \boldsymbol{T} v_2 \rangle = -\langle v_1, \lambda_2 v_2 \rangle = -\overline{\lambda}_2 \langle v_1, v_2 \rangle.$$

This implies that either $\langle v_1, v_2 \rangle = 0$ or $\lambda_1 = -\overline{\lambda}_2$. But by Exercise 5.3.18, λ_1 and λ_2 are purely imaginary, so $-\overline{\lambda}_2 = \lambda_2 \neq \lambda_1$. Therefore $\langle v_1, v_2 \rangle = 0$.

5.3.20. (a) Since $(f_1, \dots, f_r)$ is orthonormal, we just need to check that $\operatorname{range} \boldsymbol{T} = \langle f_1, \dots, f_r \rangle$. Since $f_j = \boldsymbol{T}(\sigma_j^{-1} e_j)$ for $1 \leq j \leq r$, $\langle f_1, \dots, f_r \rangle \subseteq \operatorname{range} \boldsymbol{T}$. If $v \in V$, then $v = \sum_{j=1}^n \langle v, e_j \rangle e_j$, and so

$$\boldsymbol{T} v = \sum_{j=1}^n \langle v, e_j \rangle \boldsymbol{T} e_j = \sum_{j=1}^r \langle v, e_j \rangle \sigma_j f_j \in \langle f_1, \dots, f_r \rangle,$$

so $\operatorname{range} \boldsymbol{T} \subseteq \langle f_1, \dots, f_r \rangle$.

(b) Since $(e_{r+1}, \ldots, e_n)$ is orthonormal, we just need to check that $\ker \boldsymbol{T} = \langle e_{r+1}, \ldots, e_n \rangle$. Since $\boldsymbol{T} e_j = 0$ for $j > r$, $\langle e_{r+1}, \ldots, e_n \rangle \subseteq \ker \boldsymbol{T}$. If $v \in \ker \boldsymbol{T}$, then as above $v = \sum_{j=1}^{n} \langle v, e_j \rangle e_j$ and

$$0 = \boldsymbol{T} v = \sum_{j=1}^{r} \langle v, e_j \rangle \sigma_j f_j.$$

Since $(f_1, \ldots, f_r)$ is linearly independent, this implies that $\langle v, e_j \rangle \sigma_j = 0$ for each $j = 1, \ldots, r$. Since $\sigma_j > 0$ for $1 \le j \le r$, this implies $\langle v, e_j \rangle = 0$, and so $v = \sum_{j=r+1}^{n} \langle v, e_j \rangle e_j$. Therefore $\ker \boldsymbol{T} \subseteq \langle e_{r+1}, \ldots, e_n \rangle$.

(c) This follows from part (a) and Proposition 5.16.

(d) This follows from part (b) and Proposition 5.16.

5.3.21. Given any $u \in U$ and $w \in W$,

$$\langle u, \boldsymbol{T}^* \boldsymbol{R}^* w \rangle = \langle \boldsymbol{T} u, \boldsymbol{R}^* w \rangle = \langle \boldsymbol{R} \boldsymbol{T} u, w \rangle .$$

This implies that $(\boldsymbol{R}\boldsymbol{T})^* = \boldsymbol{T}^* \boldsymbol{R}^*$.

If $\boldsymbol{T}$ is invertible, then $\operatorname{range} \boldsymbol{T} = V$, so by), then $\ker \boldsymbol{T}^* = (\operatorname{range} \boldsymbol{T})^\perp = \{0\}$, and so $\boldsymbol{T}^*$ is invertible as well. Given $u \in U$ and $v \in V$, write $v' = (\boldsymbol{T}^*)^{-1} u$ and $u' = \boldsymbol{T}^{-1} v$. Then

$$\left\langle \boldsymbol{T}^{-1} v, u \right\rangle = \left\langle u', \boldsymbol{T}^* v' \right\rangle = \left\langle \boldsymbol{T} u', v' \right\rangle = \left\langle v, (\boldsymbol{T}^*)^{-1} u \right\rangle .$$

Therefore $(\boldsymbol{T}^{-1})^* = (\boldsymbol{T}^*)^{-1}$.

5.3.22.

- Find the eigenvalues of $\mathbf{A}\mathbf{A}^*$. The square roots of the largest $p = \min\{m, n\}$ of them are the singular values $\sigma_1 \ge \cdots \ge \sigma_p$ of $\mathbf{A}$.
- Find an orthonormal basis for each eigenspace of $\mathbf{A}\mathbf{A}^*$; put the resulting collection of vectors $(\mathbf{u}_1, \ldots, \mathbf{u}_m)$ in a list so that the corresponding eigenvalues are in decreasing order. Take

$$\mathbf{U} := \begin{bmatrix} | & & | \\ \mathbf{u}_1 & \cdots & \mathbf{u}_m \\ | & & | \end{bmatrix}.$$

- Let $r = \operatorname{rank} \mathbf{A}$. For $1 \le j \le r$, define $\mathbf{v}_j := \frac{1}{\sigma_j} \mathbf{A}^* \mathbf{v}_j$, and extend $(\mathbf{v}_1, \ldots, \mathbf{v}_r)$ to an orthonormal basis $(\mathbf{v}_1, \ldots, \mathbf{v}_m)$ of $\mathbb{F}^n$. Take

$$\mathbf{V} := \begin{bmatrix} | & & | \\ \mathbf{v}_1 & \cdots & \mathbf{v}_n \\ | & & | \end{bmatrix}.$$

Then if $\Sigma \in \mathrm{M}_{m,n}(\mathbb{F})$ has σ_j in the (j, j) position for $1 \le j \le p$ and 0 otherwise,

$$\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^*$$

is an SVD of $\mathbf{A}$.

This can be proved directly, but it is easier to note that Algorithm 5.17 applied to $\mathbf{A}^*$ shows that $\mathbf{A}^* = \mathbf{V}\boldsymbol{\Sigma}^*\mathbf{U}^*$ is an SVD of $\mathbf{A}^*$; it follows that $\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^*$.

5.4 The Spectral Theorems

5.4.1. (a) First find the eigenvalues:

$$\begin{bmatrix} 1-\lambda & 2 \\ 2 & 4-\lambda \end{bmatrix} \rightsquigarrow \begin{bmatrix} 2 & 4-\lambda \\ 0 & 2-\frac{(1-\lambda)(4-\lambda)}{2} \end{bmatrix},$$

which is singular if and only if $(1-\lambda)(4-\lambda) = 4$, i.e., $\lambda \in \{5, 0\}$. A vector $\begin{bmatrix} x \\ y \end{bmatrix}$ is an eigenvector with eigenvalue 5 if and only if it is in the null space of $\begin{bmatrix} -4 & 2 \\ 2 & -1 \end{bmatrix}$; we may thus take $\mathbf{u}_1 = \frac{1}{\sqrt{5}}\begin{bmatrix} 1 \\ 2 \end{bmatrix}$ as a unit eigenvector. Similarly, $\begin{bmatrix} x \\ y \end{bmatrix}$ is an eigenvector with eigenvalue 0 if and only if it is in the null space of $\mathbf{A}$, and so we may choose $\mathbf{u}_2 = \frac{1}{\sqrt{5}}\begin{bmatrix} 2 \\ -1 \end{bmatrix}$. Then

$$\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} = \left(\frac{1}{\sqrt{5}}\begin{bmatrix} 1 & 2 \\ 2 & -1 \end{bmatrix}\right)\begin{bmatrix} 5 & 0 \\ 0 & 0 \end{bmatrix}\left(\frac{1}{\sqrt{5}}\begin{bmatrix} 1 & 2 \\ 2 & -1 \end{bmatrix}\right)$$

is a spectral decomposition.

(b) Following the same procedure as in part (a),

$$\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ -i & i \end{bmatrix}\begin{bmatrix} 1+i & 0 \\ 0 & 1-i \end{bmatrix}\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & i \\ 1 & -i \end{bmatrix}.$$

(c) Similarly,

$$\begin{bmatrix} 2 & 0 & 1 \\ 0 & -1 & 0 \\ 1 & 0 & 2 \end{bmatrix} = \begin{bmatrix} 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 1 & 0 & 0 \\ 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}\begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{bmatrix}\begin{bmatrix} 0 & 1 & 0 \\ \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \end{bmatrix}.$$

(d) Similarly,

$$\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & -\frac{2}{\sqrt{6}} \end{bmatrix}\begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}\begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} \end{bmatrix}.$$

5.4.2. (a) First find the eigenvalues:

$$\begin{bmatrix} 7-\lambda & 6 \\ 6 & -2-\lambda \end{bmatrix} \rightsquigarrow \begin{bmatrix} 6 & -2-\lambda \\ 0 & 6-\frac{(2+\lambda)(\lambda-7)}{6} \end{bmatrix},$$

which is singular if and only if $(2+\lambda)(\lambda-7) = 36$, i.e., $\lambda \in \{10, -5\}$. A vector $\begin{bmatrix} x \\ y \end{bmatrix}$ is an eigenvector with eigenvalue 10 if and only if it is in the null space of $\begin{bmatrix} 6 & -12 \\ 0 & 0 \end{bmatrix}$; we may thus take $\mathbf{u}_1 = \frac{1}{\sqrt{5}} \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ as a unit eigenvector. Similarly, $\begin{bmatrix} x \\ y \end{bmatrix}$ is an eigenvector with eigenvalue -5 if and only if it is in the null space of $\begin{bmatrix} 6 & 3 \\ 0 & 0 \end{bmatrix}$, and so we may choose $\mathbf{u}_2 = \frac{1}{\sqrt{5}} \begin{bmatrix} 1 \\ -2 \end{bmatrix}$. Then

$$\begin{bmatrix} 7 & 6 \\ 6 & -2 \end{bmatrix} = \left(\frac{1}{\sqrt{5}} \begin{bmatrix} 2 & 1 \\ 1 & -2 \end{bmatrix} \right) \begin{bmatrix} 10 & 0 \\ 0 & -5 \end{bmatrix} \left(\frac{1}{\sqrt{5}} \begin{bmatrix} 2 & 1 \\ 1 & -2 \end{bmatrix} \right)$$

is a spectral decomposition.

(b) Following the same procedure as in part (a),

$$\begin{bmatrix} 1 & i \\ i & 1 \end{bmatrix} = \left(\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \right) \begin{bmatrix} 1+i & 0 \\ 0 & 1-i \end{bmatrix} \left(\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \right).$$

(c) Similarly,

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

(d) Similarly,

$$\begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & 1 \\ -1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} = \left(\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ -1 & 0 & 1 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix} \right) \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \left(\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix} \right)$$

5.4.3. (a) Since $\mathbf{A}$ is lower triangular, the eigenvalues are equal to the diagonal entries 3 and 1. We find that $\mathbf{u}_1 = \mathbf{e}_2$ is a unit eigenvector corresponding to the eigenvalue 3, and choose $\mathbf{u}_2 = \mathbf{e}_1$ to complete an orthonormal basis. Then we compute

$$\mathbf{T} = \mathbf{U}^*\mathbf{A}\mathbf{U} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 3 & 2 \\ 0 & 1 \end{bmatrix}.$$

(b) We first find that 2 and -1 are the eigenvalues of $\mathbf{A}$. We find that $\mathbf{u}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ is an eigenvector corresponding to the eigenvalue 2, and choose $\mathbf{u}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ to complete an orthonormal basis. Then we compute

$$\mathbf{T} = \mathbf{U}^*\mathbf{A}\mathbf{U} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ -2 & 0 \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 0 & -1 \end{bmatrix}.$$

5.4.4. (a) Since $\mathbf{A}$ is lower triangular, the eigenvalues are equal to the diagonal entries 3 and 4. We find that $\mathbf{u}_1 = \mathbf{e}_2$ is a unit eigenvector corresponding to the eigenvalue 4, and choose $\mathbf{u}_2 = \mathbf{e}_1$ to complete an orthonormal basis. Then we compute

$$\mathbf{T} = \mathbf{U}^*\mathbf{A}\mathbf{U} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 4 & 1 \\ 0 & 3 \end{bmatrix}.$$

(b) We first find that 2 is the only eigenvalue of $\mathbf{A}$. We find that $\mathbf{u}_1 = \frac{1}{\sqrt{2}}\begin{bmatrix}1\\1\end{bmatrix}$ is an eigenvector corresponding to the eigenvalue 2, and choose $\mathbf{u}_2 = \frac{1}{\sqrt{2}}\begin{bmatrix}1\\-1\end{bmatrix}$ to complete an orthonormal basis. Then we compute

$$\mathbf{T} = \mathbf{U}^*\mathbf{A}\mathbf{U} = \frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\\1 & -1\end{bmatrix}\begin{bmatrix}3 & -1\\1 & 1\end{bmatrix}\frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\\1 & -1\end{bmatrix} = \begin{bmatrix}2 & 2\\0 & 2\end{bmatrix}.$$

5.4.5. By the Spectral Theorem, there exist an orthogonal matrix $\mathbf{U} \in \mathrm{M}_n(\mathbb{R})$ and a diagonal matrix $\Lambda = \mathbf{diag}(\lambda_1, \ldots, \lambda_n) \in \mathrm{M}_n(\mathbb{R})$ such that $\mathbf{A} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^*$. Let

$$\mathbf{B} = \mathbf{U}\,\mathbf{diag}(\lambda_1^{1/3}, \ldots, \lambda_n^{1/3})\mathbf{U}^*.$$

Then $\mathbf{B} \in \mathrm{M}_n(\mathbb{R})$ and

$$\mathbf{B} = \mathbf{U}\,\mathbf{diag}(\lambda_1^{1/3}, \ldots, \lambda_n^{1/3})^3\mathbf{U}^* = \mathbf{A}.$$

5.4.6. Let $\mathbf{A}$ be invertible and Hermitian. Then by the spectral theorem, there is a unitary matrix $\mathbf{U}$ and a diagonal matrix $\boldsymbol{\Lambda}$ with real entries given by the eigenvalues of $\mathbf{A}$, such that $\mathbf{A} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^*$. Since $\mathbf{A}$ is invertible, none of the diagonal entries $\lambda_1, \ldots \lambda_n$ of Λ are zero. Defining the matrix $\frac{1}{\mathbf{A}}$ via the functional calculus means

$$\frac{1}{\mathbf{A}} = \mathbf{U}\,\mathbf{diag}(\lambda_1^{-1}, \ldots, \lambda_n^{-1})\mathbf{U}^*,$$

and so

$$\begin{aligned}\mathbf{A}\left(\frac{1}{\mathbf{A}}\right) &= \mathbf{U}\,\mathbf{diag}(\lambda_1, \ldots, \lambda_n)\mathbf{U}^*\mathbf{U}\,\mathbf{diag}(\lambda_1^{-1}, \ldots, \lambda_n^{-1})\mathbf{U}^*\\ &= \mathbf{U}\,\mathbf{diag}(\lambda_1, \ldots, \lambda_n)\,\mathbf{diag}(\lambda_1^{-1}, \ldots, \lambda_n^{-1})\mathbf{U}^* = \mathbf{U}\mathbf{I}_n\mathbf{U}^* = \mathbf{I}_n.\end{aligned}$$

That is, $\frac{1}{\mathbf{A}}$ is a matrix inverse of $\mathbf{A}$.

5.4.7. Suppose that $\mathbf{A}$ and $\mathbf{B}$ commute. Then

$$\begin{aligned}e^{\mathbf{A}+\mathbf{B}} &= \sum_{k=0}^{\infty}\frac{1}{k!}(\mathbf{A}+\mathbf{B})^k = \sum_{k=0}^{\infty}\frac{1}{k!}\sum_{j=0}^{k}\binom{k}{j}\mathbf{A}^k\mathbf{B}^{j-k}\\ &= \sum_{j=0}^{\infty}\sum_{k=j}^{\infty}\frac{1}{j!(k-j)!}\mathbf{A}^j\mathbf{B}^{k-j} = \sum_{j=0}^{\infty}\sum_{k=j}^{\infty}\frac{1}{j!(k-j)!}\mathbf{A}^j\mathbf{B}^{k-j}\\ &= \sum_{j=0}^{\infty}\sum_{\ell=0}^{\infty}\frac{1}{j!\ell!}\mathbf{A}^j\mathbf{B}^{\ell} = \left(\sum_{j=0}^{\infty}\frac{1}{j!}\mathbf{A}^j\right)\left(\sum_{\ell=0}^{\infty}\frac{1}{\ell!}\mathbf{B}^{\ell}\right) = e^{\mathbf{A}}e^{\mathbf{B}}.\end{aligned}$$

For a counterexample with noncommuting matrices, consider

$$\mathbf{A} = \begin{bmatrix}1 & 0\\0 & 0\end{bmatrix} \qquad \text{and} \qquad \mathbf{B} = \begin{bmatrix}0 & 1\\1 & 0\end{bmatrix} = \frac{1}{2}\begin{bmatrix}1 & 1\\1 & -1\end{bmatrix}\begin{bmatrix}1 & 0\\0 & -1\end{bmatrix}\begin{bmatrix}1 & 1\\1 & -1\end{bmatrix}.$$

Then

$$e^{\mathbf{A}} = \begin{bmatrix} e & 1 \\ 1 & 1 \end{bmatrix}$$

and

$$e^{\mathbf{B}} = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} e & 1 \\ 1 & e^{-1} \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} e+e^{-1}+2 & e-e^{-1} \\ e-e^{-1} & e+e^{-1}-2 \end{bmatrix}.$$

From these we can compute that $e^{\mathbf{A}}e^{\mathbf{B}}$ is not Hermitian, and therefore cannot be equal to $e^{\mathbf{A}+\mathbf{B}}$.

5.4.8. We first show that $e^{\mathbf{A}}$ is Hermitian. Let $\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^*$ be a spectral decomposition of $\mathbf{A}$. Then $e^{\mathbf{A}} = \mathbf{U}\,\mathbf{diag}(e^{\lambda_1}, \dots, e^{\lambda_n})\mathbf{U}^*$, and so it follows (since the λ_i are all real) that $(e^{\mathbf{A}})^* = e^{\mathbf{A}}$. By Theorem 5.20 and the definition following it, $e^{\mathbf{A}}$ is positive definite if and only if all of its eigenvalues are strictly positive. But the expression for $e^{\mathbf{A}}$ above shows that the eigenvalues are exactly e^{λ_j} for $j = 1, \dots, n$, all of which are indeed strictly positive.

5.4.9. This is essentially the same as the proof of Theorem 5.20.

5.4.10. (a) For any $x \in \mathbb{C}^n$, $\langle \mathbf{A}^*\mathbf{A}\mathbf{x}, \mathbf{x}\rangle = \langle \mathbf{A}\mathbf{x}, \mathbf{A}\mathbf{x}\rangle = \|\mathbf{A}\mathbf{x}\|^2 \geq 0$, so $\mathbf{A}^*\mathbf{A}$ is positive semidefinite by the second characterization in Exercise 5.4.9.

(b) By Exercise 5.4.9, every eigenvalue of $\mathbf{A}^*\mathbf{A}$ is nonnegative. By Exercise 5.2.7, $\operatorname{rank} \mathbf{A}^*\mathbf{A} = \operatorname{rank} \mathbf{A} = n$, so by the Rank–Nullity Theorem, $\ker \mathbf{A}^*\mathbf{A} = \{\mathbf{0}\}$, and therefore 0 is not an eigenvalue of $\mathbf{A}^*\mathbf{A}$. By Theorem 5.20, $\mathbf{A}^*\mathbf{A}$ is therefore positive definite.

5.4.11. Let $\mathbf{B} = \sqrt{\mathbf{A}}$, as defined in Example 1 on page 324. Then $\mathbf{B} \in \mathrm{M}_n(\mathbb{F})$ is Hermitian and positive definite, so $\mathbf{A} = \mathbf{B}^2 = \mathbf{B}^*\mathbf{B}$. Now let $\mathbf{B} = \mathbf{Q}\mathbf{X}$ be a QR decomposition of $\mathbf{B}$, with $\mathbf{Q} \in \mathrm{M}_n(\mathbb{F})$ unitary and $\mathbf{X} \in \mathrm{M}_n(\mathbb{F})$ upper triangular. Then

$$\mathbf{A} = \mathbf{B}^*\mathbf{B} = (\mathbf{Q}\mathbf{X})^*\mathbf{Q}\mathbf{X} = \mathbf{X}^*\mathbf{Q}^*\mathbf{Q}\mathbf{X} = \mathbf{X}^*\mathbf{X}.$$

5.4.12. By the Spectral Theorem there is an orthonormal basis $(e_1, \dots, e_n)$ of V consisting of eigenvectors of $\boldsymbol{T}$; say $\boldsymbol{T}e_j = \lambda_j$. Then for any $v \in V$, $v = \sum_{j=1}^n \langle v, e_j\rangle\, e_j$, and so

$$\langle \boldsymbol{T}v, v\rangle = \sum_{j,k=1}^n \langle v, e_j\rangle\,\overline{\langle v, e_k\rangle}\,\langle \boldsymbol{T}e_j, e_k\rangle = \sum_{j,k=1}^n \langle v, e_j\rangle\,\overline{\langle v, e_k\rangle}\,\langle \lambda_j e_j, e_k\rangle = \sum_{j=1}^n |\langle v, e_j\rangle|^2\,\lambda_j \geq 0.$$

It follows similarly that $\langle \boldsymbol{S}v, v\rangle \geq 0$. Therefore for any $v \in V$,

$$\langle (\boldsymbol{S}+\boldsymbol{T})v, v\rangle = \langle \boldsymbol{S}v, v\rangle + \langle \boldsymbol{T}v, v\rangle \geq 0.$$

In particular, if λ is an eigenvalue of $\boldsymbol{S}+\boldsymbol{T}$ with corresponding eigenvector v, then

$$0 \leq \langle (\boldsymbol{S}+\boldsymbol{T})v, v\rangle = \langle \lambda v, v\rangle = \lambda\,\|v\|^2\,,$$

and so $\lambda \geq 0$.

5.4.13. The linearity properties are simple, and $\langle y, x\rangle = \mathbf{x}^*\mathbf{A}\mathbf{y} = (\mathbf{y}^*\mathbf{A}^*\mathbf{x})^* = (\mathbf{y}^*\mathbf{A}\mathbf{x})^* = \overline{\mathbf{y}^*\mathbf{A}\mathbf{x}}$ since $\mathbf{A}$ is Hermitian, and $\mathbf{y}^*\mathbf{A}\mathbf{x} \in \mathrm{M}_1(\mathbb{C})$ is just a complex number, so its adjoint is just its complex conjugate.

Finally, by Theorem 5.20, $\langle \mathbf{x}, \mathbf{y}\rangle = \mathbf{x}^*\mathbf{A}\mathbf{x} > 0$ for $\mathbf{x} \neq \mathbf{0}$.

5.4.14. By the Spectral Theorem for normal operators, there exist an orthonormal basis $(e_1, \ldots, e_n)$ of V and numbers $\lambda_1, \ldots, \lambda_n \in \mathbb{C}$ such that for each j, $\boldsymbol{T} e_j = \lambda_j e_j$. For each j, pick a $\mu_j \in \mathbb{C}$ such that $\mu_j^2 = \lambda_j$, and define $\boldsymbol{S} \in \mathcal{L}(V)$ by setting $\boldsymbol{S} e_j = \mu_j e_j$ and extending by linearity. Then

$$\boldsymbol{S}^2 e_j = \boldsymbol{S}(\mu_j e_j) = \mu_j^2 e_j = \lambda_j e_j = \boldsymbol{T} e_j.$$

Thus $\boldsymbol{S}^2$ and $\boldsymbol{T}$ agree on a basis, so by the uniqueness part of the theorem on extending by linearity, $\boldsymbol{S}^2 = \boldsymbol{T}$.

Remark: We can't simply say $\mu_j = \sqrt{\lambda_j}$ above, since the square root of a complex number is not uniquely defined.

5.4.15. (a) Say

$$\mathbf{A} = \begin{bmatrix} a_1 & a_2 & a_3 & \cdots & a_{n-1} & a_n \\ a_n & a_1 & a_2 & \cdots & a_{n-2} & a_{n-1} \\ a_{n-1} & a_n & a_1 & \ddots & & \vdots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ a_3 & a_4 & & \ddots & a_1 & a_2 \\ a_2 & a_3 & \cdots & \cdots & a_n & a_1 \end{bmatrix}, \mathbf{B} = \begin{bmatrix} b_1 & b_2 & b_3 & \cdots & b_{n-1} & b_n \\ b_n & b_1 & b_2 & \cdots & b_{n-2} & b_{n-1} \\ b_{n-1} & b_n & b_1 & \ddots & & \vdots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ b_3 & b_4 & & \ddots & b_1 & b_2 \\ b_2 & b_3 & \cdots & \cdots & b_n & b_1 \end{bmatrix}.$$

If we define $a_0 = a_n$, $a_{-1} = a_{n-1}$, and so on until $a_{-(n-1)} = a_1$, and similarly for $b_0, \ldots, b_{-(n-1)}$, then we can more succinctly write

$$[\mathbf{A}]_{jk} = a_{k-j+1} \qquad \text{and} \qquad [\mathbf{B}]_{jk} = b_{k-j+1}.$$

Then

$$[\mathbf{AB}]_{jk} = \sum_{\ell=1}^{n} [\mathbf{A}]_{j\ell} [\mathbf{B}]_{\ell k} = \sum_{\ell=1}^{n} a_{\ell-j+1} b_{k-\ell+1}$$

and

$$[\mathbf{BA}]_{jk} = \sum_{\ell=1}^{n} [\mathbf{B}]_{j\ell} [\mathbf{A}]_{\ell k} = \sum_{\ell=1}^{n} b_{\ell-j+1} a_{k-\ell+1}.$$

In each of these sums, we have n terms, in which each distinct a and each distinct b appear one time. Moreover, in each sum a_m is multiplied by $b_{k-j+m+2}$. Therefore

$$[\mathbf{AB}]_{jk} = [\mathbf{BA}]_{jk}$$

for each j and k.

(b) If $\mathbf{A}$ is a circulant matrix with first row $a_1, \ldots, a_n$, then $\mathbf{A}^*$ is a circulant matrix with first row $\overline{a_1}, \overline{a_n}, \ldots, \overline{a_2}$. Therefore $\mathbf{A}\mathbf{A}^* = \mathbf{A}^*\mathbf{A}$ by part (a).

5.4.16. This is equivalent to showing that each column of $\mathbf{F}$ is an eigenvector of $\mathbf{C}$. Suppose that the first row of $\mathbf{C}$ has entries $c_1, \ldots, c_n$ and define $c_0, \ldots, c_{-(n-1)}$ as in the solution to Exercise 5.4.16. If $\mathbf{f}_j$ is the jth column of $\mathbf{F}$, then the kth entry of $\mathbf{C}\mathbf{f}_j$ is

$$\frac{1}{\sqrt{n}} \sum_{l=1}^{n} [\mathbf{C}]_{k\ell} \omega^{\ell j} = \frac{\omega^{jk}}{\sqrt{n}} \sum_{l=1}^{n} c_{\ell-k+1} \omega^{j(\ell-k)} = \frac{\omega^{jk}}{\sqrt{n}} \sum_{m=1-k}^{n-k} c_{m-1} \omega^{jm} = \frac{\omega^{jk}}{\sqrt{n}} \sum_{m=1}^{n} c_{m-1} \omega^{jm},$$

where the last equality holds because $c_{m-1} = c_{m-1+n}$ and $\omega^{jm} = \omega^{j(m+n)}$. Therefore

$$\mathbf{C}\mathbf{f}_j = \left(\sum_{m=1}^{n} c_{m-1}\omega^{jm} \right) \mathbf{f}_j.$$

5.4.17. (a) Suppose $(\mathbf{u}_1, \dots, \mathbf{u}_n)$ is an orthonormal basis of eigenvectors of $\mathbf{A}$ with corresponding eigenvalues $\lambda_1, \dots, \lambda_n$. Then the matrix $\mathbf{U}$ with columns $\mathbf{u}_1, \dots, \mathbf{u}_n$ is unitary and then $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{U}^{-1} = \mathbf{U}\mathbf{D}\mathbf{U}^*$, where $\mathbf{D} = \mathbf{diag}(\lambda_1, \dots, \lambda_n)$. Then $\mathbf{A}^* = \mathbf{U}\mathbf{D}^*\mathbf{U}^* = \mathbf{U}\mathbf{D}\mathbf{U}^*$ since the λ_j are real.

(b) Suppose $(\mathbf{u}_1, \dots, \mathbf{u}_n)$ is an orthonormal basis of eigenvectors of $\mathbf{A}$ with corresponding eigenvalues $\lambda_1, \dots, \lambda_n$. Then the matrix $\mathbf{U} \in \mathbb{R}^n$ with columns $\mathbf{u}_1, \dots, \mathbf{u}_n$ is orthogonal and then $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{U}^{-1} = \mathbf{U}\mathbf{D}\mathbf{U}^{\mathrm{T}}$, where $\mathbf{D} = \mathbf{diag}(\lambda_1, \dots, \lambda_n)$. Then $\mathbf{A}^{\mathrm{T}} = \mathbf{U}\mathbf{D}^{\mathrm{T}}\mathbf{U}^{\mathrm{T}} = \mathbf{U}\mathbf{D}\mathbf{U}^{\mathrm{T}}$.

5.4.18. Since $\mathbf{U}$ is unitary, $\mathbf{U}\mathbf{U}^* = \mathbf{U}^*\mathbf{U} = \mathbf{I}$, so $\mathbf{U}$ is normal. The Spectral Theorem then guarantees the existence of a unitary $\mathbf{V}$ such that $\mathbf{U} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^*$, where $\mathbf{\Lambda}$ is diagonal and its diagonal entries are the eigenvalues of $\mathbf{U}$. Since $\mathbf{U}$ is unitary, any eigenvalue has modulus by Exercise 4.5.12.

We cannot expect that $\mathbf{V}$ will be orthogonal if $\mathbf{U}$ is orthogonal: if $\mathbf{V}$ is orthogonal (so has real entries), then $\mathbf{V}^*\mathbf{U}\mathbf{V}$ has real entries. The eigenvalues of $\mathbf{U}$ need not be real. In particular, we have seen that the orthogonal matrix $\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ has no real eigenvalues.

5.4.19. (a) If $\boldsymbol{T}$ is self-adjoint, then by the Spectral Theorem, there is an orthonormal basis $(v_1, \dots, v_n)$ of V consisting of eigenvectors of $\boldsymbol{T}$; therefore the eigenspaces of $\boldsymbol{T}$ span V. Since $\boldsymbol{T}$ is self-adjoint, the distinct eigenspaces are mutually orthogonal by Proposition 5.15. Therefore V is the orthogonal direct sum of the eigenspaces of $\boldsymbol{T}$.

Now suppose that V is the orthogonal direct sum of the eigenspaces of $\boldsymbol{T}$. Each eigenspace then has an orthonormal basis; putting those bases together produces an orthonormal basis $\mathcal{B}$ of V consisting of eigenvectors of $\boldsymbol{T}$. The matrix $[\boldsymbol{T}]_{\mathcal{B}}$ is diagonal, hence symmetric, so by Theorem 5.13, $\boldsymbol{T}$ is self-adjoint.

(b) The proof is the same as in part (a), using Exercise 5.4.22 in place of Proposition 5.15.

(c) The proof is the same as in part (a), in this case using the fact that $[\boldsymbol{T}]_{\mathcal{B}}$ is not only diagonal but also has real entries.

5.4.20. **Solution 1:** Since $\mathbf{A}$ is Hermitian, it represents a self-adjoint linear map on $\mathbb{F}^n$. By the spectral theorem for linear maps, there are an orthonormal basis $(\mathbf{v}_1, \dots, \mathbf{v}_n)$ of $\mathbb{F}^n$ and real numbers $\lambda_1, \dots, \lambda_n$ such that $\mathbf{A}\mathbf{v}_k = \lambda_k\mathbf{v}_k$ for each j. By orthonormality,

$$\left(\sum_{j=1}^{n} \lambda_j \mathbf{v}_j \mathbf{v}_j^* \right) \mathbf{v}_k = \sum_{j=1}^{n} \lambda_j \mathbf{v}_j \mathbf{v}_j^* \mathbf{v}_k = \sum_{j=1}^{n} \lambda_j \langle \mathbf{v}_k, \mathbf{v}_j \rangle \mathbf{v}_j = \lambda_k \mathbf{v}_k.$$

Thus $\mathbf{A}$ and $\sum_{j=1}^{n} \lambda_j \mathbf{v}_j \mathbf{v}_j^*$ have the same action on a basis of $\mathbb{F}^n$, so they are the same.

Solution 2: By the spectral theorem for Hermitian matrices, there are a unitary matrix $\mathbf{V} \in \mathrm{M}_n(\mathbb{F})$ and real numbers $\lambda_1, \dots, \lambda_n$ such that

$$\mathbf{A} = \mathbf{V}\,\mathbf{diag}(\lambda_1, \dots, \lambda_n)\mathbf{V}^*.$$

From this it follows that the entries of $\mathbf{A}$ are

$$a_{jk} = \sum_{\ell=1}^{n}\sum_{m=1}^{n} v_{j\ell}[\mathbf{diag}(\lambda_1, \dots, \lambda_n)]_{\ell m}\overline{v_{km}} = \sum_{\ell=1}^{n} \lambda_\ell v_{j\ell}\overline{v_{m\ell}},$$

which are the same as the entries of $\sum_{\ell=1}^{n} \lambda_\ell \mathbf{v}_\ell \mathbf{v}_\ell^*$ if $\mathbf{v}_1, \dots, \mathbf{v}_n$ are the (orthonormal) columns of $\mathbf{V}$.

5.4.21. Let $\lambda_1, \dots, \lambda_m$ be the distinct eigenvalues of $\boldsymbol{T}$, which are all real since $\boldsymbol{T}$ is self-adjoint, and let $U_j = \mathrm{Eig}_{\lambda_j}(\boldsymbol{T})$. By the spectral theorem for self-adjoint operators there is an orthonormal basis of V consisting of eigenvectors of $\boldsymbol{T}$. All the basis vectors corresponding to the eigenvalue $lambda_j$ lie in U_j, and are orthonormal. Furthermore, any vector in U_j is perpendicular to any vector in all the other U_ks. Thus the spaces $U_1, \dots, U_m$ are orthogonal to each other. Since every vector in V is a sum of vectors in these subspaces, $V = U_1 \oplus \cdots \oplus U_m$ is an orthogonal direct sum.

Furthermore, if e is one of the orthonormal basis vectors given by the spectral theorem, corresponding to eigenvalue λ_k, then

$$\boldsymbol{T}e = \lambda_k e \qquad \text{and} \qquad \sum_{j=1}^{m} \lambda_j \boldsymbol{P}_{U_j} e = \lambda_k e,$$

since $e \in U_k$ and e is orthogonal to every other U_j. So by the uniqueness of extensions by linearity,

$$\boldsymbol{T} = \sum_{j=1}^{m} \lambda_j \boldsymbol{P}_{U_j}.$$

5.4.22. By the Spectral Theorem for normal matrices, $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{U}^*$ for a unitary matrix $\mathbf{U} \in \mathrm{M}_n(\mathbb{C})$ and a diagonal matrix $\mathbf{D} = \mathbf{diag}(d_1, \dots, d_n)$ whose entries are eigenvalues of $\mathbf{A}$. Suppose that $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$ and $\mathbf{A}\mathbf{w} = \mu\mathbf{w}$. Define $\mathbf{x} = \mathbf{U}^*\mathbf{v}$ and $\mathbf{y} = \mathbf{U}^*\mathbf{w}$. Then $\mathbf{D}\mathbf{x} = \lambda\mathbf{x}$ and $\mathbf{D}\mathbf{y} = \mu\mathbf{y}$, so for each j,

$$d_j x_j = \lambda x_j \qquad \text{and} \qquad d_j y_j = \mu y_j.$$

This means that $x_j = 0$ for all j except those for which $d_j = \lambda$, and $y_j = 0$ for all j except those for which $d_j = \mu$. Therefore if $\lambda \neq \mu$, then

$$0 = \langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{U}^*\mathbf{v}, \mathbf{U}^*\mathbf{w} \rangle = \langle \mathbf{v}, \mathbf{w} \rangle,$$

so $\mathbf{v}$ and $\mathbf{w}$ are orthogonal.

5.4.23. (a) If $v \in U_j$, then $v = \sum_{k=j}^{n} \langle v, e_k \rangle e_k$, so

$$\begin{aligned}
\langle \boldsymbol{T}v, v \rangle &= \sum_{k,\ell=j}^{n} \langle v, e_k \rangle \overline{\langle v, e_\ell \rangle} \langle \boldsymbol{T}e_k, e_\ell \rangle = \sum_{k,\ell=j}^{n} \langle v, e_k \rangle \overline{\langle v, e_\ell \rangle} \langle \lambda_k e_k, e_\ell \rangle \\
&= \sum_{k=j}^{n} \left| \langle v, e_k \rangle \right|^2 \lambda_k \leq \lambda_j \sum_{k=j}^{n} \left| \langle v, e_k \rangle \right|^2 = \lambda_j \left\| v \right\|^2.
\end{aligned}$$

(b) If $v \in V_j$, then $v = \sum_{k=1}^{j} \langle v, e_k \rangle e_k$, so

$$\begin{aligned}
\langle \boldsymbol{T}v, v \rangle &= \sum_{k,\ell=1}^{j} \langle v, e_k \rangle \overline{\langle v, e_\ell \rangle} \langle \boldsymbol{T}e_k, e_\ell \rangle = \sum_{k,\ell=1}^{j} \langle v, e_k \rangle \overline{\langle v, e_\ell \rangle} \langle \lambda_k e_k, e_\ell \rangle \\
&= \sum_{k=1}^{j} \left| \langle v, e_k \rangle \right|^2 \lambda_k \geq \lambda_j \sum_{k=1}^{j} \left| \langle v, e_k \rangle \right|^2 = \lambda_j \left\| v \right\|^2.
\end{aligned}$$

(c) $\dim V_j + \dim U = n + 1$, so by Lemma 3.22 there exists a nonzero vector $v \in V_j \cap U$. By part (b), any such vector satisfies $\langle \boldsymbol{T}v, v \rangle \geq \lambda_j \left\| v \right\|^2$.

(d) By part (c), if U is any $(n-j+1)$-dimensional subspace of V, then $\max_{v \in U,\ \|v\|=1} \langle \boldsymbol{T}v, v \rangle \geq \lambda_j$. Therefore

$$\min_{\dim U = n-j+1} \max_{v \in U,\ \|v\|=1} \langle \boldsymbol{T}v, v \rangle \geq \lambda_j.$$

On the other hand, $\dim U_j = n - j + 1$, and by part (a), $\max_{v \in U_j,\ \|v\|=1} \langle \boldsymbol{T}v, v \rangle \leq \lambda_j$. Therefore

$$\min_{\dim U = n-j+1} \max_{v \in U,\ \|v\|=1} \langle \boldsymbol{T}v, v \rangle \leq \max_{v \in U_j,\ \|v\|=1} \langle \boldsymbol{T}v, v \rangle \leq \lambda_j.$$

5.4.24. Let $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{U}^*$ be a spectral decomposition of $\mathbf{A}$. Then $\mathbf{A} - \lambda\mathbf{I}_n = \mathbf{U}(\mathbf{D} - \lambda\mathbf{I}_n)\mathbf{U}^*$, and so

$$\dim \ker(\mathbf{A} - \lambda\mathbf{I}_n) = \dim \ker(\mathbf{D} - \lambda\mathbf{I}_n).$$

Since $\mathbf{D} - \lambda\mathbf{I}_n$ is diagonal, the dimension of its kernel is the number of 0s on the diagonal, which is the number of times λ appears on the diagonal of $\mathbf{D}$.

5.4.25. By Schur decomposition, $\mathbf{A} = \mathbf{U}\mathbf{T}\mathbf{U}^*$ for some unitary $\mathbf{U} \in \mathrm{M}_n(\mathbb{C})$ and upper triangular $\mathbf{T} \in \mathrm{M}_n(\mathbb{C})$. Since $\mathbf{A}$ and $\mathbf{T}$ are similar, they have the same eigenvalues, so by Theorem 4.2 the diagonal entries of $\mathbf{T}$ are $\lambda_1, \ldots, \lambda_n$ (though not necessarily in that order). Therefore

$$\left\| \mathbf{T} \right\|_F = \sqrt{\sum_{j,k=1}^{n} \left| t_{jk} \right|^2} \geq \sqrt{\sum_{j=1}^{n} \left| t_{jj} \right|^2} = \sqrt{\sum_{j=1}^{n} \left| \lambda_j \right|^2}.$$

On the other hand,

$$\begin{aligned}
\left\| \mathbf{A} \right\|_F^2 &= \operatorname{tr} \mathbf{A}\mathbf{A}^* = \operatorname{tr} \mathbf{U}\mathbf{T}\mathbf{U}^*(\mathbf{U}\mathbf{T}\mathbf{U}^*)^* = \operatorname{tr} \mathbf{U}\mathbf{T}\mathbf{U}^*\mathbf{U}\mathbf{T}^*\mathbf{U}^* \\
&= \operatorname{tr} \mathbf{U}\mathbf{T}\mathbf{T}^*\mathbf{U}^* = \operatorname{tr} \mathbf{T}\mathbf{T}^* = \left\| \mathbf{T} \right\|_F^2.
\end{aligned}$$

Putting these together,

$$\sqrt{\sum_{j=1}^{n} |\lambda_j|^2} \leq \|\mathbf{A}\|_F .$$

5.4.26. Suppose that $\mathbf{T} \in \mathrm{M}_n(\mathbb{C})$ is upper triangular. We can find distinct numbers $s_{11}, \ldots, s_{nn}$ such that for each j, $|t_{jj} - s_{jj}| \leq \frac{\varepsilon}{\sqrt{n}}$. If we define $\mathbf{S} \in \mathrm{M}_n(C)$ with these diagonal entries and $s_{jk} = t_{jk}$ for $j \neq k$, then $\mathbf{S}$ is upper triangular with *distinct* eigenvalues $s_{11}, \ldots, s_{nn}$, and

$$\|\mathbf{T} - \mathbf{S}\| = \sqrt{\sum_{j=1}^{n} |t_{jj} - s_{jj}|^2} \leq \sqrt{\sum_{j=1}^{n} \frac{\varepsilon^2}{n}} = \varepsilon.$$

Now for an arbitrary $\mathbf{A} \in \mathrm{M}_n(\mathbb{C})$, let $\mathbf{A} = \mathbf{U}\mathbf{T}\mathbf{U}^*$ be the Schur decomposition of $\mathbf{A}$. Define $\mathbf{S}$ as above, and define $\mathbf{B} = \mathbf{U}\mathbf{S}\mathbf{U}^*$. Then $\mathbf{B}$ is similar to $\mathbf{S}$, so it has n distinct eigenvalues because $\mathbf{S}$ does, and

$$\begin{aligned}\|\mathbf{A} - \mathbf{B}\|_F^2 &= \|\mathbf{U}(\mathbf{T} - \mathbf{S})\mathbf{U}^*\|^2 = \mathrm{tr}[\mathbf{U}(\mathbf{T} - \mathbf{S})\mathbf{U}^*][\mathbf{U}(\mathbf{T} - \mathbf{S})^*\mathbf{U}^*] \\ &= \mathrm{tr}(\mathbf{T} - \mathbf{S})(\mathbf{T} - \mathbf{S})^* = \|\mathbf{T} - \mathbf{S}\|_F^2 \leq \varepsilon^2.\end{aligned}$$

5.4.27. (a) Proof by induction: If $\mathbf{A} \in \mathrm{M}_1(\mathbb{C})$ is upper triangular (which it always is), then $\mathbf{A}$ is diagonal.

Now suppose that if $\mathbf{A} \in \mathrm{M}_{n-1}(\mathbb{C})$ is upper triangular and normal, then $\mathbf{A}$ is diagonal. Let $\mathbf{A} \in \mathrm{M}_n(\mathbb{C})$ be upper triangular and normal. Observe that if the top-left entries of $\mathbf{A}\mathbf{A}^*$ and $\mathbf{A}^*\mathbf{A}$ are the same, then

$$\sum_{k=1}^{n} a_{1k}\overline{a}_{1k} = \sum_{k=1}^{n} a_{k1}\overline{a}_{k1}.$$

But $a_{k1} = 0$ for $k > 1$ since $\mathbf{A}$ is upper triangular, and so the right-hand side is just $|a_{11}|^2$. Since the left-hand side is

$$|a_{11}|^2 + \sum_{k=2}^{n} |a_{1k}|^2 ,$$

it must be that $a_{1k} = 0$ for $k \geq 2$.

For $b, c \in \mathbb{F}$ and $\mathbf{B}, \mathbf{C} \in \mathrm{M}_{n-1}(\mathbb{C})$,

$$\begin{bmatrix} b & \mathbf{0} \\ \mathbf{0} & \mathbf{B} \end{bmatrix} \begin{bmatrix} c & \mathbf{0} \\ \mathbf{0} & \mathbf{C} \end{bmatrix} = \begin{bmatrix} bc & \mathbf{0} \\ \mathbf{0} & \mathbf{BC} \end{bmatrix}.$$

We have just seen that our matrix $\mathbf{A}$ has the form

$$\begin{bmatrix} a_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}' \end{bmatrix}$$

for an $(n-1)\times(n-1)$ upper-triangular matrix $\mathbf{A}'$. Moreover, by the equation above and the fact that $\mathbf{A}$ is normal, it follows that $\mathbf{A}'$ is itself normal; therefore, by the induction hypothesis, $\mathbf{A}$ is diagonal.

(b) By the Schur decomposition, if $\mathbf{A} \in \mathrm{M}_n(\mathbb{C})$, then there is a unitary matrix $\mathbf{U}$ and an upper triangular matrix $\mathbf{T}$ such that $\mathbf{A} = \mathbf{U}\mathbf{T}\mathbf{U}^*$. If $\mathbf{A}$ is normal,

$$\mathbf{U}\mathbf{T}\mathbf{T}^*\mathbf{U}^* = \mathbf{U}\mathbf{T}\mathbf{U}^*\mathbf{U}\mathbf{T}^*\mathbf{U}^* = \mathbf{A}\mathbf{A}^* = \mathbf{A}^*\mathbf{A} = \mathbf{U}\mathbf{T}^*\mathbf{U}^*\mathbf{U}\mathbf{T}\mathbf{U}^* = \mathbf{U}\mathbf{T}^*\mathbf{T}\mathbf{U}^*.$$

Multiplying on the left by $\mathbf{U}^*$ and the right by $\mathbf{U}$ shows that $\mathbf{T}\mathbf{T}^* = \mathbf{T}^*\mathbf{T}$. By the previous part, this means that $\mathbf{T}$ is diagonal, and hence we are done.

Chapter 6

Determinants

6.1 Determinants

6.1.1. (a) $\det \mathbf{A} = -2$ by formula (6.1).

(b) $\det \mathbf{A} = 6 + 6i$ by formula (6.1).

(c) By formulas (6.1) and (6.5),

$$\det \begin{bmatrix} 0 & 2 & -1 \\ 3 & 0 & 4 \\ -2 & 1 & 2 \end{bmatrix} = -2 \det \begin{bmatrix} 3 & 4 \\ -2 & 2 \end{bmatrix} - \det \begin{bmatrix} 3 & 0 \\ -2 & 1 \end{bmatrix} = -2(14) - 3 = -31.$$

(d) By formulas (6.1) and (6.5),

$$\det \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{bmatrix} = \det \begin{bmatrix} 2 & 2 \\ 2 & 3 \end{bmatrix} - \det \begin{bmatrix} 1 & 2 \\ 1 & 3 \end{bmatrix} + \det \begin{bmatrix} 1 & 2 \\ 2 & 2 \end{bmatrix} = 2 - 1 + 0 = 1.$$

6.1.2. (a) $\det \mathbf{A} = 2$ by formula (6.1).

(b) $\det \mathbf{A} = -5 + i$ by formula (6.1).

(c) By formulas (6.1) and (6.5),

$$\det \begin{bmatrix} -1 & 2 & 3 \\ 4 & -5 & 6 \\ 7 & -8 & 9 \end{bmatrix} = -\det \begin{bmatrix} -5 & 6 \\ -8 & 9 \end{bmatrix} - 2 \det \begin{bmatrix} 4 & 6 \\ 7 & 9 \end{bmatrix} + 3 \det \begin{bmatrix} 4 & -5 \\ 7 & -8 \end{bmatrix} = -3 - 2(-6) + 3(3) = 18.$$

(d) By formulas (6.1) and (6.5),

$$\det \begin{bmatrix} 1 & 0 & -2 \\ -3 & 0 & 2 \\ 0 & 4 & -1 \end{bmatrix} = \det \begin{bmatrix} 0 & 2 \\ 4 & -1 \end{bmatrix} - 2 \det \begin{bmatrix} -3 & 0 \\ 0 & 4 \end{bmatrix} = -8 - 2(-12) = 16.$$

6.1.3. (a) By multilinearity applied first to the first column and then to the second column,

$$\begin{aligned} D\left(\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}\right) &= D\left(\begin{bmatrix} 1 & 2 \\ 0 & 4 \end{bmatrix}\right) + 3D\left(\begin{bmatrix} 0 & 2 \\ 1 & 4 \end{bmatrix}\right) \\ &= 2D\left(\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}\right) + 4D\left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right) + 6D\left(\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\right) + 8D\left(\begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix}\right) \end{aligned}$$

By the alternating property and Lemma 6.2, this is

$$2 \cdot 0 + 4D(\mathbf{I}_2) - 6D(\mathbf{I}_2) + 8 \cdot 0 = -14.$$

(b) By multilinearity applied first to the first column and then to the second column,

$$\begin{aligned} D\left(\begin{bmatrix} -2 & 3 \\ 5 & -4 \end{bmatrix}\right) &= -2D\left(\begin{bmatrix} 1 & 3 \\ 0 & -4 \end{bmatrix}\right) + 5D\left(\begin{bmatrix} 0 & 3 \\ 1 & -4 \end{bmatrix}\right) \\ &= -6D\left(\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}\right) + 8D\left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right) + 15D\left(\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\right) - 20D\left(\begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix}\right) \end{aligned}$$

By the alternating property and Lemma 6.2, this is

$$-6 \cdot 0 + 8D(\mathbf{I}_2) - 15D(\mathbf{I}_2) - 20 \cdot 0 = -49.$$

6.1.4. (a) By multilinearity applied to the second column,

$$D\left(\begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{bmatrix}\right) = 2D\left(\begin{bmatrix} 1 & 1 & 3 \\ 0 & 0 & 5 \\ 0 & 0 & 6 \end{bmatrix}\right) + 4D\left(\begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & 5 \\ 0 & 0 & 6 \end{bmatrix}\right) = 4D\left(\begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & 5 \\ 0 & 0 & 6 \end{bmatrix}\right).$$

Similarly, applying multilinearity to the third column gives

$$\begin{aligned} D\left(\begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & 5 \\ 0 & 0 & 6 \end{bmatrix}\right) &= 3D\left(\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}\right) + 5D\left(\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}\right) + 6D\left(\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\right) \\ &= 6D\left(\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\right), \end{aligned}$$

and so

$$D\left(\begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{bmatrix}\right) = 24D(\mathbf{I}_3) = 144.$$

(b) By Corollary 6.4,

$$D\left(\begin{bmatrix} 3 & 2 & 4 \\ 2 & 1 & 2 \\ 2 & 0 & 1 \end{bmatrix}\right) = 6\det\left(\begin{bmatrix} 3 & 2 & 4 \\ 2 & 1 & 2 \\ 2 & 0 & 1 \end{bmatrix}\right).$$

The latter determinant can be computed using formulas (6.5) and (6.1); using $i = 3$ in (6.5), we get

$$\det\begin{bmatrix} 3 & 2 & 4 \\ 2 & 1 & 2 \\ 2 & 0 & 1 \end{bmatrix} = 2\det\begin{bmatrix} 2 & 4 \\ 1 & 2 \end{bmatrix} - 0\det\begin{bmatrix} 3 & 4 \\ 2 & 2 \end{bmatrix} + \det\begin{bmatrix} 3 & 2 \\ 2 & 1 \end{bmatrix} = -1,$$

and so $D(\mathbf{A}) = -6$.

6.1.5. (a) Writing $\mathbf{A} = \begin{bmatrix}\mathbf{a}_1 & \mathbf{a}_2\end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix}\mathbf{b}_1 & \mathbf{b}_2\end{bmatrix}$, then

$$\begin{aligned} D(\mathbf{A}+\mathbf{B}) &= D\left(\begin{bmatrix}\mathbf{a}_1+\mathbf{b}_1 & \mathbf{a}_2+\mathbf{b}_2\end{bmatrix}\right) \\ &= D\left(\begin{bmatrix}\mathbf{a}_1 & \mathbf{a}_2+\mathbf{b}_2\end{bmatrix}\right) + D\left(\begin{bmatrix}\mathbf{b}_1 & \mathbf{a}_2+\mathbf{b}_2\end{bmatrix}\right) \\ &= D\left(\begin{bmatrix}\mathbf{a}_1 & \mathbf{a}_2\end{bmatrix}\right) + D\left(\begin{bmatrix}\mathbf{a}_1 & \mathbf{b}_2\end{bmatrix}\right) + D\left(\begin{bmatrix}\mathbf{b}_1 & \mathbf{a}_2\end{bmatrix}\right) + D\left(\begin{bmatrix}\mathbf{b}_1 & \mathbf{b}_2\end{bmatrix}\right) \\ &= D(\mathbf{A}) + D\left(\begin{bmatrix}\mathbf{a}_1 & \mathbf{b}_2\end{bmatrix}\right) + D\left(\begin{bmatrix}\mathbf{b}_1 & \mathbf{a}_2\end{bmatrix}\right) + D(\mathbf{B}), \end{aligned}$$

where the second equality follows from linearity in the first column, and the third equality follows from linearity in the second column.

(b) In the same way, if $\mathbf{A} = \begin{bmatrix}\mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3\end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix}\mathbf{b}_1 & \mathbf{b}_2 & \mathbf{b}_3\end{bmatrix}$, then

$$\begin{aligned} D(\mathbf{A}+\mathbf{B}) &= D\left(\begin{bmatrix}\mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3\end{bmatrix}\right) + D\left(\begin{bmatrix}\mathbf{a}_1 & \mathbf{a}_2 & \mathbf{b}_3\end{bmatrix}\right) + D\left(\begin{bmatrix}\mathbf{a}_1 & \mathbf{b}_2 & \mathbf{a}_3\end{bmatrix}\right) \\ &\quad + D\left(\begin{bmatrix}\mathbf{a}_1 & \mathbf{b}_2 & \mathbf{b}_3\end{bmatrix}\right) + D\left(\begin{bmatrix}\mathbf{b}_1 & \mathbf{a}_2 & \mathbf{a}_3\end{bmatrix}\right) + D\left(\begin{bmatrix}\mathbf{b}_1 & \mathbf{a}_2 & \mathbf{b}_3\end{bmatrix}\right) \\ &\quad + D\left(\begin{bmatrix}\mathbf{b}_1 & \mathbf{b}_2 & \mathbf{a}_3\end{bmatrix}\right) + D\left(\begin{bmatrix}\mathbf{b}_1 & \mathbf{b}_2 & \mathbf{b}_3\end{bmatrix}\right). \end{aligned}$$

6.1.6. The first matrix has determinant -19 and the second has determinant 23, so by Corollary 6.7, they are not similar.

6.1.7. By part 5 of Theorem 4.16, there is a basis of V such that $[\boldsymbol{P}_U] = \mathbf{diag}(1, 1, \ldots, 1, 0, \ldots, 0)$. (Since U is a proper subspace of V, there are in fact some 0 entries on the diagonal.) By Example 3 on page 337, this implies that $\det \boldsymbol{P}_U = 0$.

6.1.8. There is a basis in which the matrix of $\boldsymbol{R}$ is $\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & -1\end{bmatrix}$ (choose the first two basis vectors to be in the plane, and the last one to be orthogonal to it). This means that $\det \boldsymbol{R} = -1$ by Example 3 on page 337.

6.1.9. Because $\boldsymbol{T}$ has n distinct eigenvalues, it has n linearly independent eigenvectors, which therefore form a basis $\mathcal{B}$ of V. Then $[\boldsymbol{T}]_{\mathcal{B}} = \mathbf{diag}(\lambda_1, \ldots, \lambda_n)$. By Example 3 on page 337, $\det \boldsymbol{T} = \det \mathbf{diag}(\lambda_1, \ldots, \lambda_n) = \lambda_1 \cdots \lambda_n \det(\mathbf{I}_n) = \lambda_1 \cdots \lambda_n$.

6.1.10. If $\mathbf{A} = \mathbf{BC}$ with $\mathbf{B} \in \mathrm{M}_{n,m}(\mathbb{C})$ and $\mathbf{C} \in \mathrm{M}_{m,n}(\mathbb{C})$, and $m < n$, then $\mathbf{C}$ must have a nontrivial kernel ($\operatorname{rank} \mathbf{C} \le m < n$, so $\operatorname{null} \mathbf{C} > 0$ by the Rank–Nullity Theorem). This means that $\mathbf{A}$ also has a non-trivial null space: if $\mathbf{x} \neq \mathbf{0}$ has $\mathbf{Cx} = \mathbf{0}$, then $\mathbf{Ax} = \mathbf{BCx} = \mathbf{0}$ as well. The matrix $\mathbf{A}$ is thus singular, and so $\det \mathbf{A} = 0$.

6.1.11. Since $\mathbf{A}$ is Hermitian, $\mathbf{A} = \mathbf{UDU}^*$ for some $\mathbf{U}$ unitary and $\mathbf{D}$ diagonal with real entries. By Corollary 6.7 and Example 3 on page 337, $\det(\mathbf{A}) = \det(\mathbf{D}) \in \mathbb{R}$.

6.1.12. By formula (6.5) with $i = 1$ and Example 1 on page 343,

$$\begin{aligned} \det \begin{bmatrix} a_{11} & a_{12} & a_{13}\\ a_{21} & a_{22} & a_{23}\\ a_{31} & a_{32} & a_{33}\end{bmatrix} &= a_{11} \det \begin{bmatrix} a_{22} & a_{23}\\ a_{32} & a_{33}\end{bmatrix} - a_{12} \det \begin{bmatrix} a_{21} & a_{23}\\ a_{31} & a_{33}\end{bmatrix} + a_{13} \det \begin{bmatrix} a_{21} & a_{22}\\ a_{31} & a_{32}\end{bmatrix} \\ &= a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{12}(a_{21}a_{33} - a_{23}a_{31}) + a_{13}(a_{21}a_{32} - a_{22}a_{31}). \end{aligned}$$

6.1.13. If there is some pair $i \neq j$ with $a_{ki} = a_{kj}$ for each k, then the inner sum in $f(\mathbf{A})$ is zero for that value of i and j. It follows that the product is 0.

$f(\mathbf{I}_2) = 2$, and $f\left(\begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}\right) = 3 \neq 2f(\mathbf{I}_2)$, so f is not multilinear.

6.1.14. Let $\mathbf{a}_1, \dots, \mathbf{a}_n, \mathbf{b}_1, \dots, \mathbf{b}_n \in \mathbb{R}^n$ and $c \in \mathbb{R}$. Then

$$\begin{aligned} g(\begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_k + \mathbf{b}_k & \cdots \mathbf{a}_n \end{bmatrix}) &= \left[\prod_{j \neq k} \left(\sum_{i=1}^n a_{ij}\right)\right] \times \left(\sum_{i=1}^n (a_{ik} + b_{ik})\right) \\ &= \left[\prod_{j=1}^n \left(\sum_{i=1}^n a_{ij}\right)\right] \times \left[\prod_{j \neq k} \left(\sum_{i=1}^n a_{ij}\right)\right] \times \left(\sum_{i=1}^n b_{ik}\right) \\ &= g(\begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_k & \cdots \mathbf{a}_n \end{bmatrix}) g(\begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{b}_k & \cdots \mathbf{a}_n \end{bmatrix}) \end{aligned}$$

and

$$\begin{aligned} g(\begin{bmatrix} \mathbf{a}_1 & \cdots & c\mathbf{a}_k & \cdots \mathbf{a}_n \end{bmatrix}) &= \left[\prod_{j \neq k} \left(\sum_{i=1}^n a_{ij}\right)\right] \times \left(\sum_{i=1}^n c a_{ik}\right) \\ &= c\left[\prod_{j=1}^n \left(\sum_{i=1}^n a_{ij}\right)\right] \\ &= cg(\begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_k & \cdots \mathbf{a}_n \end{bmatrix}). \end{aligned}$$

If $\mathbf{J} \in \mathrm{M}_n(\mathbb{R})$ is the matrix whose entries are all 1, then $g(\mathbf{J}) = n^n \neq 0$, so g is not isoscopic.

6.1.15. For any $j \in \{1, \dots, n\}$,

$$\begin{aligned} &D\left(\begin{bmatrix} | & & | & & | \\ \mathbf{a}_1 & \cdots & \mathbf{a}_j + \sum_{k \neq j} c_k \mathbf{a}_k & \cdots & \mathbf{a}_n \\ | & & | & & | \end{bmatrix}\right) \\ &\qquad = D(\mathbf{A}) + \sum_{k \neq j} c_k D\left(\begin{bmatrix} | & & | & & | \\ \mathbf{a}_1 & \cdots & \mathbf{a}_k & \cdots & \mathbf{a}_n \\ | & & | & & | \end{bmatrix}\right) = D(\mathbf{A}), \end{aligned}$$

since $D\left(\begin{bmatrix} | & & | & & | \\ \mathbf{a}_1 & \cdots & \mathbf{a}_k & \cdots & \mathbf{a}_n \\ | & & | & & | \end{bmatrix}\right) = 0$ for each k because the argument has a repeated column.

6.1.16. (a) Suppose $f, g \in V$ and $a \in \mathbb{F}$. If $\mathbf{a}_1, \dots, \mathbf{a}_n, \mathbf{b}_1, \dots, \mathbf{b}_n \in \mathbb{R}^n$ and $c \in \mathbb{R}$, then

$$\begin{aligned} &(f+g)(\begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_j + \mathbf{b}_j & \cdots & \mathbf{a}_n \end{bmatrix}) \\ &\quad = f(\begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_j + \mathbf{b}_j & \cdots & \mathbf{a}_n \end{bmatrix}) + g(\begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_j + \mathbf{b}_j & \cdots & \mathbf{a}_n \end{bmatrix}) \\ &\quad = f(\begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_j & \cdots & \mathbf{a}_n \end{bmatrix}) + f(\begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{b}_j & \cdots & \mathbf{a}_n \end{bmatrix}) \\ &\qquad + g(\begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_j & \cdots & \mathbf{a}_n \end{bmatrix}) + g(\begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{b}_j & \cdots & \mathbf{a}_n \end{bmatrix}) \\ &\quad = (f+g)(\begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_j & \cdots & \mathbf{a}_n \end{bmatrix}) + (f+g)(\begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{b}_j & \cdots & \mathbf{a}_n \end{bmatrix}) \end{aligned}$$

and

$$\begin{aligned}
&(f+g)(\begin{bmatrix}\mathbf{a}_1 & \cdots & c\mathbf{a}_j & \cdots & \mathbf{a}_n\end{bmatrix})\\
&\quad= f(\begin{bmatrix}\mathbf{a}_1 & \cdots & c\mathbf{a}_j & \cdots & \mathbf{a}_n\end{bmatrix}) + g(\begin{bmatrix}\mathbf{a}_1 & \cdots & c\mathbf{a}_j & \cdots & \mathbf{a}_n\end{bmatrix})\\
&\quad= cf(\begin{bmatrix}\mathbf{a}_1 & \cdots & \mathbf{a}_j & \cdots & \mathbf{a}_n\end{bmatrix}) + cg(\begin{bmatrix}\mathbf{a}_1 & \cdots & \mathbf{a}_j & \cdots & \mathbf{a}_n\end{bmatrix})\\
&\quad= c(f+g)(\begin{bmatrix}\mathbf{a}_1 & \cdots & \mathbf{a}_j & \cdots & \mathbf{a}_n\end{bmatrix}).
\end{aligned}$$

Therefore $f+g \in V$. We can show $af \in V$ similarly, and clearly $0 \in V$. Therefore V is a subspace of the vector space of all functions $\mathrm{M}_n(\mathbb{F}) \to \mathbb{F}$.

(b) For each $i = 1, \ldots, n$, define $g_i : \mathbb{F}^n \to \mathbb{F}$ by setting $g_i(\mathbf{e}_i) = 1$, $g_i(\mathbf{e}_j) = 0$, and extending by linearity. Then define $f_{i_1,\ldots,i_n} : \mathrm{M}_n(\mathbb{F}) \to \mathbb{F}$ for $1 \le i_j \le n$ by

$$f_{i_1,\ldots,i_n}\left(\begin{bmatrix}\mathbf{a}_1 & \cdots & \mathbf{a}_n\end{bmatrix}\right) = g_{i_1}(\mathbf{a}_1)\cdots g_{i_n}(\mathbf{a}_n).$$

We claim that the n^n functions $f_{i_1,\ldots,i_n}$ form a basis of V.

First observe that

$$\begin{aligned}
f_{i_1,\ldots,i_n}\left(\begin{bmatrix}\mathbf{a}_1 & \cdots & \mathbf{a}_n\end{bmatrix}\right) &= g_{i_1}(\mathbf{a}_1)\cdots g_{i_n}(\mathbf{a}_n)\\
&= g_{i_1}\left(\sum_{i=1}^n a_{i1}\mathbf{e}_i\right)\cdots g_{i_n}\left(\sum_{i=1}^n a_{in}\mathbf{e}_i\right)\\
&= a_{i_1 1}\cdots a_{i_n n}.
\end{aligned}$$

Now let $f \in V$. Then for any $\mathbf{A} \in \mathrm{M}_n(\mathbb{F})$,

$$\begin{aligned}
f(\mathbf{A}) &= f\left(\begin{bmatrix}\sum_{i=1}^n a_{i1}\mathbf{e}_i & \cdots & \sum_{i=1}^n a_{in}\mathbf{e}_i\end{bmatrix}\right)\\
&= \sum_{i_1=1}^n \cdots \sum_{i_n=1}^n a_{i_1 1}\cdots a_{i_n n} f\left(\begin{bmatrix}\mathbf{e}_{i_1} & \cdots & \mathbf{e}_{i_n}\end{bmatrix}\right)\\
&= \sum_{i_1=1}^n \cdots \sum_{i_n=1}^n f\left(\begin{bmatrix}\mathbf{e}_{i_1} & \cdots & \mathbf{e}_{i_n}\end{bmatrix}\right) f_{i_1,\ldots,i_n}\left(\begin{bmatrix}\mathbf{a}_1 & \cdots & \mathbf{a}_n\end{bmatrix}\right).
\end{aligned}$$

Therefore f is a linear combination of the $f_{i_1,\ldots,i_n}$, so those functions span V.

Now suppose that $c_{i_1,\ldots,i_n} \in \mathbb{F}$ for $1 \le i_j \le n$ satisfy

$$\sum_{i_1=1}^n \cdots \sum_{i_n=1}^n c_{i_1,\ldots,i_n} f_{i_1,\ldots,i_n} = 0.$$

Then for each choice of $1 \le j_k \le n$,

$$0 = \sum_{i_1=1}^n \cdots \sum_{i_n=1}^n c_{i_1,\ldots,i_n} f_{i_1,\ldots,i_n}\left(\begin{bmatrix}\mathbf{e}_{j_1} & \cdots & \mathbf{e}_{j_n}\end{bmatrix}\right) = c_{j_1,\ldots,j_n}.$$

Therefore the functions $f_{i_1,\ldots,i_n}$ are linearly independent, so they form a basis of V.

6.1.17. (a) Exercise 6.1.16 shows that the set V of multilinear functions $\mathrm{M}_n(\mathbb{F}) \to \mathbb{F}$ is a vector space, so it suffices to show that the set W of alternating multilinear functions is a subspace of V.
Suppose that $f, g \in W$, $a \in \mathbb{F}$, and that $\mathbf{A} \in \mathrm{M}_n(\mathbb{F})$ has two equal columns. Then

$$(f+g)(\mathbf{A}) = f(\mathbf{A}) + g(\mathbf{A}) = 0 \qquad \text{and} \qquad (af)(\mathbf{A}) = a(f(\mathbf{A})) = 0,$$

so $f+g$ and af are alternating. Clearly $0 \in W$. Therefore W is a subspace of V.

(b) Corollary 6.4 shows that each element of W is a scalar multiple of the determinant. Therefore the single function $\det$ forms a basis of W.

6.2 Computing determinants

6.2.1. (a) By Laplace expansion along the first column,

$$\det \begin{bmatrix} 1 & 3 & -1 \\ 0 & 2 & 2 \\ 4 & 1 & 3 \end{bmatrix} = \det \begin{bmatrix} 2 & 2 \\ 1 & 3 \end{bmatrix} + 4 \det \begin{bmatrix} 3 & -1 \\ 2 & 2 \end{bmatrix} = 6 - 2 + 4(6+2) = 36.$$

(b) We follow the idea of Algorithm 6.15 in the next three parts. In this part we can stop after the first couple row operations, observing that two rows are equal so the matrix must be singular.

$$\det \begin{bmatrix} 1 & -1 & 0 & 2 \\ 3 & 3 & -1 & 1 \\ 2 & 4 & -1 & -1 \\ 1 & 1 & 1 & 1 \end{bmatrix} = \det \begin{bmatrix} 1 & -1 & 0 & 2 \\ 0 & 6 & -1 & -5 \\ 0 & 6 & -1 & -5 \\ 1 & 1 & 1 & 1 \end{bmatrix} = 0$$

since the second and third rows are equal, and therefore the rank of the matrix is smaller than 4.

(c)

$$\det \begin{bmatrix} 1 & 0 & -1 & 3 \\ 2 & -3 & -2 & 5 \\ 3 & 0 & -1 & 9 \\ 2 & -3 & -2 & 6 \end{bmatrix} = \det \begin{bmatrix} 1 & 0 & -1 & 3 \\ 0 & -3 & 0 & -1 \\ 0 & 0 & 2 & 0 \\ 0 & -3 & 0 & 0 \end{bmatrix} = \det \begin{bmatrix} 1 & 0 & -1 & 3 \\ 0 & -3 & 0 & -1 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = -6.$$

(d)

$$\det \begin{bmatrix} 1 & 0 & 1 & 0 & -1 \\ 1 & 2 & 1 & -1 & 1 \\ 1 & 2 & 4 & 0 & 0 \\ 1 & 2 & 4 & -1 & 0 \\ 1 & 2 & 4 & -1 & 1 \end{bmatrix} = \det \begin{bmatrix} 1 & 0 & 1 & 0 & -1 \\ 0 & 2 & 0 & -1 & 2 \\ 0 & 2 & 3 & 0 & 1 \\ 0 & 2 & 3 & -1 & 1 \\ 0 & 2 & 3 & -1 & 2 \end{bmatrix} = \det \begin{bmatrix} 1 & 0 & 1 & 0 & -1 \\ 0 & 2 & 0 & -1 & 2 \\ 0 & 0 & 3 & 1 & -1 \\ 0 & 0 & 3 & 0 & -1 \\ 0 & 0 & 3 & 0 & 0 \end{bmatrix}$$

$$= \det \begin{bmatrix} 1 & 0 & 1 & 0 & -1 \\ 0 & 2 & 0 & -1 & 2 \\ 0 & 0 & 3 & 1 & -1 \\ 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 & 1 \end{bmatrix} = \det \begin{bmatrix} 1 & 0 & 1 & 0 & -1 \\ 0 & 2 & 0 & -1 & 2 \\ 0 & 0 & 3 & 1 & -1 \\ 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} = -6$$

by Corollary 6.10.

(e) By Laplace expansion along the first row,

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

(f) We can do the computation just as in part (e), but in $\mathbb{F}_2$ the result turns out to be the same as 0.

6.2.2. (a) By Laplace expansion along the first row,

$$\det \begin{bmatrix} 2 & -7 & 1 \\ 8 & 2 & 7 \\ 1 & 8 & 2 \end{bmatrix} = 2\det \begin{bmatrix} 2 & 8 \\ 8 & 2 \end{bmatrix} + 7\det \begin{bmatrix} 8 & 8 \\ 1 & 2 \end{bmatrix} + \det \begin{bmatrix} 8 & 2 \\ 1 & 8 \end{bmatrix} = 2(-60) + 7(8) + 62 = -2.$$

(b) We follow the idea of Algorithm 6.15 in the next four parts.

$$\det \begin{bmatrix} -1 & 4 & 1 \\ 2 & -1 & -2 \\ 3 & 1 & -1 \end{bmatrix} = \det \begin{bmatrix} -1 & 4 & 1 \\ 0 & 7 & 0 \\ 0 & 13 & 2 \end{bmatrix} = \det \begin{bmatrix} -1 & 4 & 1 \\ 0 & 7 & 0 \\ 0 & 0 & 2 \end{bmatrix} = -14.$$

(c)

$$\det \begin{bmatrix} 1 & -2 & 3 & 0 \\ -2 & 1 & 0 & -1 \\ 3 & -4 & 5 & -2 \\ 0 & -1 & 2 & -3 \end{bmatrix} = \det \begin{bmatrix} 1 & -2 & 3 & 0 \\ 0 & -3 & 6 & -1 \\ 0 & 2 & -4 & -2 \\ 0 & -1 & 2 & -3 \end{bmatrix} = -\det \begin{bmatrix} 1 & -2 & 3 & 0 \\ 0 & -1 & 2 & -3 \\ 0 & 2 & -4 & -2 \\ 0 & -3 & 6 & -1 \end{bmatrix}$$

$$= -\det \begin{bmatrix} 1 & -2 & 3 & 0 \\ 0 & -1 & 2 & -3 \\ 0 & 0 & 0 & -8 \\ 0 & 0 & 0 & 8 \end{bmatrix} = 0.$$

(d)

$$\det \begin{bmatrix} 2 & 0 & -1 & 1 & 3 \\ 0 & -1 & 0 & -1 & 0 \\ 2 & 1 & 4 & 0 & 3 \\ 2 & -1 & -1 & -3 & 3 \\ 0 & 2 & 0 & 2 & 1 \end{bmatrix} = \det \begin{bmatrix} 2 & 0 & -1 & 1 & 3 \\ 0 & -1 & 0 & -1 & 0 \\ 0 & 1 & 5 & -1 & 0 \\ 0 & -1 & 0 & -4 & 0 \\ 0 & 2 & 0 & 2 & 1 \end{bmatrix}$$

$$= \det \begin{bmatrix} 2 & 0 & -1 & 1 & 3 \\ 0 & -1 & 0 & -1 & 0 \\ 0 & 0 & 5 & -2 & 0 \\ 0 & 0 & 0 & -3 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} = 30.$$

(e)

$$\det\begin{bmatrix} 1 & 0 & 0 & 0 & 3 & -2 \\ 0 & 2 & 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 2 & 0 \\ -1 & 0 & 0 & 4 & 0 & 1 \\ 1 & 0 & -2 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 & 2 & 0 \end{bmatrix} = \det\begin{bmatrix} 1 & 0 & 0 & 0 & 3 & -2 \\ 0 & 2 & 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 2 & 0 \\ 0 & 0 & 0 & 4 & 3 & -1 \\ 0 & 0 & -2 & 0 & -3 & 1 \\ 0 & 1 & 0 & -1 & 2 & 0 \end{bmatrix}$$

$$= -\det\begin{bmatrix} 1 & 0 & 0 & 0 & 3 & -2 \\ 0 & 1 & 0 & -1 & 2 & 0 \\ 0 & 0 & 1 & 0 & 2 & 0 \\ 0 & 0 & 0 & 4 & 3 & -1 \\ 0 & 0 & -2 & 0 & -3 & 1 \\ 0 & 2 & 0 & -1 & 0 & 0 \end{bmatrix} = -\det\begin{bmatrix} 1 & 0 & 0 & 0 & 3 & -2 \\ 0 & 1 & 0 & -1 & 2 & 0 \\ 0 & 0 & 1 & 0 & 2 & 0 \\ 0 & 0 & 0 & 4 & 3 & -1 \\ 0 & 0 & -2 & 0 & -3 & 1 \\ 0 & 0 & 0 & 1 & -4 & 0 \end{bmatrix}$$

$$= -\det\begin{bmatrix} 1 & 0 & 0 & 0 & 3 & -2 \\ 0 & 1 & 0 & -1 & 2 & 0 \\ 0 & 0 & 1 & 0 & 2 & 0 \\ 0 & 0 & 0 & 4 & 3 & -1 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & -4 & 0 \end{bmatrix} = \det\begin{bmatrix} 1 & 0 & 0 & 0 & 3 & -2 \\ 0 & 1 & 0 & -1 & 2 & 0 \\ 0 & 0 & 1 & 0 & 2 & 0 \\ 0 & 0 & 0 & 1 & -4 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 4 & 3 & -1 \end{bmatrix}$$

$$= \det\begin{bmatrix} 1 & 0 & 0 & 0 & 3 & -2 \\ 0 & 1 & 0 & -1 & 2 & 0 \\ 0 & 0 & 1 & 0 & 2 & 0 \\ 0 & 0 & 0 & 1 & -4 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 19 & -1 \end{bmatrix} = \det\begin{bmatrix} 1 & 0 & 0 & 0 & 3 & -2 \\ 0 & 1 & 0 & -1 & 2 & 0 \\ 0 & 0 & 1 & 0 & 2 & 0 \\ 0 & 0 & 0 & 1 & -4 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & -20 \end{bmatrix} = -20.$$

6.2.3. (a) If $\mathbf{U} \in \mathrm{M}_n(\mathbb{C})$ is unitary, then

$$1 = \det \mathbf{I}_n = \det(\mathbf{U}^*\mathbf{U}) = (\det \mathbf{U}^*)(\det \mathbf{U}) = \left(\overline{(\det \mathbf{U})}\right)(\det \mathbf{U}) = |\det \mathbf{U}|^2 .$$

(b) Let $\mathbf{A} = \mathbf{U\Sigma V}^*$ be the singular value decomposition of $\mathbf{A}$, so $\Sigma = \mathbf{diag}(\sigma_1, \dots, \sigma_n)$. Then by part (a),

$$|\det \mathbf{A}| = |\det(\mathbf{U\Sigma V}^*)| = |\det \mathbf{U}|\,|\det \mathbf{\Sigma}|\,|\det \mathbf{V}^*| = \sigma_1 \cdots \sigma_n.$$

since $\mathbf{U}$ and $\mathbf{V}^*$ are both unitary.

6.2.4. (a) By Theorem 6.5, $\det \mathbf{A} = (\det \mathbf{L})(\det \mathbf{U})$. Since $\mathbf{L}$ and $\mathbf{U}$ are both triangular, $\det \mathbf{L} = \ell_{11} \cdots \ell_{nn} = 1$ (since the diagonal entries of $\mathbf{L}$ are all 1) and $\det \mathbf{U} = u_{11} \cdots u_{nn}$.

(b) By Theorem 6.5 and part (a),

$$\operatorname{sgn}(\sigma) \det \mathbf{A} = (\det \mathbf{P})(\det \mathbf{A}) = \det(\mathbf{PA}) = \det(\mathbf{LU}) = u_{11} \cdots u_{nn}.$$

Since $\operatorname{sgn}(\sigma) = \pm 1$, this implies that $\det \mathbf{A} = (\operatorname{sgn} \sigma) u_{11} \cdots u_{nn}$.

6.2.5. By Theorem 6.5, $\det(\mathbf{LDU}) = \det(\mathbf{L})\det(\mathbf{D})\det(\mathbf{U})$. Since $\mathbf{D}$ is diagonal, $\det(\mathbf{D}) = d_{11}\cdots d_{nn}$, and since $\mathbf{L}$ and $\mathbf{U}$ are triangular with ones on the diagonals, so $\det(\mathbf{L}) = \det(\mathbf{U}) = 1$.

6.2.6. By Theorem 6.5, $|\det \mathbf{A}| = |\det(\mathbf{QR})| = |(\det \mathbf{Q})(\det \mathbf{R})| = |\det \mathbf{Q}|\,|\det \mathbf{R}|$. By Exercise 6.2.3(a), $|\det \mathbf{A}| = 1$, and by Corollary 6.10, $\det \mathbf{R} = r_{11}\cdots r_{nn}$.

6.2.7. Let $\mathbf{A} = \mathbf{QR}$ be a QR decomposition. Then $|\det \mathbf{A}| = |\det \mathbf{Q}|\,|\det \mathbf{R}|$ by Theorem 6.5. Since $\mathbf{Q}$ is unitary, $|\det \mathbf{Q}| = 1$ by Exercise 6.2.3(a). Since $\mathbf{R}$ is triangular, $\det \mathbf{R} = r_{11}\cdots r_{nn}$. Finally, by Exercise 4.5.21, $|r_{jj}| \le \|\mathbf{a}_j\|$ for each j.

6.2.8. Write

$$\mathbf{A}_n = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & 2 & 2 & \cdots & 2 \\ 1 & 2 & 3 & \cdots & 3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 2 & 3 & \cdots & n \end{bmatrix} \in \mathrm{M}_n(\mathbb{R}).$$

We prove that $\det \mathbf{A}_n = 1$ by induction on n. If $n = 1$, then $\det \mathbf{A}_1 = \det \begin{bmatrix} 1 \end{bmatrix} = 1$.

Now suppose that we know that $\det \mathbf{A}_n = 1$. We need to show that $\det \mathbf{A}_{n+1} = 1$. Begin by subtracting the first row of $\mathbf{A}_{n+1}$ (which consists of all ones) from each of the other rows; we obtain that

$$\det \mathbf{A}_{n+1} = \det \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 0 & 1 & 1 & \cdots & 1 \\ 0 & 1 & 2 & \cdots & 2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 1 & 2 & \cdots & n \end{bmatrix}.$$

Next subtract the first column of the matrix above from each of the other columns; we then obtain that

$$\det \mathbf{A}_{n+1} = \det \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 1 & \cdots & 1 \\ 0 & 1 & 2 & \cdots & 2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 1 & 2 & \cdots & n \end{bmatrix}.$$

Now Laplace expansion along either the first row or first column of the latter matrix shows that $\det \mathbf{A}_{n+1} = \det \mathbf{A}_n = 1$.

6.2.9. By the the Schur decomposition there is a basis $\mathcal{B}$ of V such that $[\boldsymbol{T}]_{\mathcal{B}}$ is upper triangular. Then the diagonal entries of $[\boldsymbol{T}]_{\mathcal{B}}$ are the eigenvalues of $\boldsymbol{T}$, which are real by assumption. Furthermore, since $[\boldsymbol{T}]_{\mathcal{B}}$ is triangular, its determinant is equal to the product of its diagonal entries. It follows that $\det \boldsymbol{T} = \det\,[\boldsymbol{T}]_{\mathcal{B}}$ is real.

6.2.10. Let m be fixed. The result is clearly true for $n = 0$. Now suppose that $n \ge 1$ and we know the result is true for matrices of the given form in $\mathrm{M}_{m+(n-1)}(\mathbb{F})$. Now if $1 \le j \le m$, then

$a_{m+n,j} = 0$, so by Laplace expansion along the last row of $\mathbf{A}$,

$$\det \mathbf{A} = \sum_{j=m+1}^{m+n+j} a_{m+n,j} \det \mathbf{A}_{m+n,j}.$$

Now for $m+1 \le j \le m+n$, $a_{m+n,j} = d_{n,j-m}$ and

$$\mathbf{A}_{m+n,j} = \begin{bmatrix} \mathbf{B} & \mathbf{C}_{j-m} \\ \mathbf{0} & \mathbf{D}_{n,j-m} \end{bmatrix},$$

where $\mathbf{C}_j$ represents the matrix obtained by removing the j^{th} column from $\mathbf{C}$. By the induction hypothesis,

$$\det \mathbf{A}_{m+n,j} = \det \mathbf{B} \det \mathbf{D}_{n,j-m},$$

and so

$$\begin{aligned} \det \mathbf{A} &= \sum_{j=m+1}^{n+m} (-1)^{m+n+j} d_{n,j-m} \det \mathbf{B} \det \mathbf{D}_{n,j-m} \\ &= (\det \mathbf{B}) \sum_{k=1}^{n} (-1)^{n+k} d_{n,k} \det \mathbf{D}_{n,k} \\ &= \det \mathbf{B} \det \mathbf{D}. \end{aligned}$$

Therefore, by induction, the result is true for all n.

6.2.11. Let $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{U}^*$ be a spectral decomposition of $\mathbf{A}$, which exists since $\mathbf{A}$ is normal. Then the diagonal entries of the diagonal factor $\mathbf{D}$ are the eigenvalues of $\mathbf{A}$. Let $\lambda_1, \dots, \lambda_r$ be the distinct eigenvalues of $\mathbf{A}$, and let k_j be the number of times λ_j appears on the diagonal of $\mathbf{D}$. Then $\det \mathbf{A} = \det \mathbf{D} = \lambda_1^{k_1} \cdots \lambda_r^{k_r}$.

6.2.12. (a) Let $\mathcal{B} = (v_1, \dots, v_n)$. Then

$$[\boldsymbol{T}]_{\mathcal{B}} = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 0 & 2 & 2 & \cdots & 2 \\ 0 & 0 & 3 & \cdots & 3 \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \cdots & n \end{bmatrix}$$

is upper triangular with diagonal entries $1, 2, \dots, n$. So by Corollary 6.10, $\det \boldsymbol{T} = n!$.

(b) Since $\det \boldsymbol{T} \neq 0$, $\boldsymbol{T}$ is invertible, hence an isomorphism, and so by Theorem 3.15, $(\boldsymbol{T}v_1, \dots, \boldsymbol{T}v_n)$ is also a basis of V.

6.2.13. **Solution 1:** By Theorem 6.17,

$$f(t) = \sum_{\sigma \in S_n} (\operatorname{sgn} \sigma) \prod_{i=1}^{n} (\delta_{i,\sigma(i)} + t a_{i\sigma(i)}),$$

where $\delta_{i,j} = 1$ if $i = j$ and $\delta_{i,j} = 0$ otherwise. Expanding the product, this implies that

$$f(t) = \sum_{\sigma \in S_n} (\operatorname{sgn} \sigma) \left[\prod_{i=1}^{n} \delta_{i,\sigma_i} + \sum_{i=1}^{n} (t a_{i\sigma(i)}) \left(\prod_{j \neq i} \delta_{j,\sigma_j} \right) + g(t) \right],$$

where $g(t)$ is a polynomial in which all the terms are of degree at least 2. Therefore $g'(0) = 0$, and so

$$f'(0) = \sum_{\sigma \in S_n} (\operatorname{sgn} \sigma) \sum_{i=1}^{n} a_{i\sigma(i)} \left(\prod_{j \neq i} \delta_{j,\sigma_j} \right)$$

The product in the above expression is only nonzero if $\sigma_j = j$ for each $j \neq i$; since σ is a bijection this implies that also $\sigma(i) = i$, and therefore $\sigma = \iota$. The above expression therefore simplifies to

$$f'(0) = \sum_{i=1}^{n} a_{ii} = \operatorname{tr} \mathbf{A}.$$

Solution 2: Let $\mathbf{A} = \mathbf{U}\mathbf{T}\mathbf{U}^*$ be a Schur decomposition of $\mathbf{A}$, and let $\lambda_1, \ldots, \lambda_n$ be the diagonal entries of $\mathbf{T}$. Then

$$f(t) = \det(\mathbf{I}_n + t\mathbf{U}\mathbf{T}\mathbf{U}^*) = \det \mathbf{U}(\mathbf{I}_n + t\mathbf{T})\mathbf{U}^* = \det(\mathbf{I}_n + t\mathbf{T}) = (1 + t\lambda_1) \cdots (1 + t\lambda_n).$$

This implies that

$$f'(0) = \lambda_1 + \cdots + \lambda_n = \operatorname{tr} \mathbf{T} = \operatorname{tr} \mathbf{A}.$$

6.2.14. (a) It is trivial to check that the permanent is multilinear. If $\mathbf{A}$ is the matrix whose entries are all 1, then $\operatorname{per} \mathbf{A} = n!$, despite the fact that the columns are all equal.

Remark: Actually, if $\mathbb{F} = \mathbb{F}_2$ then the permanent is alternating, since it is the same as the determinant.

(b) Almost any two matrices will do. For example, if $\mathbf{A} = \mathbf{B} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$, then $\operatorname{per} \mathbf{A} = \operatorname{per} \mathbf{B} = 2$, but $\mathbf{AB} = 2\mathbf{A}$ so $\operatorname{per}(\mathbf{AB}) = 2^2 \operatorname{per} \mathbf{A} = 8$.

6.2.15. Suppose first that $\mathbf{A} = \mathbf{A}_\sigma$ for some $\sigma \in S_n$. That is,

$$a_{ij} = \begin{cases} 1 & \text{if } \sigma(i) = j, \\ 0 & \text{otherwise.} \end{cases}$$

Since $\sigma : \{1, \ldots, n\} \to \{1, \ldots, n\}$ is a well-defined function, for each i there is exactly one j such that $a_{ij} = 1$ and $a_{ij} = 0$ for all other j. Thus each row of $\mathbf{A}$ has exactly one entry equal to 1 and the others are 0. Since σ is a bijection, for each j there is exactly one i such that $a_{ij} = 1$ and $a_{ij} = 0$ for all other i. Thus each column of $\mathbf{A}$ has exactly one entry equal to 1 and the others are 0.

Conversely, suppose that $\mathbf{A}$ has the stated properties. Define $\sigma : \{1, \ldots, n\} \to \{1, \ldots, n\}$ by setting $\sigma(i) = j$ if $a_{ij} = 1$. By the given properties, for each i there is exactly one such j, so σ is a well-defined function. Since there is exactly one such value of i for each j, σ is actually a bijection. Thus $\sigma \in S_n$, and then $\mathbf{A} = \mathbf{A}_\sigma$ for this permutation σ.

6.2.16. Given $\sigma \in S_n$, the permutation matrix $\mathbf{A}_\sigma$ can be reduced to $\mathbf{I}_n$ via some number of row operations of type **R3**. That means that $\mathbf{A}_\sigma$ can be written as a product of elementary matrices $\mathbf{R}_{i,j}$. Each such matrix is itself a permutation matrix corresponding to a transposition τ. Therefore $\mathbf{A}_\sigma = \mathbf{A}_{\tau_1} \cdots \mathbf{A}_{\tau_k}$ for some collection of transpositions, so $\sigma = \tau_1 \circ \cdots \circ \tau_k$.

6.2.17. If τ is a transposition and $\mathbf{A}_\tau$ is the corresponding permutation matrix, then $\mathbf{A}_\tau$ can be reduced to $\mathbf{I}_n$ by one row operation of type **R3**. Therefore $\det \mathbf{A}_\tau = -1$. Theorem 6.5 and formula (6.7) then imply that

$$\begin{aligned}(-1)^k &= (\det \mathbf{A}_{\tau_1}) \cdots (\det \mathbf{A}_{\tau_k}) = \det(\mathbf{A}_{\tau_1} \cdots \mathbf{A}_{\tau_k}) = \det \mathbf{A}_{\tau_1 \circ \cdots \circ \tau_k} \\ &= \det \mathbf{A}_{\widetilde{\tau}_1 \circ \cdots \circ \widetilde{\tau}_\ell} = \det(\mathbf{A}_{\widetilde{\tau}_1} \cdots \mathbf{A}_{\widetilde{\tau}_\ell}) = (\det \mathbf{A}_{\widetilde{\tau}_1}) \cdots (\det \mathbf{A}_{\widetilde{\tau}_k}) = (-1)^\ell,\end{aligned}$$

which proves the claim.

6.2.18. We prove the claim by induction on n. If $n = 0$ then the claim holds trivially. Suppose now that the claim is known for $n \times n$ matrices. Then subtracting x_0 times column n from column $n+1$, subtracting x_0 times column $n-1$ from column n, and so on shows that

$$\det \mathbf{V} = \det \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & x_1 - x_0 & x_1(x_1 - x_0) & \cdots & x_1^{n-1}(x_1 - x_0) \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_n - x_0 & x_n(x_n - x_0) & \cdots & x_n^{n-1}(x_n - x_0) \end{bmatrix}$$

Next, subtracting the first row of the latter matrix from each of the other rows and factoring out $x_i - x_0$ from row $i+1$ shows that

$$\begin{aligned}\det \mathbf{V} &= \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & x_1 - x_0 & x_1(x_1 - x_0) & \cdots & x_1^{n-1}(x_1 - x_0) \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & x_n - x_0 & x_n(x_n - x_0) & \cdots & x_n^{n-1}(x_n - x_0) \end{bmatrix} \\ &= (x_1 - x_0) \cdots (x_n - x_0) \det \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & x_1 & \cdots & x_1^{n-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 1 & x_n & \cdots & x_n^{n-1} \end{bmatrix}.\end{aligned}$$

Now using Laplace expansion along the first row or column and the induction hypothesis shows that

$$\det \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & x_1 & \cdots & x_1^{n-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 1 & x_n & \cdots & x_n^{n-1} \end{bmatrix} = \det \begin{bmatrix} 1 & x_1 & \cdots & x_1^{n-1} \\ \vdots & \vdots & & \vdots \\ 1 & x_n & \cdots & x_n^{n-1} \end{bmatrix} = \prod_{1 \le i < j \le n} (x_j - x_i).$$

Combining this with the above expression for $\det \mathbf{V}$ proves the claim.

Finally, $\mathbf{V}$ is invertible iff $\det \mathbf{V} \neq 0$, which is the case iff $x_j \neq x_i$ for each $i < j$.

6.2.19. (a) Define $\mathbf{A} \in \mathrm{M}_n(\mathbb{F})$ by $a_{ij} = x_i^{j-1}$. Then Exericse 6.2.18 implies that

$$\det \mathbf{A} = \prod_{1 \le i < j \le n} (x_j - x_i).$$

On the other hand, Theorem 6.17 implies that

$$\det \mathbf{A} = \sum_{\sigma \in S_n} (\operatorname{sgn} \sigma) \prod_{j=1}^n x_i^{(\sigma(j)-1)}.$$

(b) Since $|z_j| = 1$, $\overline{|z_j|} = z_j^{-1}$. Then applying part (a) to both $x_j = z_j$ and $x_j = \overline{z_j}$,

$$\begin{aligned}
\prod 1 \le j < k \le n \, |z_j - z_k|^2 &= \left(\prod 1 \le j < k \le n(z_j - z_k) \right) \left(\prod 1 \le j < k \le n(\overline{z_j} - \overline{z_k}) \right) \\
&= \left(\sum_{\sigma \in S_n} (\operatorname{sgn} \sigma) \prod_{j=1}^n z_j^{\sigma(j)-1} \right) \left(\sum_{\rho \in S_n} (\operatorname{sgn} \rho) \prod_{k=1}^n \overline{z_k}^{\rho(j)-1} \right) \\
&= \sum_{\sigma, \rho \in S_n} (\operatorname{sgn} \sigma)(\operatorname{sgn} \rho) \prod_{j=1}^n z_j^{\sigma(j)-1} \overline{z_j}^{\rho(j)-1} \\
&= \sum_{\sigma, \rho \in S_n} (\operatorname{sgn} \sigma)(\operatorname{sgn} \rho^{-1}) \prod_{j=1}^n z_j^{\sigma(j)-\rho(j)} \\
&= \sum_{\sigma, \rho \in S_n} (\operatorname{sgn} \sigma\rho^{-1}) \prod_{j=1}^n z_j^{\sigma(j)-\rho(j)}.
\end{aligned}$$

6.3 Characteristic polynomials

6.3.1. (a) $p(\lambda) = \det \begin{bmatrix} 1-\lambda & 1 \\ 2 & -\lambda \end{bmatrix} = \lambda^2 - \lambda - 2 = (\lambda - 2)(\lambda + 1)$ (or use Quick Exercise #11). The eigenvalues are 2 and -1, and

$$\operatorname{Eig}_2(\mathbf{A}) = \ker \begin{bmatrix} -1 & 1 \\ 2 & -2 \end{bmatrix} = \left\langle \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\rangle$$

and

$$\operatorname{Eig}_{-1}(\mathbf{A}) = \ker \begin{bmatrix} 2 & 1 \\ 2 & 2 \end{bmatrix} = \left\langle \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right\rangle.$$

(b) $p(\lambda) = \det \begin{bmatrix} \cos(\theta) - \lambda & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) - \lambda \end{bmatrix} = \lambda^2 - 2\cos(\theta)\lambda + 1$ (or use Quick Exercise #11). The eigenvalues of $\mathbf{A}$ are the roots of this polynomial, which we can find to be $\cos(\theta) \pm i \sin(\theta)$ via the quadratic formula. Furthermore,

$$\operatorname{Eig}_{\cos(\theta)+i\sin(\theta)}(\mathbf{A}) = \ker \begin{bmatrix} -i\sin(\theta) & -\sin(\theta) \\ \sin(\theta) & -i\sin(\theta) \end{bmatrix} = \left\langle \begin{bmatrix} 1 \\ -i \end{bmatrix} \right\rangle$$

and

$$\operatorname{Eig}_{\cos(\theta)-i\sin(\theta)}(\mathbf{A}) = \ker \begin{bmatrix} i\sin(\theta) & -\sin(\theta) \\ \sin(\theta) & i\sin(\theta) \end{bmatrix} = \left\langle \begin{bmatrix} 1 \\ i \end{bmatrix} \right\rangle.$$

(c) $p(\lambda) = \det \begin{bmatrix} -1-\lambda & -3 & 1 \\ 3 & 3-\lambda & 1 \\ 3 & 0 & 4-\lambda \end{bmatrix} = -\lambda^3+6\lambda^2-11\lambda+6 = (1-\lambda)(2-\lambda)(3-\lambda)$, so the eigenvalues are 1, 2, and 3, and

$$\mathrm{Eig}_1(\mathbf{A}) = \ker \begin{bmatrix} -2 & -3 & 1 \\ 3 & 2 & 1 \\ 3 & 0 & 3 \end{bmatrix} = \left\langle \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix} \right\rangle,$$

$$\mathrm{Eig}_2(\mathbf{A}) = \ker \begin{bmatrix} -3 & -3 & 1 \\ 3 & 1 & 1 \\ 3 & 0 & 2 \end{bmatrix} = \left\langle \begin{bmatrix} -2 \\ 3 \\ 3 \end{bmatrix} \right\rangle,$$

and

$$\mathrm{Eig}_3(\mathbf{A}) = \ker \begin{bmatrix} -4 & -3 & 1 \\ 3 & 0 & 1 \\ 3 & 0 & 1 \end{bmatrix} = \left\langle \begin{bmatrix} -3 \\ 7 \\ 9 \end{bmatrix} \right\rangle.$$

(d) $p(\lambda) = \det \begin{bmatrix} 1-\lambda & 0 & 3 \\ 0 & -2-\lambda & 0 \\ 3 & 0 & 1-\lambda \end{bmatrix} - \lambda^3 + 12\lambda + 16 = (4-\lambda)(2-\lambda)^2$, so the eigenvalues are 4 and 2 and

$$\mathrm{Eig}_4(\mathbf{A}) = \ker \begin{bmatrix} -3 & 0 & 3 \\ 0 & -6 & 0 \\ 3 & 0 & -3 \end{bmatrix} = \left\langle \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \right\rangle$$

and

$$\mathrm{Eig}_2(\mathbf{A}) = \ker \begin{bmatrix} -1 & 0 & 3 \\ 0 & -4 & 0 \\ 3 & 0 & -1 \end{bmatrix} = \left\langle \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\rangle.$$

6.3.2. (a) $p(\lambda) = \det \begin{bmatrix} 1-\lambda & 2 \\ 3 & 2-\lambda \end{bmatrix} = \lambda^2 - 3\lambda - 4 = (\lambda+1)(\lambda-4)$ (or use Quick Exercise #11). The eigenvalues are 4 and -1, and

$$\mathrm{Eig}_4(\mathbf{A}) = \ker \begin{bmatrix} -3 & 2 \\ 3 & -2 \end{bmatrix} = \left\langle \begin{bmatrix} 2 \\ 3 \end{bmatrix} \right\rangle$$

and

$$\mathrm{Eig}_{-1}(\mathbf{A}) = \ker \begin{bmatrix} 2 & 2 \\ 3 & 3 \end{bmatrix} = \left\langle \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right\rangle.$$

(b) $p(\lambda) = \det \begin{bmatrix} 2-\lambda & 2i \\ i & -1-\lambda \end{bmatrix} = \lambda^2 - \lambda = \lambda(\lambda-1)$ (or use Quick Exercise #11). The eigenvalues are 0 and 1, and

$$\mathrm{Eig}_0(\mathbf{A}) = \ker \begin{bmatrix} 2 & 2i \\ i & -1 \end{bmatrix} = \left\langle \begin{bmatrix} 1 \\ i \end{bmatrix} \right\rangle$$

and

$$\mathrm{Eig}_1(\mathbf{A}) = \ker \begin{bmatrix} 1 & 2i \\ i & -2 \end{bmatrix} = \left\langle \begin{bmatrix} 2 \\ i \end{bmatrix} \right\rangle.$$

(c) $p(\lambda) = \det \begin{bmatrix} 2-\lambda & 1 & 1 \\ -4 & -3-\lambda & 0 \\ -2 & -2 & 1-\lambda \end{bmatrix} = \lambda - \lambda^3 = \lambda(1-\lambda)(1+\lambda)$, so the eigenvalues are 0, 1, and -1, and

$$\mathrm{Eig}_0(\mathbf{A}) = \ker \begin{bmatrix} 2 & 1 & 1 \\ -4 & -3 & 0 \\ -2 & -2 & 1 \end{bmatrix} = \left\langle \begin{bmatrix} 3 \\ -4 \\ -2 \end{bmatrix} \right\rangle,$$

$$\mathrm{Eig}_1(\mathbf{A}) = \ker \begin{bmatrix} 1 & 1 & 1 \\ -4 & -4 & 0 \\ -2 & -2 & 0 \end{bmatrix} = \left\langle \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \right\rangle,$$

and

$$\mathrm{Eig}_{-1}(\mathbf{A}) = \ker \begin{bmatrix} 3 & 1 & 1 \\ -4 & -2 & 0 \\ -2 & -2 & 2 \end{bmatrix} = \left\langle \begin{bmatrix} 1 \\ -2 \\ -1 \end{bmatrix} \right\rangle.$$

(d) $p(\lambda) = \det \begin{bmatrix} -\lambda & 2 & 1 \\ -1 & -2-\lambda & 0 \\ 1 & 1 & -1-\lambda \end{bmatrix} = -\lambda^3 - 3\lambda^2 - 3\lambda - 1 = -(1+\lambda)^3$, so the only eigenvalue is -1 and

$$\mathrm{Eig}_{-1}(\mathbf{A}) = \ker \begin{bmatrix} 1 & 2 & 1 \\ -1 & -1 & 0 \\ 1 & 1 & 0 \end{bmatrix} = \left\langle \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \right\rangle.$$

6.3.3. (a) Both matrices have trace 2, determinant 1, and characteristic polynomial $x^2 - 2x + 1$ (which you can see by direct computation, or using Quick Exercise #11).

(b) For any invertible $\mathbf{S} \in \mathrm{M}_2(\mathbb{F})$, $\mathbf{S}\mathbf{I}_2\mathbf{S}^{-1} = \mathbf{S}\mathbf{S}^{-1} = \mathbf{I}_2$. Therefore the only matrix similar to $\mathbf{I}_2$ is $\mathbf{I}_2$. Since $\mathbf{A} \neq \mathbf{I}_2$, $\mathbf{A}$ is not similar to $\mathbf{I}_2$.

Remark: Notice that we didn't even have to say what field we're working over in this problem. (Although if $\mathbb{F} = \mathbb{F}_2$, the trace turns out to be 0.)

6.3.4. (a) The first matrix has distinct eigenvalues 1 and 2 (which we can read off since the matrix is triangular), so it is diagonalizable and similar to $\mathbf{diag}(1, 2)$. The second matrix has characteristic polynomial $\lambda^2 - 3\lambda + 2 = (\lambda - 1)(\lambda - 2)$; therefore it also has distinct eigenvalues 1 and 2 and is similar to $\mathbf{diag}(1, 2)$.

(b) The first matrix has eigenvalues 1.000000001 and 2 as eigenvalues, whereas the second matrix has 1 and 2 as eigenvalues.

6.3.5. Let λ_1 and λ_2 be the other eigenvalues. By Corollary 6.24, we know that $2 + \lambda_1 + \lambda_2 = -1$ and $2\lambda_1\lambda_2 = 6$. That is, $\lambda_1 + \lambda_2 = -3$ and $\lambda_1\lambda_2 = 3$. This implies that λ_1 and λ_2 are roots of the quadratic equation $x^2 + 3x + 3 = 0$, which has solutions $-\frac{3}{2} \pm \frac{\sqrt{3}}{2}i$.

6.3.6. Let λ_1 and λ_2 be the other eigenvalues. By Corollary 6.24, we know that $2 + \lambda_1 + \lambda_2 = -1$ and $2\lambda_1\lambda_2 = -6$. That is, $\lambda_1 + \lambda_2 = -3$ and $\lambda_1\lambda_2 = -3$. This implies that λ_1 and λ_2 are roots of the quadratic equation $x^2 + 3x - 3 = 0$, which has solutions $-\frac{3}{2} \pm \frac{\sqrt{21}}{2}$.

6.3.7. (a) $p(\lambda) = \lambda^2 - 5\lambda - 2$ by Quick Exercise #11.

(b) By the Cayley–Hamilton Theorem, $\mathbf{A}^2 - 5\mathbf{A} - 2\mathbf{I}_2 = \mathbf{0}$. This implies that $\mathbf{A}^2 = 5\mathbf{A} + 2\mathbf{I}_2$, so

$$\mathbf{A}^3 = 5\mathbf{A}^2 + 2\mathbf{A} = 5\,(5\mathbf{A} + 2\mathbf{I}_2) + 2\mathbf{A} = 27\mathbf{A} + 10\mathbf{I}_2 = \begin{bmatrix} 37 & 54 \\ 81 & 118 \end{bmatrix}.$$

6.3.8. (a) $p(\lambda) = \lambda^2 - 10\lambda + 9$ by Quick Exercise #11.

(b) By the Cayley–Hamilton Theorem, $\mathbf{A}^2 - 10\mathbf{A} + 9\mathbf{I}_2 = \mathbf{0}$. This implies that $\mathbf{A}^2 = 10\mathbf{A} - 9\mathbf{I}_2$, so

$$\mathbf{A}^3 = 10\mathbf{A}^2 - 9\mathbf{A} = 10\,(10\mathbf{A} - 9\mathbf{I}_2) - 9\mathbf{A} = 91\mathbf{A} - 90\mathbf{I}_2$$

and then

$$\begin{aligned} \mathbf{A}^4 &= 91\mathbf{A}^2 - 90\mathbf{A} = 91\,(10\mathbf{A} - 9\mathbf{I}_2) - 90\mathbf{A} \\ &= 820\mathbf{A} - 819\mathbf{I}_2 = \begin{bmatrix} 821 & 5740 \\ 820 & 5741 \end{bmatrix}. \end{aligned}$$

6.3.9. By Proposition 6.18, $p_{\mathbf{A}}(0) = 0$ if and only if 0 is a root of $p_{\mathbf{A}}$. an eigenvalue, which is true if and only if $\mathbf{A}$ has a non-trivial nullspace. By the Rank–Nullity Theorem, this is true if and only if $\operatorname{rank} \mathbf{A} < n$.

6.3.10. First observe that if $\mathbf{B}$ is upper triangular, then the list of its columns containing a nonzero diagonal entry is linearly independent by the Linear Dependence Lemma, since none of them is in the span of the previous columns. Thus the rank of an upper triangular matrix is greater than or equal to the number of nonzero diagonal entries. By the Rank–Nullity Theorem, this means that the nullity of an upper triangular matrix is less than or equal to the number of 0s on its diagonal. Hence

$$\dim \ker(\lambda \mathbf{I}_n - \mathbf{B})$$

is less than or equal to the number of times λ appears on the diagonal of $\mathbf{B}$, which is the algebraic multiplicity of λ as an eigenvalue of $\mathbf{B}$.

Now in general, given $\mathbf{A}$ there is an invertible matrix $\mathbf{S}$ (over a larger field, if necessary) such that $\mathbf{B} = \mathbf{SAS}^{-1}$ is upper triangular. Then $\mathbf{B}$ has the same characteristic polynomial as $\mathbf{A}$, hence the same eigenvalues with the same algebraic multiplicities. Furthermore,

$$\lambda \mathbf{I}_n - \mathbf{B} = \mathbf{S}(\lambda \mathbf{I}_n - \mathbf{A})\mathbf{S}^{-1},$$

so

$$\dim \ker(\lambda \mathbf{I}_n - \mathbf{B}) = \dim \ker(\lambda \mathbf{I}_n - \mathbf{A})$$

and thus the geometric multiplicities are the same as well. Thus the result in general follows from the upper triangular case above.

6.3.11. Let $\mathbf{A} = \mathbf{UDU}^*$ be a spectral decomposition of $\mathbf{A}$, and let $\mathbf{D} = \mathbf{diag}(\lambda_1, \ldots, \lambda_n)$. By Lemma 6.23, the multiplicity of an eigenvalue λ of $\mathbf{D}$ (hence of $\mathbf{A}$) is equal to the number of times λ appears among the λ_j. On the other hand, $\operatorname{Eig}_\lambda(\mathbf{A}) = \operatorname{Eig}_\lambda(\mathbf{D}) = \ker(\mathbf{D} - \lambda \mathbf{I}_n)$ is spanned by those $\mathbf{e}_j$ for which $\lambda_j = \lambda$. Thus the geometric multiplicity of λ is equal to the number of such j.

6.3.12. By Proposition 6.22, if $\lambda_1, \dots, \lambda_m$ are the distinct eigenvalues of $\mathbf{A}$ with multiplicities $k_1, \dots, k_m$, then $p_{\mathbf{A}}(x) = \prod_{j=1}^m (\lambda_j - x)^{k_j}$. Since the λ_j are all real, expanding this product results in a polynomial whose coefficients are all real.

6.3.13. Let $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^*$ be an SVD for $\mathbf{A}$. Then $\mathbf{A}\mathbf{A}^* = \mathbf{U}\mathbf{\Sigma}\mathbf{\Sigma}^*\mathbf{U}^*$ is similar to $\mathbf{\Sigma}\mathbf{\Sigma}^* = \mathbf{diag}(\sigma_1^2, \dots, \sigma_m^2)$. By Lemma 6.23, $\sigma_1^2, \dots, \sigma_m^2$ are the eigenvalues of $\mathbf{\Sigma}\mathbf{\Sigma}^*$, and hence of $\mathbf{A}\mathbf{A}^*$, listed with multiplicity.

6.3.14. Since $\mathbb{F}$ is algebraically closed, $\mathbf{A} = \mathbf{S}\mathbf{T}\mathbf{S}^{-1}$ for some invertible $\mathbf{S} \in \mathrm{M}_n(\mathbb{F})$ and some upper triangular $\mathbf{T} \in \mathrm{M}_n(\mathbb{F})$. Then $\mathbf{A}^{-1} = \mathbf{S}\mathbf{T}^{-1}\mathbf{S}^{-1}$, and $\mathbf{T}^{-1}$ is upper triangular, with diagonal entries equal to the reciprocols of the diagonal entries of $\mathbf{T}$ (see Exercises 2.4.19 and 2.3.12). By Lemma 6.23, this implies that the eigenvalues of $\mathbf{A}^{-1}$ are, when listed with multiplicity, the same as the reciprocols of the eigenvalues of $\mathbf{A}$ listed with multiplicity.

6.3.15. Let

$$p_{\mathbf{A}}(x) = \det(\mathbf{A} - x\mathbf{I}_n) = \sum_{k=0}^{n} a_k x^k$$

be the characteristic polynomial of $\mathbf{A}$. Recall that $a_n = (-1)^n$. By the Cayley–Hamilton theorem, $p_{\mathbf{A}}(\mathbf{A}) = \mathbf{0}$, and so

$$\mathbf{A}^n = (-1)^{n+1} \sum_{k=0}^{n-1} a_k \mathbf{A}^k \in \langle \mathbf{I}_n, \mathbf{A}, \dots, \mathbf{A}^{n-1} \rangle.$$

Using this, we can prove by induction that $\mathbf{A}^m$ is in $\langle \mathbf{I}_n, \mathbf{A}, \dots, \mathbf{A}^{n-1} \rangle$ for integer $m \geq 0$. This is obvious for $m < n$, and the case $m = n$ follows from the above observation. Now suppose that $m \geq n$ and we know

$$\mathbf{A}^m = \sum_{k=0}^{n-1} c_k \mathbf{A}^k$$

for some $c_0, \dots, c_{n-1} \in \mathbb{F}$. Then

$$\mathbf{A}^{m+1} = \sum_{k=0}^{n-1} c_k \mathbf{A}^{k+1} = \sum_{k=0}^{n-2} c_k \mathbf{A}^{k+1} + \mathbf{A}^n$$

which is in $\langle \mathbf{I}_n, \mathbf{A}, \dots, \mathbf{A}^{n-1} \rangle$ since the sum is by definition, and $\mathbf{A}^n$ is by the above observation.

Finally, given any polynomial $p \in \mathcal{P}(\mathbb{F})$, $p(\mathbf{A})$ is a linear combination of matrices of the form $\mathbf{A}^m$, and is therefore in $\langle \mathbf{I}_n, \mathbf{A}, \dots, \mathbf{A}^{n-1} \rangle$ since each $\mathbf{A}^m$ is.

6.3.16. (a) Let

$$p_{\mathbf{A}}(x) = \det(\mathbf{A} - x\mathbf{I}_n) = \sum_{k=0}^{n} a_k x^k$$

be the characteristic polynomial of $\mathbf{A}$. Since $\mathbf{A}$ is invertible, $a_0 = p_{\mathbf{A}}(0) = \det \mathbf{A} \neq 0$. By the Cayley–Hamilton theorem, $p_{\mathbf{A}}(\mathbf{A}) = \mathbf{0}$. This implies that

$$a_0 \mathbf{I}_n = -\sum_{k=1}^{n} a_k \mathbf{A}^k$$

and so

$$\mathbf{A}^{-1} = \mathbf{A}^{-1}\left(-a_0^{-1}\sum_{k=1}^{n} a_k \mathbf{A}^k\right) = -a_0^{-1}\sum_{k=1}^{n} a_k \mathbf{A}^{k-1}.$$

(b) By Exercise 2.3.12, $\mathbf{A}^k$ is upper triangular for each $k \geq 0$, and hence so is any linear combination of them. So by part (a), $\mathbf{A}^{-1}$ is upper triangular.

6.3.17. Let $\mathbf{A} = \mathbf{U}\mathbf{T}\mathbf{U}^*$ be a Schur decomposition. Then the diagonal entries of $\mathbf{T}$ are $\lambda_1, \ldots, \lambda_n$ in some order by Lemma 6.23, and so

$$\|\mathbf{A}\|_F^2 = \|\mathbf{T}\|_F^2 = \sum_{j,k=1}^{n} |t_{jk}|^2 \geq \sum_{j=1}^{n} |t_{jj}|^2 = \sum_{j=1}^{n} |\lambda_j|^2 .$$

6.3.18. We start by performing row operations on

$$\mathbf{A} - x\mathbf{I}_n = \begin{bmatrix} -x & 0 & 0 & \cdots & 0 & -c_0 \\ 1 & -x & 0 & \cdots & 0 & -c_1 \\ 0 & 1 & -x & \cdots & 0 & -c_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -x & -c_{n-2} \\ 0 & 0 & 0 & \cdots & 1 & -c_{n-1} - x \end{bmatrix}.$$

We add x times the last row to the second-last row, then add x times the second-last row to the third-last row, and so on. We obtain that

$$p_{\mathbf{A}}(x) = \det(\mathbf{A} - x\mathbf{I}_n) = \det \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & -c_0 - c_1 x - \cdots - x^n \\ 1 & 0 & 0 & \cdots & 0 & -c_1 - c_2 x - \cdots - x^{n-1} \\ 0 & 1 & 0 & \cdots & 0 & -c_2 - c_3 x - \cdots - x^{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & -c_{n-2} - c_{n-1}x - x^2 \\ 0 & 0 & 0 & \cdots & 1 & -c_{n-1} - x \end{bmatrix}.$$

Note that the top-right entry of the matrix above is $-p(x)$. Next swap the first two rows of the matrix, then swap the second and third row, and so on until the original first row reaches the bottom. This takes a total of $n-1$ swaps, so

$$p_{\mathbf{A}}(x) = (-1)^{n-1} = \det \begin{bmatrix} 1 & 0 & \cdots & 0 & -c_1 - c_2 x - \cdots - x^{n-1} \\ 0 & 1 & \cdots & 0 & -c_2 - c_3 x - \cdots - x^{n-2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -c_{n-1} - x \\ 0 & 0 & \cdots & 0 & -p(x) \end{bmatrix}.$$

Since this last matrix is upper triangular, its determinant is the product of the diagonal entries, and so finally

$$p_{\mathbf{A}}(x) = (-1)^{n-1}(-p(x)) = (-1)^n p(x).$$

6.3.19. By definition

$$t_{\mathbf{A}}(x) = \operatorname{tr}(\mathbf{A} - x\mathbf{I}_n) = \operatorname{tr}\mathbf{A} - x\operatorname{tr}\mathbf{I}_n = \operatorname{tr}\mathbf{A} - nx.$$

Therefore $t_{\mathbf{A}}(\mathbf{A}) = (\operatorname{tr}\mathbf{A})\mathbf{I}_n - n\mathbf{A}$. This is $\mathbf{0}$ iff $\mathbf{A} = \frac{\operatorname{tr}\mathbf{A}}{n}\mathbf{I}_n$.

6.4 Applications of determinants

6.4.1. Let $\mathcal{C}$ be the unit circle. Then $\begin{bmatrix} x \\ y \end{bmatrix} \in \mathcal{C}$ if and only if $\begin{bmatrix} ax \\ by \end{bmatrix} \in \mathcal{E}$. Therefore $\mathcal{E} = \boldsymbol{T}(\mathcal{C})$, where $\boldsymbol{T} \in \mathcal{L}(\mathbb{R}^2)$ has matrix $\mathbf{A} = \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}$. So the area of $\mathcal{E}$ is the area of $\mathcal{C}$ times $|\det\mathbf{A}| = ab$, or $ab\pi$.

6.4.2. Let $\mathcal{B}$ be the unit circle. Then $\begin{bmatrix} x \\ y \\ z \end{bmatrix} \in \mathcal{B}$ if and only if $\begin{bmatrix} ax \\ by \\ cz \end{bmatrix} \in \mathcal{E}$. Therefore $\mathcal{E} = \boldsymbol{T}(\mathcal{B})$, where $\boldsymbol{T} \in \mathcal{L}(\mathbb{R}^2)$ has matrix $\mathbf{A} = \mathbf{diag}(a, b, c)$. So the volume of $\mathcal{E}$ is the volume of $\mathcal{B}$ times $|\det\mathbf{A}| = abc$, or $\frac{4}{3}abc\pi$.

6.4.3. Let $\mathbf{A} = \begin{bmatrix} 1/2 & 1/2 \\ 1 & -1 \end{bmatrix}$. Then $\mathbf{x} \in \mathcal{E}$ if and only if $\|\mathbf{A}\mathbf{x}\|^2 \le 1$; that is, if $\mathbf{A}\mathbf{x}$ lies in the unit circle $\mathcal{C}$. Therefore $\mathcal{E} = \boldsymbol{T}(\mathcal{C})$, where $\boldsymbol{T} \in \mathcal{L}(\mathbb{R}^2)$ has matrix $\mathbf{A}^{-1} = \begin{bmatrix} 1 & 1/2 \\ 1 & -1/2 \end{bmatrix}$, and so the area of $\mathcal{E}$ is $\left|\det\mathbf{A}^{-1}\right|\pi = \pi$.

6.4.4.
(a) $\det\mathbf{A} = -8$, $\det\mathbf{A}_1 = -10 + 3\sqrt{2}$, $\det\mathbf{A}_2 = 1 + 5\sqrt{2}$. Therefore $(x, y) = \left(\frac{10-3\sqrt{2}}{8}, \frac{-1-5\sqrt{2}}{8}\right)$.

(b) $\det\mathbf{A} = -3 + 2i$, $\det\mathbf{A}_1 = -7 + i$, $\det\mathbf{A}_2 = -3 + 4i$. Therefore $(x, y) = \left(\frac{-7+i}{-3+2i}, \frac{-3+4i}{-3+2i}\right) = \left(\frac{23}{13} + \frac{11}{13}i, \frac{17}{13} - \frac{6}{13}i\right)$.

(c) $\det\mathbf{A} = 2$, $\det\mathbf{A}_1 = -2$, $\det\mathbf{A}_2 = 0$, $\det\mathbf{A}_3 = 4$. Therefore $(x, y, z) = (-1, 0, 2)$.

(d) $\det\mathbf{A} = 1$, $\det\mathbf{A}_1 = 1$, $\det\mathbf{A}_2 = 1$, $\det\mathbf{A}_3 = 0$. Therefore $(x, y, z) = (1, 1, 0)$.

6.4.5.
(a) $\det\mathbf{A} = 3$, $\det\mathbf{A}_1 = 11$, $\det\mathbf{A}_2 = -1$. Therefore $(x, y) = \left(\frac{11}{3}, -\frac{1}{3}\right)$

(b) $\det\mathbf{A} = 2 + 2i$, $\det\mathbf{A}_1 = -i$, $\det\mathbf{A}_2 = 1 + 4i$. Therefore $(x, y) = \left(\frac{-i}{2+2i}, \frac{1+4i}{2+2i}\right) = \left(-\frac{1}{4} - \frac{1}{4}i, \frac{5}{4} + \frac{3}{4}i\right)$.

(c) $\det\mathbf{A} = -4$, $\det\mathbf{A}_1 = 12$, $\det\mathbf{A}_2 = -20$, $\det\mathbf{A}_3 = -40$. Therefore $(x, y, z) = (-3, 5, 10)$.

(d) $\det\mathbf{A} = 1$, $\det\mathbf{A}_1 = 2$, $\det\mathbf{A}_2 = 3$, $\det\mathbf{A}_3 = 2$, $\det\mathbf{A}_4 = 2$. Therefore $(x, y, z, w) = (2, 3, 2, 2)$.

6.4.6. (a) $\begin{bmatrix} 0 & 1 & -1 \\ 1 & -1 & 1 \\ -1 & 1 & 0 \end{bmatrix}$

(b) $\begin{bmatrix} -1 & -2 & 4 & -2 \\ -1 & -1 & 2 & -1 \\ \frac{1}{2} & \frac{1}{2} & -1 & 1 \\ -\frac{1}{2} & -\frac{1}{2} & 2 & -1 \end{bmatrix}$

(c) $\begin{bmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 1 \end{bmatrix}$

(d) $\begin{bmatrix} 1 & 1 & -1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & -1 & 1 & 1 \end{bmatrix}$

6.4.7. (a) $\begin{bmatrix} \frac{2}{3} & 0 & -\frac{1}{3} \\ -\frac{11}{3} & 2 & \frac{4}{3} \\ -2 & 1 & 1 \end{bmatrix}$

(b) $\begin{bmatrix} 1 & 0 & -1 & 1 \\ 0 & 0 & 1 & -1 \\ -1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 1 \end{bmatrix}$

(c) $\begin{bmatrix} -\frac{3}{2} & \frac{5}{2} & 1 & -\frac{1}{2} \\ \frac{9}{2} & -\frac{11}{2} & -2 & \frac{1}{2} \\ \frac{5}{2} & -\frac{7}{2} & -1 & \frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix}$

(d) $\begin{bmatrix} -\frac{1}{2} & -\frac{1}{2} & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ -\frac{1}{4} & -\frac{1}{4} & \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 & -1 & -1 \\ 0 & -1 & 0 & 3 & 3 \end{bmatrix}$

6.4.8. Since $\boldsymbol{T}$ is an isometry, its matrix (with respect to the standard basis) is orthogonal, so by Exercise 6.2.3(a), $|\det \boldsymbol{T}| = 1$. So by Theorem 6.27, $\mathrm{vol}(\boldsymbol{T}(\Omega)) = \mathrm{vol}(\Omega)$.

6.4.9. If the matrix of $\boldsymbol{T}$ is $\Sigma = \mathbf{diag}(\sigma_1, \dots, \sigma_n)$, then for a cube C with each face in some $\langle \mathbf{e}_j \rangle^\perp$, it is clear that $\mathrm{vol}(\boldsymbol{T}(C)) = \sigma_1 \cdots \sigma_n \,\mathrm{vol}(C)$; this is then true for Jordan measurable sets by approximation. If $\boldsymbol{T}$ is an isometry, then it is reasonable to expect that $\mathrm{vol}(\boldsymbol{T}(\Omega)) = \mathrm{vol}(\Omega)$ (this is the most heuristic part of this argument; a careful proof of this statement is no easier than the proof of Theorem 6.27). By the Singular Value Decomposition, it then follows that $\mathrm{vol}(\boldsymbol{T}(\Omega)) = \sigma_1 \cdots \sigma_n \,\mathrm{vol}(\Omega)$, where $\sigma_1, \dots, \sigma_n$ are the singular values of $\boldsymbol{T}$. By Exercise 6.2.3(b), $|\det \boldsymbol{T}| = \sigma_1 \cdots \sigma_n$.

6.4.10. By Exercise 6.2.3(b), $|\det \boldsymbol{T}| = \sigma_1 \cdots \sigma_n$, and therefore $\sigma_n^n \leq |\det \boldsymbol{T}| \leq \sigma_1^n$. The claim now follows from Theorem 6.27.

6.4.11. Let $\mathcal{C}$ be the parallelepiped (the cube) spanned by $\mathbf{e}_1, \ldots, \mathbf{e}_n$, and let $\boldsymbol{T} \in \mathcal{L}(\mathbb{R}^n)$ be given by the matrix $\mathbf{A} = \begin{bmatrix} | & & | \\ \mathbf{a}_1 & \cdots & \mathbf{a}_n \\ | & & | \end{bmatrix}$. The n-dimensional parallelepiped $\mathcal{P}$ spanned by $\mathbf{a}_1, \ldots, \mathbf{a}_n$ is exactly $\boldsymbol{T}(\mathcal{C})$, so by Theorem 6.27, $\operatorname{vol}(\mathcal{P}) = |\det \boldsymbol{T}| \operatorname{vol}(\mathcal{C}) = |\det \mathbf{A}|$.

6.4.12. By Theorem 6.29, we need to show that if $\mathbf{A}$ is invertible and upper triangular, then its cofactor matrix $\mathbf{C}$ is lower triangular. That is, if $i < j$, then $c_{ij} = 0$, or equivalently $\det \mathbf{A}_{ij} = 0$.

Since $\mathbf{A}$ is upper triangular, if $i < j$ then $\mathbf{A}_{ij}$ is also upper triangular, with ith diagonal entry equal to 0. Therefore $\det \mathbf{A}_{ij} = 0$, which proves the claim.

Furthermore, $\mathbf{A}_{ii}$ is upper triangular with diagonal entries $a_{11}, \ldots, a_{i-1,i-1}, a_{i+1,i+1}, \ldots, a_{nn}$, so $\det \mathbf{A}_{ii} = \frac{a_{11} \cdots a_{nn}}{a_{ii}} = \frac{\det \mathbf{A}}{a_{ii}}$. It follows from Theorem 6.29 that the (i,i) entry of $\mathbf{A}^{-1}$ is a_{ii}^{-1}.

6.4.13. If $\mathbf{A}$ and $\mathbf{b}$ have only integer entries, then in Cramer's rule $\mathbf{A}_i$ has only integer entries for each i, and so $\det \mathbf{A}_i$ is an integer (this follows from either Theorem 6.17 or Laplace expansion). Since $\det \mathbf{A} = \pm 1$, Cramer's rule then implies that each x_i is an integer as well.

6.4.14. If $\mathbf{A}$ and $\mathbf{A}^{-1}$ have only integer entries, theny by either Theorem 6.17 or Laplace expansion, $\det \mathbf{A}$ and $\det \mathbf{A}^{-1}$ are integers. Since $(\det \mathbf{A})(\det \mathbf{A}^{-1}) = \det \mathbf{I}_n = 1$, this implies that $\det \mathbf{A} = \pm 1$.

Now suppose that $\det \mathbf{A} = \pm 1$. For each i, $\mathbf{x} = \mathbf{A}^{-1}\mathbf{e}_i$ is the unique solution to the linear system $\mathbf{A}\mathbf{x} = \mathbf{e}_i$. By Exercise 6.4.13, $\mathbf{A}^{-1}\mathbf{e}_i$ has only integer entries. Thus each column of $\mathbf{A}^{-1}$ has only integer entries, and so $\mathbf{A}^{-1}$ has only integer entries.

6.4.15. Let ℓ denote the length of the line segment L. Subdivide L into n pieces of length $\frac{\ell}{n}$; cover each piece with a rectangle whose opposite corners lie on the line. The area of any such rectangle is bounded by $\left(\frac{\ell}{n}\right)^2$ and there are n rectangles, so the total area is bounded by $\frac{\ell^2}{n}$.