Solutions to Selected Problems In: Fundamentals of Matrix Computations: Second Edition by David S. Watkins.

John L. Weatherwax*

March 13, 2012

*wax@alum.mit.edu

Chapter 1 (Gaussian Elimination and Its Variants)

Exercise 1.1.20 (an example of block partitioning)

We first compute Ax with (ignoring the partitioning for the time being) is given by

$$Ax = \begin{bmatrix} 1 & 3 & 2 \\ 2 & 1 & 1 \\ -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ 2 & 1 & 1 \\ -1 & 2 & 8 \end{bmatrix} = \begin{bmatrix} 4 & 7 & 4 \\ 3 & 3 & 3 \\ -2 & 2 & -1 \end{bmatrix}$$

Now to show that $A_{i1}X_{1j} + A_{i2}X_{2j} = B_{ij}$ for $i, j = 1, 2$ consider each of the possible four pairs in tern. First consider $(i, j) = (1, 1)$ which is

$$A_{11}X_{11} + A_{12}X_{21} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} [1] + \begin{bmatrix} 3 & 2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} + \begin{bmatrix} 4 \\ 1 \end{bmatrix} = \begin{bmatrix} 5 \\ 3 \end{bmatrix}$$

Which is the same as B_{11}. Now $(i, j) = (1, 2)$ gives

$$A_{11}X_{12} + A_{12}X_{22} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \begin{bmatrix} 0 & 1 \end{bmatrix} + \begin{bmatrix} 3 & 2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 2 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 2 \end{bmatrix} + \begin{bmatrix} 7 & 3 \\ 3 & 1 \end{bmatrix} = \begin{bmatrix} 7 & 4 \\ 3 & 3 \end{bmatrix}$$

Which is the same as B_{12}. Now $(i, j) = (2, 1)$ gives

$$A_{21}X_{11} + A_{22}X_{21} = [-1] [1] + \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ -1 \end{bmatrix} = -1 + -1 = -2$$

Which is the same as B_{21}. Finally, considering $(i, j) = (2, 2)$, we have

$$A_{21}X_{12} + A_{22}X_{22} = [-1] \begin{bmatrix} 0 & 1 \end{bmatrix} + \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 2 & 0 \end{bmatrix} = \begin{bmatrix} 0 & -1 \end{bmatrix} + \begin{bmatrix} 2 & 0 \end{bmatrix} = \begin{bmatrix} 2 & -1 \end{bmatrix}$$

Which is the same as B_{22}. Thus we have shown the equivalence requested.

Exercise 1.1.25

Let A be n by m and consider A partitioned by columns so that

$$A = \left[a^1 | a^2 | \cdots | a^m\right] .$$

Here a^i is the i-th column of A of dimension n by 1. Consider x partitioned into m scalar blocks as

$$x = \left[\begin{array}{c} x_1 \\ \hline x_2 \\ \hline \vdots \\ \hline x_m \end{array}\right]$$

Then

$$Ax = \left[a^1 | a^2 | \cdots | a^m\right] \left[\begin{array}{c} x_1 \\ \hline x_2 \\ \hline \vdots \\ \hline x_m \end{array}\right] = x_1 a^1 + x_2 a^2 + x_3 a^3 + \cdots + x_m a^m = b \tag{1}$$

Showing that b is a linear combination of the columns of A.

Exercise 1.2.4

Lets begin by assuming that there exists a nonzero y such that $Ay = 0$. But applying A^{-1} to both sides gives

$$A^{-1}Ay = A^{-1}0 = 0$$

or $y = 0$ which is a contradiction to our initial assumption. Therefore no nonzero y can exist.

Exercise 1.2.5

If A^{-1} exists then we have

$$AA^{-1} = I\,.$$

Taking the determinant of both sides of this expression gives

$$|A||A^{-1}| = 1\,, \tag{2}$$

but since $|A^{-1}| = |A|^{-1}$ it is not possible for $|A| = 0$ or else Equation 2 would be in contradiction.

Exercise 1.2.11

The equation for the car at x_1 is

$$-4x_1 + 1 + 4(x_2 - x_1) = 0\,,$$

or

$$8x_1 - 4x_2 = 1\,.$$

Which is the same as given in the book. The equation for the car at x_2 is given by

$$-4(x_2 - x_1) + 2 + 4(x_3 - x_2) = 0\,.$$

Which is the same as in the book. The equation for the car at x_3 is given by

$$-4(x_3 - x_2) + 3 - 4x_3 = 0\,,$$

or

$$-4x_2 - 8x_3 = -3\,.$$

Which is the same as given in the book. Since the determinant of the coefficient matrix A has value $\det(A) = 256 \neq 0$, A is nonsingular.

Exercise 1.3.7

A pseudo-code implementation would look like the following

```
% Find the last element of b that is zero ($b_k=0$ and $b_{k+1} \ne 0$)
for i=1,2,...,n
  if( b_i \ne 0 ) break
endfor
k=i-1 % we now have b_1 = b_2 = b_3 = \ldots = b_k = 0
for i=k+1,...,n
  for j=k+1,...,n
    b_i = b_i - g_{ij} b_j
  endfor
  if( g_{ii} = 0 ) set error flag, exit
  b_i = b_i/g_{ii}
endfor
```

Exercise 1.3.14

Part (a): To count the operations we first consider the inner for loop which has 2 flops and is executed $n-(j+1)+1 = n-j$ times. The outer loop is executed once for every $j = 1, 2, \ldots, n$ therefore the total flop count is given by

$$\sum_{j=1}^{n} \sum_{i=j+1}^{n} 2 = \sum_{j=1}^{n} 2(n-j) = 2 \sum_{j=1}^{n-1} j = 2\frac{1}{2} n(n-1) = n(n-1)\,,$$

the same as row oriented substitution.

Part (b): Row oriented forward substitution subtracts just the columns of the row we are working on as we get to each row. Column oriented forward substitution subtracts from all rows before moving to the next unknown (row).

Exercise 1.3.14

Part (b): In row-oriented substitution we first solve for x_1, then x_2 using the known value of x_1, then x_3 using the known values of x_1 and x_2. This pattern continues until finally we solve for x_n using all of then known x_1, x_2, down to x_{n-1}. The column subtractions (in column oriented substitution) occurs one at a time i.e. in row-oriented substitution we have

$$a_{11}x_1 = b_1 \Rightarrow x_1 = \frac{b_1}{a_{11}}\,.$$

$$a_{21}x_1 + a_{22}x_2 = b_2 \quad \Rightarrow \quad x_2 = \frac{b_2 - a_{21}x_1}{a_{22}}$$
$$\Rightarrow \quad x_2 = \frac{b_2 - a_{21}\left(\frac{b_1}{a_{11}}\right)}{a_{22}}$$

while in column oriented substitution we solve for x_1, with

$$x_1 = \frac{b_1}{a_{11}} .$$

and then we get

$$a_{22}x_2 = b_2 - a_{21}\left(\frac{b_1}{a_{11}}\right) ,$$

and solve for x_2. In summary using column oriented substitution we do some of the subtractions in $x_i = \frac{b_i - \sum_{j=1}^{i-1} a_{ij}y_j}{a_{ij}}$, each time we go through the loop. In row-oriented does all subtractions at once.

Exercise 1.3.15

A pseudocode implementation of row-oriented back substitution would look something like the following

```
for i=n,n-1,...,2,1 do
    for j=n,n-1,...,i+2,i+1 do
        b(i) = b(i) - A(i,j)*b(j)
    end
    if A(i,i)=0 then
        // return error; system is singular
    end
    b(i) = b(i)/A(i,i)
end
```

Exercise 1.3.16 (column-oriented back substitution)

A pseudocode implementation of column-oriented back substitution would look something like the following

Note that the i loop above could be written backwards as $i = j-1, j-2, \ldots, 2, 1$ if this helps maintain consistency.

```
for j=n,n-1,...,2,1 do
|  if  A(j,j)=0 then
|  |  // return error; system is singular
|  end
|  b(j) = b(j)/A(j,j)
|  for i=1,2,...,j-1 do
|  |  b(i) = b(i) - A(i,j)*b(j)
|  end
end
```

Exercise 1.3.16

Column oriented substitution we factor our problem $Ux = y$ as

$$\begin{bmatrix} \hat{U} & h \\ 0 & u_{nn} \end{bmatrix} \begin{bmatrix} \hat{x} \\ x_n \end{bmatrix} = \begin{bmatrix} \hat{y} \\ y_n \end{bmatrix},$$

which written out in each

$$\begin{aligned} \hat{U}\hat{x} + hx_n &= \hat{y} \\ u_{nn}x_n &= y_n \end{aligned}$$

We first solve for x_n then solve for $\hat{x}$ in

$$\hat{U}x = \hat{y} - hx_n\,.$$

This is a $n-1 \times n-1$ upper triangular system where we make the same substitutions, so we let

$$\begin{aligned} y_1 &= y_1 - u_{1n}x_n \\ y_2 &= y_2 - u_{2n}x_n \\ &\vdots \\ y_{n-1} &= \cdots \end{aligned}$$

Put x_n in y_n. The algorithm is given by

```
for i:-1:n do
  if u(i,i)=0 set flag
  y(i) = y(i)/u(i,i)
  for j = i-1:-1:1 do
    y(j) = y(j) - u(j,i) y(i)
```

The values of x stored in $y_1, y_2, y_3, \cdots y_n$. We can check this for the index n. We have

$$\begin{aligned} y_n &\leftarrow y_n/u_{nn} \\ y_{n-1} &\leftarrow y_{n-1} - u_{n-1,n}x_n \\ &\vdots \\ y_1 &\leftarrow y_1 - u_{1,n}x_n\,. \end{aligned}$$

Exercise 1.3.17

Part (a): Performing row oriented back-substitution on

$$\begin{bmatrix} 3 & 2 & 1 & 0 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & -2 & 1 \\ 0 & 0 & 0 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} -10 \\ 10 \\ 1 \\ 12 \end{bmatrix},$$

we have

$$\begin{aligned} x_4 &= 3 \\ x_3 &= \frac{1-3}{-2} = 1 \\ x_2 &= 10 - 2(1) - 3(3) = -1 \\ x_1 &= \frac{-10 - 2(-1) - 1(1)}{3} = \frac{-10+2-1}{3} = \frac{-9}{3} = -3 \end{aligned}$$

Part (b): Column oriented back-substitution, would first solve for x_4 giving $x_4 = 3$, and then reduce the order of the system by one giving

$$\begin{bmatrix} 3 & 2 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -10 \\ 10 \\ 1 \end{bmatrix} - 3 \begin{bmatrix} 0 \\ 3 \\ 1 \end{bmatrix} = \begin{bmatrix} -10 \\ 1 \\ -2 \end{bmatrix}.$$

This general procedure is then repeated by solving for x_3. The last equation above gives $x_3 = 1$, and then reducing the order of the system above gives

$$\begin{bmatrix} 3 & 2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -10 \\ 1 \end{bmatrix} - 1 \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} -11 \\ -1 \end{bmatrix}.$$

Using the last equation to solve for x_2 gives $x_2 = -1$, and reducing the system one final time gives

$$[3]x_1 = [-11] - 2[-1] = -9\,,$$

which has as its solution $x_1 = 3$.

Exercise 1.3.20 (block column oriented forward substitution)

The block variant of column oriented forward substitution on

$$\begin{bmatrix} G_{11} & & & & \\ G_{21} & G_{22} & & & \\ G_{31} & G_{32} & G_{33} & & \\ \vdots & \vdots & & \ddots & \\ G_{s1} & G_{s2} & G_{s3} & & G_{ss} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_s \end{bmatrix}$$

would look like

```
for j=1,2,...,s do
    if G(j,j) is singular then
        // return error; system is singular
    end
    // matrix inverse of G(j,j) taken here
    b(j)=inverse(G(j,j))*b(j)
    for i=j+1,j+2,...,s do
        // matrix multiplication performed here
        b(i) = b(i) - G(i,j)*b(j)
    end
end
```

Exercise 1.3.21 (row and column back substitution)

The difference between (block) row and column back substitution is that in row forward substitution as we are processing each row we subtract out multiples of the computed unknowns. In column substitution we do all the subtractions required for each row at one time and then no longer need to remember these unknowns. The difference is simply a question of doing the calculations all at once or each time we access a row.

Exercise 1.3.23 (the determinant of a triangular matrix)

The fact that the determinant of a triangular matrix is equal to the product of the diagonal elements, can easily be proved by induction. Lets assume without loss of generality that our system is *lower* triangular (upper triangular systems are transposes of lower triangular systems) and let $n = 1$ then $|G| = g_{11}$ trivially. Now assume that for a triangular system of size $n \times n$ that the determinant is given by the product of its n diagonal elements and consider a matrix $\tilde{G}$ of size $(n+1) \times (n+1)$ partitioned into a leading matrix G_{11} of size $n \times n$.

$$G = \begin{bmatrix} G_{11} & 0 \\ h^T & g_{n+1,n+1} \end{bmatrix}.$$

Now by expanding the determinant of G along its last column we see that

$$|G| = g_{n+1,n+1}|G_{11}| = g_{n+1,n+1} \prod_{i=1}^{n} g_{ii} = \prod_{i=1}^{n+1} g_{ii},$$

proving by induction that the determinant of a triangular matrix is equal to the product of its diagonal elements.

Exercise 1.3.29

Part (b):

$$\begin{bmatrix} \hat{G} & 0 \\ h^T & g_{nn} \end{bmatrix} \begin{bmatrix} y \\ y_n \end{bmatrix} = \begin{bmatrix} \hat{b} \\ b_n \end{bmatrix},$$

Given $y_1, y_2, \cdots, y_{i-1}$ we compute y_i from

$$y_i = \frac{b_i - h^T \hat{y}}{a_{ii}}.$$

As an algorithm we have

```
for i:1:n do
  if a(i,i)=0 (set error)
  for j=1:i-1 do
    b(i) = b(i)-a(i,j) y(j)
    b(i) = b(i)/a(i,i)
  end
end
```

Exercise 1.4.15

If $A = \begin{bmatrix} 4 & 0 \\ 0 & 9 \end{bmatrix}$.

Part (a): $x^T A x = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 9 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 4x_1^2 + 9x_2^2 \geq 0$, which can bet equal to zero only if $x = 0$.

Part (b): $A = GG^T$ we have $g_{11} = \pm\sqrt{a_{11}}$, we take the positive sign, so $g_{11} = 2$. Now

$$g_{i1} = \frac{a_{i1}}{g_{11}} \quad i = 2, \cdots, n$$

so we have that

$$g_{21} = \frac{a_{21}}{g_{11}} = \frac{0}{2} = 0.$$

Multiplying out GG^T we see that

$$\begin{aligned} a_{11} &= g_{11}^2 \\ a_{12} &= g_{11} g_{21} \\ a_{22} &= g_{21}^2 + g_{22}^2. \end{aligned}$$

$$a_{i2} = g_{i1} g_{21} + g_{i2} g_{22}.$$

$$a_{22} = 0 + g_{22}^2 \Rightarrow g_{22} = \sqrt{a_{22}} = 3.$$

so

$$A = \begin{bmatrix} 4 & 0 \\ 0 & 9 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} = GG^T$$

so the Cholesky factor of A is $\begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$.

Part (c): $G_2 = \begin{bmatrix} -2 & 0 \\ 0 & 3 \end{bmatrix}$, $G_3 = \begin{bmatrix} 2 & 0 \\ 0 & -3 \end{bmatrix}$, $G_4 = \begin{bmatrix} -2 & 0 \\ 0 & -3 \end{bmatrix}$.

Part (d): We can change the sign of any of the elements on the diagonal, so there are $2 \cdot 2 \cdot 2 \cdots 2 = 2^n$ where n is the number of diagonal elements so 2^n lower triangular matrices.

Exercise 1.4.21

For $A = \begin{bmatrix} 16 & 4 & 8 & 4 \\ 4 & 10 & 8 & 4 \\ 8 & 8 & 12 & 10 \\ 4 & 4 & 10 & 12 \end{bmatrix}$.

$$\begin{aligned}
g_{11} &= \sqrt{16} = 4 \\
g_{21} &= \frac{4}{4} = 1 \\
g_{31} &= \frac{8}{4} = 2 \\
g_{41} &= \frac{4}{4} = 1 \\
g_{22} &= \sqrt{a_{22} - g_{21}^2} = \sqrt{10 - 1^2} = 3 \\
g_{32} &= \frac{a_{32} - g_{31}g_{21}}{g_{22}} = \frac{8 - 2(1)}{3} = 2 \\
g_{42} &= \frac{a_{42} - g_{41}g_{21}}{g_{22}} = \frac{4 - 1(1)}{3} = 1 \\
g_{33} &= \sqrt{a_{33} - g_{31}^2 - g_{32}^2} = \sqrt{12 - 4 - 4} = 2 \\
g_{43} &= \frac{a_{43} - g_{41}g_{31} - g_{42}g_{32}}{g_{33}} = \frac{10 - 1(2) - 1(2)}{2} = 3 \\
g_{44} &= \sqrt{a_{44} - g_{41}^2 - g_{42}^2 - g_{43}^2} = \sqrt{12 - 1^2 - 1^2 - 3^2} = 1\,.
\end{aligned}$$

$$G = \begin{bmatrix} 4 & 0 & 0 & 0 \\ 1 & 3 & 0 & 0 \\ 2 & 2 & 2 & 0 \\ 1 & 1 & 3 & 1 \end{bmatrix}.$$

Since we want to solve $GG^T x = b$, we let $y = G^T x$ to solve $Gy = b$, which is the system

$$\begin{bmatrix} 4 & 0 & 0 & 0 \\ 1 & 3 & 0 & 0 \\ 2 & 2 & 2 & 0 \\ 1 & 1 & 3 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} 32 \\ 26 \\ 38 \\ 30 \end{bmatrix}.$$

The first equation gives $y_1 = 8$, which put in the second equation gives $8 + 3y_2 = 26$ or $y_2 = 6$. When we put these two variables into the third equation we get

$$16 + 12 + 2y_3 = 38 \Rightarrow y_3 = 5\,.$$

When all of these variables are put into the fourth equation we have

$$8 + 6 + 15 + y_4 = 30 \Rightarrow y_4 = 1\,.$$

Using these values of y_i we now want solve

$$\begin{bmatrix} 4 & 1 & 2 & 1 \\ 0 & 3 & 2 & 1 \\ 0 & 0 & 2 & 3 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 8 \\ 6 \\ 5 \\ 1 \end{bmatrix}.$$

The fourth equation gives $x_4 = 1$. The third equation is then $2x_3 + 3 = 5 \Rightarrow x_3 = 1$. With both of these values into the second equation we get

$$3x_2 + 2 + 1 = 6 \Rightarrow x_2 = 1\,.$$

With these three values put into the first equation we get

$$4x_1 + 1 + 2 + 1 = 8 \Rightarrow x_1 = 1\,.$$

Thus in total we have $x = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$.

Exercise 1.4.52

Since A is positive definite we must have for all non-zero x's the condition $x^T Ax > 0$. Let $x = e_i$, a vector of all zeros but with a one in the i-th spot. Then $x^T Ax = e_i^T Ae_i = a_{ii} > 0$, proving that the diagonal elements of a positive definite matrix must be positive.

Exercise 1.4.56

Let v be a nonzero vector and consider

$$v^T X^T AXv = (Xv)^T A(Xv) = y^T Ay > 0\,,$$

where we have defined $y = Xv$, and the last inequality is from the positive definiteness of A. Thus, $X^T AX$ is positive definite. Note, if X were singular then $X^T AX$ is only positive *semi* definite since there would then exist nonzero v such that $Xv = 0$, and therefore $v^T X^T AXv = 0$, in contradiction to the requirement of positive definiteness.

Exercise 1.4.58

Part (a): Consider the expression for $\tilde{A}_{22}$, we have

$$\begin{aligned}\tilde{A}_{22} &= A_{22} - R_{12}^T R_{12} \\ &= A_{22} - (R_{11}^{-T} A_{12})^T (R_{11}^{-T} A_{12}) \\ &= A_{22} - A_{12}^T R_{11}^{-1} R_{11}^{-T} A_{12} \\ &= A_{22} - A_{12}^T (R_{11}^T R_{11})^{-1} A_{12} \\ &= A_{22} - A_{12}^T A_{11}^{-1} A_{12}\end{aligned}$$

which results from the definition of $A_{11} = R_{11}^T R_{11}$.

Part (b): Since A is symmetric we must have that $A_{21} = A_{12}^T$ and following the discussion on the previous page A has a decomposition like the following

$$\begin{aligned}A &= \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \\ &= \begin{bmatrix} R_{11}^T & 0 \\ A_{12}^T R_{11}^{-1} & I \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & \tilde{A}_{22} \end{bmatrix} \begin{bmatrix} R_{11} & R_{11}^{-T} A_{12} \\ 0 & I \end{bmatrix} . \end{aligned} \tag{3}$$

This can be checked by expanding the individual products on the right hand side (RHS) as follows

$$\text{RHS} = \begin{bmatrix} R_{11}^T & 0 \\ A_{12}^T R_{11}^{-1} & I \end{bmatrix} \begin{bmatrix} R_{11} & R_{11}^{-T} A_{12} \\ 0 & \tilde{A}_{22} \end{bmatrix} = \begin{bmatrix} R_{11}^T R_{11} & A_{12} \\ A_{12}^T & \tilde{A}_{22} + A_{12}^T R_{11}^{-1} R_{11}^{-T} A_{12} \end{bmatrix}$$

Which will equal A when

$$\tilde{A}_{22} = A_{22} - A_{12}^T R_{11}^{-1} R_{11}^{-T} A_{12} .$$

Part (c): To show that $\tilde{A}_{22}$ is positive definite we note that by first defining X to be

$$X^T \equiv \begin{bmatrix} R_{11}^T & 0 \\ A_{12}^T R_{11}^{-1} & I \end{bmatrix} ,$$

we see that X is nonsingular at from Eq. 3 we have that

$$\begin{bmatrix} I & 0 \\ 0 & \tilde{A}_{22} \end{bmatrix} = X^{-T} A X^{-1} .$$

By the fact that since X^{-1} is invertible and A is positive definite then we have that $X^{-T} A X^{-1}$ is also positive definite and therefore by the square partitioning of principle submatrices of positive definite matrices (Proposition 1.4.53) the submatrix $\tilde{A}_{22}$ is positive definite.

Exercise 1.4.60

We will do this problem using induction on n the size of our matrix A. If $n = 1$ then showing that the Cholesky factor is unique is trivial

$$[a_{11}] = [+\sqrt{a_{11}}][+\sqrt{a_{11}}] .$$

Assume that the Cholesky factor is unique for all matrices of size less than or equal to n. Let A be a positive definite of size $n+1$ and assume (to reach a contradiction) that A has two Cholesky factorizations, i.e. $A = R^T R$ and $A = S^T S$. By partitioning A and R as follows

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{12}^T & a_{n+1,n+1} \end{bmatrix} = \begin{bmatrix} R_{11}^T & 0 \\ R_{12}^T & r_{n+1,n+1} \end{bmatrix} \begin{bmatrix} R_{11} & R_{12} \\ 0 & r_{n+1,n+1} \end{bmatrix}.$$

Here A_{11} is of dimension n by n, A_{12} is n by 1, and the elements of R are partitioned conformably with A. Then by equating terms in the above block matrix equation we have three unique equations (equating terms for A_{12}^T results in a transpose of the equation for A_{12}) The three equations are

$$\begin{aligned} A_{11} &= R_{11}^T R_{11} \\ A_{12} &= R_{11}^T R_{12} \\ a_{n+1,n+1} &= R_{12}^T R_{12} + r_{n+1,n+1}^2 . \end{aligned}$$

From the induction hypotheses since A_{11} is n by n its Cholesky decomposition (i.e. R_{11}) is unique. This implys that the column vector R_{12} is unique since R_{11} and correspondingly R_{11}^T is invertible. It then follows that since $r_{n+1,n+1} > 0$ that $r_{n+1,n+1}$ must be unique since it must equal

$$r_{n+1,n+1} = +\sqrt{a_{n+1,n+1} - R_{12}^T R_{12}} .$$

We know that the expression

$$a_{n+1,n+1} - R_{12}^T R_{12} > 0$$

by an equivalent result to the Schur complement result discussed in the book. Since every component of our construction of R is unique, we have proven that the Cholesky decomposition is unique for positive definite matrices of size $n+1$. By induction it must hold for positive definite matrices of any size.

Exercise 1.4.62

Since A is positive definite it has a Cholesky factorization given by $A = R^T R$. Taking the determinant of this expression gives

$$|A| = |R^T||R| = |R|^2 = \left(\prod_{i=1}^{n} r_{ii} \right)^2 > 0 ,$$

showing the desired inequality.

Exercise 1.5.9

We wish to count the number of operations required to solve the following banded system $Rx = y$ with semiband width s or

$$\begin{bmatrix} r_{1,1} & r_{1,2} & r_{1,3} & \dots & r_{1,s} & & \\ & r_{2,2} & r_{2,3} & \dots & r_{2,s} & r_{2,s+1} & \\ & & & & & & \\ & & & & r_{n-1,n-1} & r_{n-1,n} & \\ & & & & & r_{n,n} & \end{bmatrix} x = y$$

so at row i we have non-zero elements in columns $j = i, i+1, i+2, \dots i+s$, assuming that $i+s$ is less than n. Then a pseudocode implementation of row-oriented back substitution would look something like the following

for *i=n,n-1,...,2,1* **do**
  `// the following is not executed when j=n`
  for *j=*`min`*(n,i+s),*`min`*(n,i+s)-1,...,i+2,i+1* **do**
    y(j) = y(j) - R(i,j)*y(j)
  end
  if *R(i,i)=0* **then**
    `// return error; system is singular`
  end
  y(i) = y(i)/A(i,i)
end

Counting the number of flops this requires, we have approximately two flops for every execution of the line $y(j) = y(j) - R(i,j) * y(j)$, giving the following expression for the number of flops

$$\sum_{i=n}^{1} \left[\left(\sum_{j=\min(n,i+s)}^{i+1} 2 \right) + 1 \right] .$$

Now since

$$\sum_{j=\min(n,i+s)}^{i+1} 2 = O(2s)$$

the above sum simplifies (using order notation) to

$$O(n + 2sn) = O(2sn) ,$$

as requested.

Exercise 1.6.4

Please see the Matlab file `exercise_1_6_4.m` for the evaluation of the code to perform the suggested numerical experiments. From those experiments one can see that the minimum

degree ordering produces the best ordering as far as remaining non-zero elements and time to compute the Cholesky factor. This is followed by the reverse Cuthill-McKee ordering. Since the matrix `delsq` is banded to begin with a direct Cholesky factorization can be computed rather quickly. Randomizing the ordering of the nodes produces the matrix with the most fill in (largest number of non-zero elements) and also requires the largest amount of time to compute the Cholesky factorization for.

Exercise 1.6.5

Please see the Matlab file `exercise_1_6_5.m` for the evaluation of the code to perform the suggested numerical experiments. From those experiments one can see that the minimum degree ordering produces the best ordering as far as remaining non-zero elements and time to compute the Cholesky factor. This is followed by the reverse Cuthill-McKee ordering.

Exercise 1.7.18 (solving linear systems with LU)

From exercise 1.7.10 our matrix A has the following LU decomposition

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

Then we are solving $LUx = b$ by first solving $Ly = b$ and then $Ux = y$. The first problem is $Ly = b$ or

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 2 & -1 & 1 & 0 \\ -3 & 2 & -1 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} 12 \\ -8 \\ 21 \\ -26 \end{bmatrix}$$

which upon performing forward substitution gives

$$\begin{aligned} y_1 &= 12 \\ y_2 &= -8 + y_1 = -8 + 12 = 4 \\ y_3 &= 21 - 2y_1 + y_2 = 21 - 24 + 4 = 1 \\ y_4 &= -26 + 3y_1 - 2y_2 + y_3 = 3(12) - 2(4) + 1 = 29 \end{aligned}$$

The second step is to solve $Ux = y$ or the system

$$\begin{bmatrix} 2 & 1 & -1 & 3 \\ 0 & 1 & -1 & 3 \\ 0 & 0 & -1 & 3 \\ 0 & 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 12 \\ 4 \\ 1 \\ 29 \end{bmatrix}$$

which upon performing backward substitution gives

$$\begin{aligned} x_4 &= 13 \\ -x_3 &= 1 - 3x_1 = 1 - 39 = -38 \quad \Rightarrow \quad x_3 = 38 \\ x_2 &= x_3 - 3x_4 = 38 - 39 = -1 \\ 2x_1 &= -x_2 + x_3 - 3x_4 = 1 + 38 - 29 = 10 \quad \Rightarrow \quad x_1 = 5 \end{aligned}$$

so our solution is given by

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 5 \\ -1 \\ 38 \\ 13 \end{bmatrix}.$$

Exercise 1.7.34

Part (a): By considering the given matrix M and the product MA we see that we are multiplying row j by m and adding it to row i. This is the *definition* of a row operations of type 1.

Part (b): Since the given matrix M, is also lower triangular the determinant of M is the product of the diagonal elements. In this case, since the diagonal elements are all ones the product is then also a one, and therefore $\det(M) = +1$. Since $\hat{A} = MA$ we have that

$$|\hat{A}| = |MA| = |M||A| = |A|\,.$$

Part (c): I will assume that the inverse of M is given by replacing the element m in M by its negative $-m$. We can check this for correctness by multiplying the two matrices as $M^{-1}M$ and observing that we obtain the identity. This result can also be understood as recognizing that the inverse of M (which is action of multiplying row j by m and adding it to row i), would be the action of multiplying row j by $-m$ and adding it to row i. This action is the same as replacing the (i,j)th element in I with $-m$.

Exercise 1.7.35

Part (a): By multiplying a matrix A by the matrix M described we find the ith and jth rows of A exchanged.

Part (b): Since M is obtained from the identity matrix by exchanging two rows of I the determinant change sign i.e.

$$\det(M) = -\det(I) = -1\,,$$

so

$$\det(\hat{A}) = \det(MA) = \det(M)\det(A) = -\det(A)\,.$$

Part (c): The inverse of exchanging the ith row and the jth row would be performing this operation twice so $M^{-1} = M$. Also we can see that this is true by explicitly calculating the multiplication of M with itself.

Exercise 1.7.36

Part (a): M is obtained from the identity matrix by replacing the ith row by c times the ith row, i.e. $M_{ii} = c$.

Part (b): M^{-1} is obtained from the identity matrix by replacing the ith row by $1/c$ times the ith row, i.e. $M_{ii} = 1/c$.

Part (c): We have

$$\det(M) = c\det(I) = c\,,$$

so

$$\det(\hat{A}) = \det(MA) = \det(M)\det(A) = c\det(A)\,.$$

Exercise 1.7.44 (triangular matrices have triangular inverses)

Part (a): We are told that L is non-singular and lower triangular. We want to prove that L^{-1} is lower triangular. We will do this by using induction on n the dimension of L. For $n = 1$ L is a scalar and L^{-1} is also a scalar. Trivially both are lower triangular. Now assume that if L is non-singular and lower triangular of size $n \times n$, then L^{-1} has the same property. Let L be a matrix of size $(n+1) \times (n+1)$ and partition L as follows

$$L = \begin{bmatrix} L_{11} & 0 \\ L_{21} & L_{22} \end{bmatrix}.$$

Where L_{11} and L_{22} are both lower triangular matrices of sizes less than $n \times n$, so that we can apply the induction hypothesis. Let $M = L^{-1}$ and partition M con formally i.e.

$$M = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}.$$

We want to show that M_{12} must be zero. Now since $ML = I$ by multiplying the matrices above out we obtain

$$\begin{aligned} LM &= \begin{bmatrix} L_{11} & 0 \\ L_{21} & L_{22} \end{bmatrix} \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} \\ &= \begin{bmatrix} L_{11}M_{11} & L_{11}M_{12} \\ L_{21}M_{11} + L_{22}M_{21} & L_{21}M_{12} + L_{22}M_{22} \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} \end{aligned}$$

Equating block components gives

$$\begin{aligned} L_{11}M_{11} &= I \\ L_{11}M_{12} &= 0 \\ L_{21}M_{11} + L_{22}M_{21} &= 0 \\ L_{21}M_{12} + L_{22}M_{22} &= I\,. \end{aligned}$$

By the induction hypothesis both L_{11} and L_{22} are invertible. Thus the equation $L_{11}M_{11} = I$ gives $M_{11} = L_{11}^{-1}$, and the equation $L_{11}M_{12} = 0$ gives $M_{12} = 0$. With these two conditions the equation $L_{21}M_{12} + L_{22}M_{22} = I$ becomes $L_{22}M_{22} = I$. Since L_{22} is invertible we compute that $M_{22} = L_{22}^{-1}$. As both L_{11} and L_{22} are lower triangular of size less than $n \times n$ by the induction hypothesis their inverse are lower triangular and we see that M itself is then lower triangular since

$$M = \begin{bmatrix} L_{11}^{-1} & 0 \\ M_{21} & L_{22}^{-1} \end{bmatrix}.$$

Thus by the principle of induction we have shown that the inverse of a lower triangular matrix is lower triangular.

Part (b): We can prove that the main diagonal elements of L^{-1} are given by l_{ii}^{-1} in a number of ways. One is by using mathematical induction, another is to simply compute the product of the corresponding row of L with the corresponding column of L^{-1}. For example since the ith row of L multiplied by the ith column of L^{-1} must produce unity we have

$$1 = \sum_{k=1}^{n} L_{ik}(L^{-1})_{ki}\,.$$

Since L is lower triangular L^{-1} is lower triangular so we have that their components must satisfy

$$\begin{aligned} L_{ik} &= 0 \quad \text{for} \quad k > i \\ (L^{-1})_{ki} &= 0 \quad \text{for} \quad i > k \end{aligned}$$

so that the above sum becomes

$$1 = L_{ii}(L^{-1})_{ii} \quad \text{or} \quad (L^{-1})_{ii} = \frac{1}{L_{ii}}\,.$$

Exercise 1.7.45 (product of lower triangular matrices)

Part (a): We will prove that the product of two lower triangular matrices is lower triangular by induction. We begin with $n = 2$ for which we have

$$\begin{bmatrix} l_{11} & 0 \\ l_{21} & l_{22} \end{bmatrix} \begin{bmatrix} m_{11} & 0 \\ m_{21} & m_{22} \end{bmatrix} = \begin{bmatrix} l_{11}m_{11} & 0 \\ l_{21}m_{11} + l_{22}m_{21} & l_{22}m_{22} \end{bmatrix}$$

which is lower triangular. Assume the product of two lower triangular matrices of size $\hat{n} \leq n$ is also lower triangular and consider two lower triangular matrices of size $n+1$. Performing

a "bordered" block partitioning of the two lower triangular matrices we have that

$$L_{n+1} = \begin{bmatrix} L_n & 0 \\ l^T & l_{n+1,n+1} \end{bmatrix} \quad \text{and} \quad M_{n+1} = \begin{bmatrix} M_n & 0 \\ m^T & m_{n+1,n+1} \end{bmatrix}$$

where the single subscripts denote the order of the matrices. With bordered block decomposition of our two individual matrices we have a product given by

$$L_{n+1}M_{n+1} = \begin{bmatrix} L_nM_n & 0 \\ l^TM_n + l_{n+1,n+1}m^T & l_{n+1,n+1}m_{n_1,n+1} \end{bmatrix}.$$

Since by the induction hypotheses the product L_nM_n is lower triangular we see that our product $L_{n+1}M_{n+1}$ is lower triangular.

Part (b): This can again be proved by induction. For $n = 2$ we see our diagonal elements are given by $l_{11}m_{11}$ and $l_{22}m_{22}$. Assuming this is true for $\hat{n} \leq n$. The fact that the $n+1$th diagonal element of $L_{n+1}M_{n+1}$ in the above bordered factorization is given by $l_{n+1,n+1}m_{n_1,n+1}$ and the induction hypothesis applied to L_nM_n shows that the diagonal elements of $L_{n+1}M_{n+1}$ have the correct form. As stated in the book this implies that the product of two unit lower triangular matrices is unit triangular.

One can prove that the diagonal elements of products of unit triangular matrices are ones by another method. Letting L and M be unit lower triangular we have that the ith row of L is given by

$$L_{i,:} = [l_{i,1}, l_{i,2}, \ldots l_{i,i-1}, 1, 0, \ldots 0]$$

the ith column of the matrix M is given by

$$M_{:,i} = [0, 0, \ldots, 0, 1, m_{i+1,i}, m_{i+2,i}, \ldots, m_{n,i}]^T$$

Then the (i,i) element in LM is given by the inner product of these two vectors and gives one since this is the only non-zero overlapping component.

Exercise 1.7.46 (product of upper triangular matrices)

Let matrices U and V be upper triangular, then U^T and V^T are lower triangular. Because of this U^TV^T is lower triangular. Since $U^TV^T = (VU)^T$, we see that $(VU)^T$ is lower triangular. Taking the transpose we see that VU is upper triangular. Thus the product of upper triangular matrices are again upper triangular.

Exercise 1.7.47

Part (e): Our L matrix is given by $L = M_1M_2M_3 \ldots M_{n-2}M_{n-1}$. To compute this product we will multiply the factors from the left to the right. To begin, consider the product of

$M_{n-2}M_{n-1}$ which is given by

$$
\begin{aligned}
M_{n-2}M_{n-1} &= \begin{bmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & & 1 & 0 & 0 \\ & & & m_{n-1,n-2} & 1 & 0 \\ & & & m_{n,n-2} & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & & 1 & 0 & 0 \\ & & & 0 & 1 & 0 \\ & & & 0 & m_{n,n-1} & 1 \end{bmatrix} \\
&= \begin{bmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & & 1 & 0 & 0 \\ & & & m_{n-1,n-2} & 1 & 0 \\ & & & m_{n,n-2} & m_{n,n-1} & 1 \end{bmatrix}.
\end{aligned}
$$

From which we see that the effect of this multiplication is to have inserted the two elements $m_{n-1,n-2}$ and $m_{n,n-2}$ into the lower right block matrix in M_{n-1}. Motivated by this observation we will now block decompose the product $M_{n-2}M_{n-1}$ into a matrix like

$$
\begin{bmatrix} I_{n-4} & 0 \\ 0 & A \end{bmatrix},
$$

with A defined as

$$
A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & m_{n-1,n-2} & 1 & 0 \\ 0 & m_{n,n-2} & m_{n,n-1} & 1 \end{bmatrix} \equiv \begin{bmatrix} 1 & 0 \\ 0 & M \end{bmatrix}.
$$

Here we have increased the dimension of lower right block matrix by one over the lower right block matrix annotated in the product $M_{n-2}M_{n-1}$ above. In addition, we have defined the matrix M to be

$$
M = \begin{bmatrix} 1 & 0 & 0 \\ m_{n-1,n-2} & 1 & 0 \\ m_{n,n-2} & m_{n,n-1} & 1 \end{bmatrix}
$$

Now consider the matrix M_{n-3} given by

$$
M_{n-3} = \begin{bmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & & 1 & 0 & 0 & 0 \\ & & & m_{n-2,n-3} & 1 & 0 & 0 \\ & & & m_{n-1,n-3} & 0 & 1 & 0 \\ & & & m_{n,n-3} & 0 & 0 & 1 \end{bmatrix}.
$$

which can be viewed as having the same block structure as the product $M_{n-2}M_{n-1}$ above by considering a block decomposition of this matrix as

$$
M_{n-3} = \begin{bmatrix} I & 0 \\ 0 & B \end{bmatrix},
$$

where B is given by

$$B = \begin{bmatrix} 1 & 0 & 0 & 0 \\ m_{n-2,n-3} & 1 & 0 & 0 \\ m_{n-1,n-3} & 0 & 1 & 0 \\ m_{n,n-3} & 0 & 0 & 1 \end{bmatrix} \equiv \begin{bmatrix} 1 & 0 \\ m & I \end{bmatrix}.$$

Where we have introduced the vector m defined as $m = [m_{n-2,n-3}, m_{n-1,n-3}, m_{n,n-3}]^T$. Then the block product of $M_{n-3}(M_{n-2}M_{n-1})$ is given by

$$M_{n-3}M_{n-2}M_{n-1} = \begin{bmatrix} I & 0 \\ 0 & B \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & A \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & AB \end{bmatrix}. \tag{4}$$

For the product of B with A, we have from their definitions that

$$BA = \begin{bmatrix} 1 & 0 \\ m & I \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & M \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ m & M \end{bmatrix}. \tag{5}$$

From which we see that the effect of the multiplication BA is to replace the first column of A with the first column of B. Since everything we have done simply involves viewing the various multiplications as block matrices in specific ways this procedure can be repeated at each step, i.e. we can form block matrices of the current running product $M_i M_{i+1} \dots M_{n-1}$, with the current M_{i-1} to multiply on the left by, just as is done in Eq. 4. The lower right block multiplication can be decomposed further in a form exactly like Eq. 5. By repeating this procedure $n-2$ times corresponding to all required products we will have produced a matrix L as given in the book.

Exercise 1.7.44/46

Part (a): For $n = 2$ we have

$$\begin{bmatrix} a_{11} & a_{12} \\ 0 & a_{22} \end{bmatrix}^{-1} = \begin{bmatrix} a_{11}^{-1} & -a_{11}^{-1}a_{12}a_{22}^{-1} \\ 0 & a_{22}^{-1} \end{bmatrix}.$$

So an upper triangular matrix $n = 2$ has an upper triangular matrix as its inverse. Assume this is true for $n = k$, by mathematical induction we want to show this is true for $n = k+1$.

$$\begin{bmatrix} a_{11} & & a_{1,k+1} \\ 0 & a_{22} & \\ 0 & 0 & a_{k+1,k+1} \end{bmatrix} = \begin{bmatrix} a_{11} & h^T \\ 0 & A \end{bmatrix}$$

with A a $k \times k$ matrix. Now the inverse of the above is

$$\begin{bmatrix} a_{11}^{-1} & a_{11}^{-1}h^T A^{-1} \\ 0 & A^{-1} \end{bmatrix}$$

Now A^{-1} is an upper triangular by the induction hypothesis and therefore the matrix above is upper triangular.

Part (b): V is unit upper triangular then V^{-1} is a unit upper triangular. As the inverse of any triangular matrix (upper or lower) must have its diagonal elements to be the reciprocals of the corresponding elements in the original matrix i.e.

$$a_{ii}^{-1} = \frac{1}{a_{ii}} .$$

Thus if $a_{ii} = 1$ then $a_{ii}^{-1} = 1$.

Exercise 1.7.50

Part (a): Assume that there are at least two M that will satisfy the given expression. What is important is this expression is that the block matrices A_{11} and A_{12} don't change while the transformation introduces zeros below the block matrix A_{11}. As specified in the problem the matrix $\tilde{A}_{22}$ in each case can be different. This means we will assume that there exists matrices M_1 and M_2 such that

$$\begin{bmatrix} I_k & 0 \\ -M_1 & I_{n-k} \end{bmatrix} \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ 0 & \tilde{A}_{22}^{(1)} \end{bmatrix} ,$$

where we have indicated that $\tilde{A}_{22}$ may depend on the "M" matrix by providing it with a subscript. For the matrix M_2 we have a similar expression. Multiplying on the left by the inverse of this block matrix

$$\begin{bmatrix} I_k & 0 \\ -M_1 & I_{n-k} \end{bmatrix}$$

which is

$$\begin{bmatrix} I_k & 0 \\ M_1 & I_{n-k} \end{bmatrix}$$

(this inverse can be shown by direct multiplication of the two matrices), gives the following

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} = \begin{bmatrix} I_k & 0 \\ M_1 & I_{n-k} \end{bmatrix} \begin{bmatrix} A_{11} & A_{12} \\ 0 & \tilde{A}_{22}^{(1)} \end{bmatrix} = \begin{bmatrix} I_k & 0 \\ M_2 & I_{n-k} \end{bmatrix} \begin{bmatrix} A_{11} & A_{12} \\ 0 & \tilde{A}_{22}^{(2)} \end{bmatrix} .$$

Equating the $(2,1)$ component of the block multiplication above gives $M_1 A_{11} = M_2 A_{11}$, which implys that $M_1 = M_2$, since A_{11} is nonsingular. This shows the uniqueness of this block Gaussian factorization.

Returning to a single M, by multiplying the given factorization out we have

$$\begin{bmatrix} I_k & 0 \\ -M & I_{n-k} \end{bmatrix} \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ -MA_{11} + A_{21} & -MA_{12} + A_{22} \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ 0 & \tilde{A}_{22} \end{bmatrix} ,$$

so equating the $(2,1)$ block component of the above expression we see that $-MA_{11} + A_{21} = 0$, or $M = A_{21} A_{11}^{-1}$. In the same way equating the $(2,2)$ block components of the above gives

$$-A_{21} A_{11}^{-1} A_{12} + A_{22} ,$$

which is the Schur complement of A_{11} in A.

Part (b): By multiplying on the left by the block matrix inverse we have

$$\begin{bmatrix} I_k & 0 \\ M & I_{n-k} \end{bmatrix} \begin{bmatrix} I_k & 0 \\ -M & I_{n-k} \end{bmatrix} \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} = \begin{bmatrix} I_k & 0 \\ M & I_{n-k} \end{bmatrix} \begin{bmatrix} A_{11} & A_{12} \\ 0 & \tilde{A}_{22} \end{bmatrix},$$

or

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} = \begin{bmatrix} I_k & 0 \\ M & I_{n-k} \end{bmatrix} \begin{bmatrix} A_{11} & A_{12} \\ 0 & \tilde{A}_{22} \end{bmatrix},$$

as was to be shown.

Part (c): Taking the determinant of the above expression gives since $|A| \neq 0$, that

$$|A| = \left| \begin{bmatrix} I & 0 \\ M & I \end{bmatrix} \right| \left| \begin{bmatrix} A_{11} & A_{12} \\ 0 & \tilde{A}_{22} \end{bmatrix} \right| = 1|A_{11}||\tilde{A}_{22}|\,.$$

so $|\tilde{A}_{22}| \neq 0$, (or else $|A| = 0$, which is not true) and therefore since its determinant is nonzero, $\tilde{A}_{22}$ is nonsingular.

Exercise 1.7.54

Part (a): Let H be by symmetric and invertible then by the definition of the inverse, H^{-1} satisfies $HH^{-1} = I$. Now taking the transpose of both sides and remembering that the transpose of a product is the product of the transposes but in the opposite order we have $(H^{-1})^T H^T = I^T$ which simplifies to $(H^{-1})^T H = I$, since both H and I are symmetric. By multiplying both sides on the left by H^{-1} we have that

$$(H^{-1})^T = H^{-1}$$

showing that H^{-1} is symmetric.

Part (b): The Schur complement is given by $\hat{A}_{22} = A_{22} - A_{21}A_{11}^{-1}A_{12}$ and involves the submatrices in A. To determine properties of these submatrices consider the block definition of A given by

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}.$$

Taking the transpose of A (and using the fact that it is symmetric) we have

$$A^T = \begin{bmatrix} A_{11}^T & A_{21}^T \\ A_{12}^T & A_{22}^T \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

which gives (equating elements) the following

$$\begin{aligned} A_{11}^T &= A_{11} \\ A_{21}^T &= A_{12} \\ A_{12}^T &= A_{21} \\ A_{22}^T &= A_{22}\,. \end{aligned}$$

With these components we can compute the transpose of the Schur complement given by

$$\hat{A}_{22}^T = A_{22}^T - A_{12}^T(A_{11}^{-1})^T A_{21}^T = A_{22} - A_{21}A_{11}^{-1}A_{21} = \hat{A}_{22}\,,$$

showing that $\hat{A}_{22}$ is symmetric.

Exercise 1.8.1

Taking the block determinant of the given B we have

$$|B| = |B_{11}||B_{22}|\,.$$

Since B_{22} has a column of all zeros, it has a zero determinant. Thus B has a zero determinant and is singular.

Exercise 1.8.7 (the inverse of a permutation matrix)

The fact that $P^{-1} = P^T$ can be recognized by considering the product of P with P^T. When row i of P is multiplied by any column of P^T not equal to i the result will be zero since the location of the one's in each vector won't agree in the location of their index. However, when row i of P is multiplied by column i the result will be one. Thus we see by looking at the components of PP^T that $PP^T = I$ and P^T is the inverse of P.

Chapter 2 (Sensitivity of Linear Systems)

Exercise 2.1.10 (the one norm is a norm)

We have for the one norm the following definition

$$||x||_1 = \sum_{i=1}^{n} |x_i| \,.$$

We can check to see that each norm requirement is satisfied for this norm. We have

- The condition $||x||_1 \geq 0$ for all x and $||x||_1 = 0$ if and only if $x = 0$ can be seen to be certainly true.

Exercise 2.1.10

From the definition of the one norm $||x||_1 = \sum_{i=1}^{n} |x_i|$ we see that

$$||x||_1 \geq 0 \quad \text{and} \quad ||x||_1 = 0 \,,$$

can only be true if $x = 0$. We also see that

$$||\alpha x||_1 = \sum_{i=1}^{n} |\alpha x_i| = |\alpha| \sum_i |x_i| = |\alpha| ||x||_1 \,.$$

and that

$$||x + y||_1 = \sum_{i=1}^{n} |x_i + y_i| \leq \sum_i (|x_i| + |y_i|) = ||x||_1 + ||y||_1$$

Exercise 2.1.11

The distance between two points is the distance taking only right angles.

Exercise 2.1.13

For the infinity norm has

$$||x||_\infty \geq 0 \,,$$

and $||x||_\infty =$ if and only if $x = 0$. We also have

$$||\alpha x||_\infty = \min_{1 \leq i \leq n} |\alpha x_i| = |\alpha| \min_{1 \leq i \leq n} |x_i| = |\alpha| ||x||_\infty \,.$$

The third requirement of a norm is given by

$$\begin{aligned} ||x+y||_\infty &= \max_{1\le i\le n} |x_i + y_i| \\ &\le \max_{1\le i\le n} (|x_i| + |y_i|) \\ &\le \max_{1\le i\le n} |x_i| + \max_{1\le i\le n} |y_i| = ||x||_\infty + ||y||_\infty . \end{aligned}$$

Exercise 2.2.4

Part (a): Show that $\kappa(A) = \kappa(A^{-1})$, We know that

$$\kappa(A) = ||A||||A^{-1}|| .$$

so that

$$\kappa(A^{-1}) = ||A^{-1}||||A|| = ||A||||A^{-1}|| = \kappa(A) .$$

Part (b): We have

$$\kappa(cA) = ||cA||||(cA)^{-1}|| = |c|||A||\frac{1}{|c|}||A^{-1}|| = \kappa(A) .$$

Exercise 2.2.5

If $p = 1$ we have the set $\{x \in \mathbb{R}^2 | ||x||_1 = 1\}$ is equivalent to $\{x \in \mathbb{R}^2 | |x| + |y| = 1\}$. Which looks like If $p = 3/2$ we have the set $\{x \in \mathbb{R}^2 | (|x|^{3/2} + |y|^{3/2})^{2/3} = 1\}$ is equivalent to $\{x \in \mathbb{R}^2 | |x|^{3/2} + |y|^{3/2} = 1\}$ which looks like If $p = 2$ we have the set $\{x \in \mathbb{R}^2 | x^2 + y^2 = 1\}$, which looks like If $p = 3$ we have the set $\{x \in \mathbb{R}^2 | |x|^3 + |y|^3 = 1\}$, which looks like If $p = 10$ we have the set $\{x \in \mathbb{R}^2 | |x|^1 0 + |y|^1 0 = 1\}$, which looks like If $p = \infty$ we have the set $\{x \in \mathbb{R}^2 | \max_{1\le i\le n} |x_i| = 1\}$, which looks like

Exercise 2.2.6

Part (a): The condition number of A is defined by $\kappa(A) = ||A||||A^{-1}||$, while the condition number of A^{-1} is defined by $\kappa(A^{-1}) = ||A^{-1}||||(A^{-1})^{-1}|| = ||A^{-1}||||A|| = \kappa(A)$

Part (b): If c is any nonzero scalar then we have for $\kappa(cA)$ the following simplifications proving the desired relationship

$$\kappa(cA) = ||cA||\,||(cA)^{-1}|| = |c|\,||A||\,||\frac{1}{c}A^{-1}|| = |c|\,\frac{1}{|c|}||A||\,||A^{-1}|| = \kappa(A) .$$

Exercise 2.2.7

If $A = GG^T$ then **Part (a):** $||x||_A = (x^T Ax)^{1/2} = (x^T GG^T x)^{1/2} = ((G^T x)^T (G^T x))^{1/2} = ||G^T x||_2$.

Part (b): We know that $||G^T x||_2 \geq 0$, for all $G^T x$ which is the same as for all x. In addition, if $||G^T x||_2 = 0$ if and only if $G^T x = 0$, which since G^T is invertible is true if and only if $x = 0$.

Next we have

$$\begin{aligned} ||\alpha x||_A &= ||G^T(\alpha x)||_2 \\ &= |\alpha|||G^T x||_2 = |\alpha|||x||_A \,. \end{aligned}$$

Next we have

$$\begin{aligned} ||x+y||_A &= ||G^T(x+y)||_2 = ||G^T x + G^T y||_2 \\ &\leq ||G^T x||_2 + ||G^T y||_2 \\ &= ||x||_A + ||y||_A \,. \end{aligned}$$

Exercise 2.2.7

Let $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$, then $||A|| = 1$ and $||B|| = 1$, where

$$AB = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix} \quad \text{and} \quad ||AB|| = 2 \,.$$

But $||AB|| = 2 \neq ||A||||B|| = 1 \cdot 1 = 1$.

Exercise 2.2.8

Part (a): $||A(cx)|| = ||cAx|| = |c|||Ax||$, and $||cx|| = |c|||x||$, so that the ratio is unchanged by scalar multiplication.

Part (b):

$$||A|| = \max_{x \neq 0} \frac{||Ax||}{||x||} = \max_{x \neq 0} ||A\left(\frac{x}{||x||}\right)|| \,.$$

But as x runs through all $\mathbb{R}^n$ for $x \neq 0$, so that $\frac{x}{||x||}$ runs through $\mathbb{R}^n$ with $||x|| = 1$. Thus

$$||A|| = \max_{||x||=1} ||Ax|| \,.$$

Exercise 2.2.9

The Frobenius norm is defined by

$$||A||_F = \left(\sum_{i=1}^{n} \sum_{j=1}^{n} |a_{ij}|^2 \right)^{1/2},$$

and the two norm of A is given by

$$||A||_2 = \max_{x \neq 0} \frac{||Ax||_2}{||x||_2}.$$

Part (a): We have $||I||_F = (\sum_{i=1}^{n} 1^2)^{1/2} = n^{1/2}$, and

$$||I||_2 = \max_{x \neq 0} \frac{||Ix||_2}{||x||_2} = 1.$$

Part (b): $||A||_2^2 = \max_{||x||_2=1} ||Ax||_2^2 = \max_{||x||_2=1} \sum_{i=1}^{n} \left| \sum_{j=1}^{n} a_{ij} x_j \right|^2$, by Cauchy-Schwartz inequality

$$\left| \sum_{j=1}^{n} a_{ij} x_j \right|^2 \leq \sum_{j=1}^{n} a_{ij}^2 \sum_{j=1}^{n} x_j^2 .$$

As $||x|| = 1$ or $\left(\sum_{j=1}^{n} x_j^2 \right)^{1/2} = 1$.

Exercise 2.2.10

Remembering the definition of the infinity norm. We have $||A||_\infty = \max_{1 \leq i \leq n} \sum_{j=1}^{n} |a_{ij}|$, and consider the following

$$\begin{aligned} ||Ax||_\infty &= \max_{1 \leq i \leq n} |Ax|_i \leq \max_{1 \leq i \leq n} \sum_{j=1}^{n} |a_{ij}||x_j| \\ &\leq \max_{1 \leq i \leq n} \sum_{j=1}^{n} |a_{ij}| \left(\max_{1 \leq k \leq n} |x_k| \right) \\ &\leq ||x||_\infty \max_{1 \leq i \leq n} \sum_{j=1}^{n} |a_{ij}| . \end{aligned}$$

Therefore

$$\frac{||Ax||_\infty}{||x||_\infty} \leq \max_{1 \leq i \leq n} \sum_{j=1}^{n} |a_{ij}| .$$

Let $\max_{1\le i\le n}\sum_{j=1}^{n}|a_{ij}|$ occur at at the k-th row. Then

$$Ax_k = \sum_{j=1}^{n} a_{kj}x_j = \sum_{j=1}^{n} |a_{kj}|\,,$$

If the x_j is chosen so that as the vector x has components x_i, picked as

$$x_j \begin{cases} -1 & a_{kj} < 0 \\ +1 & a_{kj} > 0\,. \end{cases}$$

So $||x||_\infty = 1$, one sees that

$$\frac{||Ax||_\infty}{||x||} = \sum_{i=1}^{n} |a_{kj}| = \max_{1\le i\le n}\sum_{i=1}^{n} |a_{ij}|\,.$$

Exercise 2.3.2

We want to prove that

$$\mathrm{MaxMag}(A) = \max_{x\neq 0}\frac{||Ax||}{||x||}\,,$$

and that

$$\mathrm{MinMag}(A^{-1}) = \min_{x\neq 0}\frac{||A^{-1}x||}{||x||}\,,$$

We have that

$$\begin{aligned} \mathrm{maxmag}(A) &= \max_{x\neq 0}\frac{||Ax||}{||x||} = \max_{v\neq 0}\frac{||v||}{||A^{-1}v||} \\ &= \frac{1}{\min_{v\neq 0}\frac{||A^{-1}v||}{||v||}} \\ &= \frac{1}{\mathrm{minmax}(A^{-1})}\,. \end{aligned}$$

$Ax = v$, so $x = A^{-1}v$.

$$\begin{aligned} \mathrm{MinMag}(A^{-1}) &= \min_{x\neq 0}\frac{||A^{-1}x||}{||x||} = \min_{v\neq 0}\frac{||v||}{|Av|} \\ &= \frac{1}{\max_{v\neq 0}\frac{||Av||}{||v||}} \\ &= \frac{1}{\mathrm{maxmag}(A^{-1})}\,. \end{aligned}$$

Then

$$\frac{\mathrm{maxmag}(A)}{\mathrm{minmag}(A)} = ||A||\mathrm{maxmag}(A^{-1}) = ||A||||A^{-1}||\,,$$

from Exercise 2.3.2.

Exercise 2.3.4

With $A_\epsilon = \begin{bmatrix} \epsilon & 0 \\ 0 & \epsilon \end{bmatrix}$, so that

$$||A_\epsilon|| = \max_{x \neq 0} \frac{||A_\epsilon x||}{||x||} = \max_{(x_1,x_2)\neq 0} \frac{||(\epsilon x_1, \epsilon x_2)||}{||x||} = |\epsilon| \max_{x \neq 0} \frac{||x||}{||x||} = |\epsilon| = \epsilon$$

$A\epsilon^{-1} = \begin{bmatrix} 1/\epsilon & 0 \\ 0 & 1/\epsilon \end{bmatrix}$, so by the above $||A_\epsilon^{-1}|| = 1/\epsilon$, to therefore

$$\kappa(A_\epsilon) = \epsilon \left(\frac{1}{\epsilon} \right) = 1 \,.$$

an obviously $\det(A_\epsilon) = \epsilon^2$.

Exercise 2.3.5 (perturbing the right-hand-side b)

We want to prove that

$$\frac{||\delta b||}{||\delta||} \leq \kappa(A) \frac{||\delta x||}{||x||} \,.$$

We begin with $Ax = b$ which gives that $x = A^{-1}b$, so that $||x|| \leq ||A^{-1}||||b||$. By perturbing the solution to $Ax = b$, we get that δx relates to δb by considering

$$A(x + \delta x) = b + \delta b \Rightarrow A\delta x = \delta b \,.$$

Thus $||A||||\delta x|| \geq ||\delta b||$. So $||\delta x|| \geq \frac{||\delta b||}{||A||}$, or equivalently that

$$\frac{1}{||\delta x||} \leq \frac{||A||}{||\delta b||} \,.$$

Multiplying these two equations we get

$$\frac{||x||}{||\delta x||} \leq ||A^{-1}||||A|| \frac{||b||}{||\delta b||} = \kappa(A) \frac{||b||}{||\delta b||} \,.$$

Equality in this expression holds when $||x|| = ||A^{-1}||||b||$ or b is in in the direction of maximum magnification of A^{-1} *and* $||A||||\delta x|| = ||\delta b||$, so that δx is in the direction of the maximal magnification of A. Then we have that

$$\frac{||\delta b||}{||b||} \leq \kappa(A) \frac{||\delta x||}{||x||} \,.$$

Exercise 2.3.6 ($\frac{||\delta b||}{||b||} \leq \kappa(A)\frac{||\delta x||}{||x||}$ as a statement about residuals)

We have $r(\hat{x}) = A\hat{x} - b = A(x + \delta x) - b = A\delta x = \delta b$. So that

$$\frac{||r(\hat{x})||}{||b||} = \frac{||\delta b||}{||b||} \leq \kappa(A)\frac{||\delta x||}{||x||}$$

As the above inequality is sharp we can get $\frac{||r(\hat{x})||}{||b||} = \kappa(A)\frac{||\delta x||}{||x||}$, and thus the residual can be very much non-zero. If the matrix is well conditioned then $\kappa(A)$ is fairly small and by the above the residual must be small and we cannot get a very large $\frac{||r(\hat{x})||}{||b||}$ with a well conditioned matrix.

Watkins: Ex 2.3.10 (solving $Ax = b$ with perturbations in A and b)

We are told to consider $Ax = b$ and $(A+\delta A)(x+\delta x) = b+\delta b$. Expanding the left-hand-side of the second equation gives

$$Ax + A\delta x + \delta A(x + \delta x) = b + \delta b\,,$$

or since $Ax = b$ this is

$$A\delta x = \delta b - \delta A(x + \delta x)\,,$$

or solving for δx we get

$$\delta x = A^{-1}(\delta b - \delta A(x + \delta x))\,.$$

Taking vector norms on both sides we get

$$\begin{aligned} ||\delta x|| &\leq ||A^{-1}||\,(||\delta b|| + ||\delta A||(||x|| + ||\delta x||)) \\ &= ||A^{-1}||||A||\left(\frac{||\delta b||}{||A||} + \frac{||\delta A||}{||A||}(||x|| + ||\delta x||)\right)\,. \end{aligned}$$

Since $b = Ax$ we have $||b|| \leq ||A||||x||$ so $\frac{1}{||A||} \leq \frac{||x||}{||b||}$ and using the definition of the condition number $\kappa(A) \equiv ||A||||A^{-1}||$ we have

$$||\delta x|| \leq \kappa(A)\left(\frac{||\delta b||}{||b||}||x|| + \frac{||\delta A||}{||A||}||x||\right) + \kappa(A)\frac{||\delta A||}{||A||}||\delta x||\,,$$

or solving for $||\delta x||$ we get

$$\left(1 - \kappa(A)\frac{||\delta A||}{||A||}\right)||\delta x|| \leq \kappa(A)\left(\frac{||\delta b||}{||b||} + \frac{||\delta A||}{||A||}\right)||x||\,.$$

Since we are assuming that our matrix perturbation is small enough that the new matrix $A + \delta A$ is still invertible or $\frac{||\delta A||}{||A||} \leq \frac{1}{\kappa(A)}$, the left-hand-side has a leading coefficient that is positive and we can divide by it to get

$$\frac{||\delta x||}{||x||} \leq \frac{\kappa(A)\left(\frac{||\delta A||}{||A||} + \frac{||\delta b||}{||b||}\right)}{1 - \kappa(A)\frac{||\delta A||}{||A||}}\,, \tag{6}$$

the desired expression.

Chapter 4 (Eigenvalues and Eigenvectors I)

Exercise 4.3.21 (quartic convergence)

Exercise 4.3.22 (the number of digits of accuracy with linear convergence)

At iteration i denote our approximation error as $||q_j - v||_2 = 10^{-s_j}$ for some s_j. Since we are told that we have linear convergence we know that

$$||q_{j+1} - v||_2 \approx \frac{1}{r}||q_j - v||_2 = \frac{1}{r}10^{-s_j} = 10^{-s_j - \log(r)}\,,$$

or we see the number of correct digits increases by $-\log(r)$ on each iteration.

Exercise 4.4.1 (geometric multiplicity)

The geometric multiplicity is the dimension of the space $\{v \in \mathbb{C}^n | Av = \lambda v\}$. This space transforms as

$$\text{Dim}\{v \in \mathbb{C}^n | Av = \lambda v\} = \text{Dim}\{v \in \mathbb{C}^n | P^{-1}APv = \lambda v\} = \text{Dim}\{Pv \in \mathbb{C}^n | A(Pv) = \lambda (Pv)\}\,.$$

Thus the geometric multiplicity of the eigenvalue of B equals the geometric multiplicity of the corresponding eigenvalue of A.

Exercise 4.4.2 (geometric and algebraic multiplicity for simple matrices)

When A is simple it has n distinct eigenvalues and thus n linearly independent eigenvectors. Thus the geometric and the algebraic multiplicity of each eigenvalue is 1.

Exercise 4.4.3 (A^{-1} is similar to B^{-1})

If A is similar to B that means that there is an invertible transformation P such that $B = P^{-1}AP$. Thus taking the "inverse" of each side and assuming that all inverses exist we have

$$B^{-1} = P^{-1}A^{-1}P\,,$$

or that B^{-1} and A^{-1} are similar.

Exercise 4.4.4 (eigenvectors of a diagonal matrix)

The vectors that are all zeros except for a single 1 at one location are the eigenvectors of a diagonal matrix. These vectors are obviously linearly independent and a diagonal matrix is therefore simple.

Exercise 4.4.5

If A satisfies $V^{-1}AV = D$ with D diagonal and V nonsingular then we want to show that the columns of V are the n linearly independent eigenvectors of A and the elements of D are the eigenvalues. From $V^{-1}AV = D$ we have $AV = VD$. If we let v_i be the ith column of V, and λ_i the ith diagonal element of D then $AV = VD$ is the same as $Av_i = \lambda_i v_i$ for $i = 1, 2, \cdot n$ or that v_i is the ith eigenvector of A. That v_i and v_j are linearly independent follows from the fact that V is nonsingular. As A has n linearly independent eigenvectors A is simple.

Exercise 4.4.6 (norms of unitary matrices)

Part (a): As U is unitary we have

$$||U||_2 = \max_{||x|| \neq 0} \frac{||Ux||_2}{||x||_2} = \max_{||x|| \neq 0} \frac{||x||_2}{||x||_2} = 1 \,,$$

since $||Ux||_2 = ||x||_2$. As U is unitary so is U^{-1}, thus $||U^{-1}||_2 = 1$ also. Thus we have

$$\kappa(U) = ||U||_2 ||U^{-1}||_2 = 1 \cdot 1 = 1 \,.$$

Part (b): We are told that A and B are unitary similar thus $B = U^{-1}AU$. Thus

$$||B||_2 \leq ||U^{-1}||_2 ||A||_2 ||U||_2 = ||A||_2 \,,$$

by the consistency property of the matrix norm. Also from $A = UBU^{-1}$ we have $||A||_2 \leq ||B||_2$ so we have that $||A||_2 = ||B||_2$. In the same way we can show that $||A^{-1}||_2 = ||B^{-1}||_2$ so we have

$$\kappa_2(A) = ||A||_2 ||A^{-1}||_2 = ||B|| ||B^{-1}|| = \kappa_2(B)$$

Part (c): Given $B = U^*AU$ and consider $B + \delta B = U^*(A + \delta A)U$ or

$$B + \delta B = U^*AU + U^*\delta AU \,,$$

or

$$\delta B = U^*\delta AU \,.$$

Thus δA and δB are unitary similar and $||\delta B||_2 = ||\delta A||_2$.

Exercise 4.4.7 (round-off errors in arbitrary similarity transforms)

From the given P we have $P^{-1} = \begin{bmatrix} 1 & -\alpha \\ 0 & 1 \end{bmatrix}$.

Part (a): With $A = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$ so $B = P^{-1}AP = \begin{bmatrix} 2 & \alpha \\ 0 & 1 \end{bmatrix}$ and

$$||B||_\infty = \max(2 + |\alpha|, 1) = 2 + |\alpha| \,,$$

if α is large enough.

Part (b): $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and $\delta A = \frac{\epsilon}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$. Thus we have

$$||A||_\infty = 1$$
$$||\delta A||_\infty = \frac{\epsilon}{2}(2) = \epsilon \,,$$

thus

$$\frac{||\delta A||_\infty}{||A||_\infty} = \epsilon \,.$$

From the matrices above we have $B = P^{-1}AP = P^{-1}P = I$, and

$$\begin{aligned} B + \delta B = P^{-1}(A + \delta A)P &= \begin{bmatrix} 1 & -\alpha \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1+\frac{\epsilon}{2} & \frac{\epsilon}{2} \\ \frac{\epsilon}{2} & 1+\frac{\epsilon}{2} \end{bmatrix}\begin{bmatrix} 1 & \alpha \\ 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & -\alpha \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1+\frac{\epsilon}{2} & \alpha(1+\frac{\epsilon}{2}) + \frac{\epsilon}{2} \\ \frac{\epsilon}{2} & \alpha\frac{\epsilon}{2} + 1 + \frac{\epsilon}{2} \end{bmatrix} \\ &= \begin{bmatrix} 1+\frac{\epsilon}{2} - \alpha\frac{\epsilon}{2} & \alpha\left(1+\frac{\epsilon}{2}\right) + \frac{\epsilon}{2} - \alpha - \alpha\frac{\epsilon}{2} - \alpha\frac{\epsilon}{2} \\ \frac{\epsilon}{2} & \alpha\frac{\epsilon}{2} \end{bmatrix} = \frac{\epsilon}{2}\begin{bmatrix} 1-\alpha & \alpha - \alpha^2 \\ 1 & 1+\alpha \end{bmatrix} . \end{aligned}$$

Thus $||B||_\infty = 1$ and

$$||\delta B||_\infty = \frac{\epsilon}{2}\max(|1-\alpha| + |\alpha - \alpha^2|, 1 + |1+\alpha|) \,,$$

which can be as large as desired by selecting α large. When α is large we have this expression $O(\alpha^2)$.

Exercise 4.4.8 (conditioning of similarity transformations)

Part (a): Writing $B = P^{-1}AP$ we have

$$||B|| \le ||P^{-1}||||A||||P|| = \kappa(P)||A|| \,.$$

Thus $||B|| \le \kappa(P)||A||$. Therefore $\frac{1}{\kappa(P)}||A|| \le ||B||$. Combining these two we get

$$\frac{1}{\kappa(P)}||A|| \le ||B|| \le \kappa(P)||A||$$

Part (b): From $B = P^{-1}AP$ and $B + \delta B = P^{-1}(A + \delta A)P$ we have that

$$\delta B = P^{-1}\delta A P\,.$$

Thus as in Part (a) we can show that

$$\frac{1}{\kappa(P)}||\delta A|| \leq ||\delta B|| \leq \kappa(P)||\delta A||\,.$$

Using $||B|| \geq \frac{1}{\kappa(P)}||A||$ and $||\delta B|| \leq \kappa(P)||\delta A||$ we have that

$$\frac{||\delta B||}{||B||} \leq \frac{\kappa(P)^2||\delta A||^2}{||A||}\,.$$

Using $||B|| \leq \kappa(P)||A||$ and $||\delta B|| \geq \frac{1}{\kappa(P)}||\delta A||$ we have that

$$\frac{||\delta B||}{||B||} \geq \frac{||\delta A||^2}{\kappa(P)^2||A||}\,.$$

Exercise 4.4.9 (arbitrary similarity transformation don't preserve $A = A^*$)

For this just pick any invertible P such that $P^* \neq P^{-1}$ like $P = \begin{bmatrix} 1 & i \\ 0 & 1 \end{bmatrix}$. Then consider a matrix A that is Hermitian or satisfies $A = A^*$ like $A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$. From the given P we have that $P^{-1} = \begin{bmatrix} 1 & -i \\ 0 & 1 \end{bmatrix}$ and we find

$$\begin{aligned} B = P^{-1}AP &= \begin{bmatrix} 1 & -i \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}\begin{bmatrix} 1 & i \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & -i \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & i+2 \\ 2 & 2i+1 \end{bmatrix} \\ &= \begin{bmatrix} 1-2i & i+2+2-i \\ 2 & 2i+1 \end{bmatrix} = \begin{bmatrix} 1-2i & 5 \\ 2 & 2i+1 \end{bmatrix}. \end{aligned}$$

Note that $B^* = \begin{bmatrix} 1+2i & 2 \\ 5 & -2i+1 \end{bmatrix} \neq B$.

Exercise 4.4.10 (observations from the proof of Schur's theorem)

Part (a): We can specify the ordering of the eigenvalues in T in any order since all that was required in the proof of Schur's theorem was to initially pick an eigenvector v of A. Since the eigenvalue associated with this eigenvector appears at the $(1,1)$ location in the matrix T. Since v was arbitrary wen can pick any eigenvector and correspondingly any eigenvalue. This argument can then be repeated recursively.

Part (b): We start the proof of Schur's theorem by picking an eigenvector v (of A) that vector then goes in the matrix U_1. Our total unitary transformation U is $U_1 U_2$ with

$$U_2 = \left[\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & \hat{U}_2 & \\ 0 & & & \end{array}\right] .$$

Note that

$$U = U_1 U_2 = \left[\begin{array}{cc} v & W \end{array}\right] \left[\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & \hat{U}_2 & \\ 0 & & & \end{array}\right] = \left[\begin{array}{cc} v & W\hat{U}_2 \end{array}\right] ,$$

Thus the first column of U is the eigenvector of A with the eigenvalue λ. Since λ (and the corresponding eigenvector v) was chosen arbitrarily we can extract any arbitrary ordering of eigenvectors values.

Exercise 4.4.13 (the Rayleigh quotient can give approximate eigenvalues)

Part (a): From $q = \sum_{i=1}^{n} c_i v_i$ we have

$$||q||_2^2 = q^* q = \left(\sum_{i=1}^{n} c_i^* v_i^* \right) \left(\sum_{j=1}^{n} c_j v_j \right) = \sum_{i=1}^{n} |c_i|^2 .$$

Since $q_i^* q_j = \delta_{ij}$ the Kronecker delta.

Part (b): Note that

$$v_1 - q = (1 - c_1) v_1 - \sum_{j=2}^{n} c_j v_j .$$

Thus using the same arguments as in the first part of this problem we have

$$||v_1 - q||_2^2 = |1 - c_1|^2 + \sum_{j=2}^{n} |c_j|^2 .$$

Thus dropping the positive term $|1 - c_1|^2$ we have

$$||v_1 - q||_2^2 \leq \sum_{j=2}^{n} |c_j|^2 ,$$

as we were to show.

Part (c): Note the Rayleigh quotient ρ is $\rho = q^* A q$ when the norm of q is one. If we expand q in terms of the eigenvectors of A we have $q = \sum_{i=1}^{n} c_i v_j$ so that $q^* = \sum_{i=1}^{n} c_i^* v_j^*$ and

$$Aq = \sum_{i=1}^{n} c_i \lambda_i v_i ,$$

giving

$$q^* A q = \left(\sum_{i=1}^n c_i^* v_j^* \right) \left(\sum_{i=1}^n c_i \lambda_i v_i \right) = \sum_{i=1}^n |c_i|^2 \lambda_i \,,$$

using the property of the orthonormal eigenvectors.

Part (d): Since $\sum_{i=1}^n |c_i|^2 = 1$ by Part (a) we can write

$$\lambda_1 - \rho = \sum_{i=1}^n \lambda_1 |c_i|^2 - \sum_{i=1}^n \lambda_i |c_i|^2 = \sum_{i=2}^n (\lambda_1 - \lambda_i)|c_i|^2 \,.$$

Using this we have

$$|\lambda_1 - \rho| \leq C \sum_{i=2}^n |c_i|^2 \leq C ||v_1 - q||_2^2 \,,$$

using Part (b). Thus the take away from this is that if we pick a vector q that is close to the eigenvector v_1 (say in norm by $O(\epsilon)$) then the Rayleigh quotient will be close to the eigenvalue (by an amount $O(\epsilon^2)$).

Exercise 4.4.14 (eigenvalues of Hermitian matrices are real)

Part (a): Note that the conjugate of the expression $x^* A x$ is itself again

$$(x^* A x)^* = x^* A^* x = x^* A x \,.$$

Only real numbers are their own complex conjugates.

Part (b): Since the numerator $x^* A x$ and the denominator $x^* x$ in the Rayleigh quotient are real the Rayleigh quotient itself must be real.

Part (c): As the Rayleigh quotients when computed using eigenvectors return eigenvalues the eigenvalues of Hermitian matrices must be real.

Exercise 4.4.15 (unitary similarity and positive definiteness)

Assume that B is unitarily similar to A and consider for an arbitrary vector x the inner product

$$x^* B x = x^* U^* A U x = (Ux)^* A (Ux) > 0$$

as A is positive definite. Thus B is positive definite.

Exercise 4.4.16 (eigenvalues of positive definite matrices are positive)

Consider the Rayleigh quotient for a positive definite matrix A or $\frac{x^* X x}{x^* x}$. Since we know that the numerator and the denominator are positive real numbers the fraction is also. Thus the eigenvalues of the positive definite matrix A must be positive.

Exercise 4.4.17 (positive definite is equivalent to positive eigenvalues)

We have shown that if A is positive definite it has positive eigenvalues. Assume that A has positive eigenvalues, then from the Spectral Theorem we can write A as $A = UDU^*$ with U unitary, and D a diagonal matrix with the positive eigenvalue of A as the diagonal elements. In that case note that

$$x^* A x = x^* U D U^* x = (U^* x)^* D (U^* x) > 0\,,$$

since the last expression is the sum of positive terms.

Exercise 4.4.18 (positive semidefinite matrices)

For results with positive semidefinite matrices all signs like $>$ become $\geq$ but all of the arguments are the same.

Exercise 4.4.19 (skew-Hermitian matrices)

Part (a): Consider $B^* = (U^* A U)^* = U^* A^* U = -U^* A U = -B$ showing that B is skew-Hermitian.

Part (b): When A is skew-Hermitian then Schur's Theorem states $T = U^* A U$ so we have that T and A are unitary similar. Thus T must be skew-Hermitian and $T^* = -T$ this means that T is a diagonal matrix with imaginary elements on the diagonal.

Part (c): Besides the argument above we can consider the Rayleigh quotient. Note that taking its conjugate we have

$$\rho^* = \frac{x^* A^* x}{x^* x} = -\frac{x^* A x}{x^* x} = -\rho\,.$$

The only numbers that are negative conjugates of them selves are pure imaginary numbers, showing that the eigenvalues of A must be pure imaginary.

Exercise 4.4.20 (unitary and unitary similarity)

Part (a): If $B = U^*AU$ then

$$BB^* = U^*AU(U^*A^*U) = U^*AA^*U = I$$

therefor $B^{-1} = B$ and B is unitary.

Part (b): Let T be an upper triangular matrix that is unitary. Then because T us unitary we know that $T^{-1} = T^*$. Since inverses of upper triangular matrices are also upper triangular we have that T^{-1} is upper triangular. At the same time T^* is lower triangular. In order that these two things equal each other we must have that they are both diagonal matrices. As T^* is a diagonal matrix so is T.

Part (c): Schur's Theorem states that there exists a unitary matrix U such that $U^*AU = T$ with T upper triangular with the eigenvalues of A on the diagonal. Since A is unitary and T is unitarily similar to A, by Part (a) we know that T is unitary. Since T is unitary and upper triangular we know by Part (b) that T must be a diagonal matrix.

Part (d): From the above discussion we have that $U^*AU = T$ where T is a unitary diagonal matrix and hold the eigenvalues of A. Since T is unitary $T^{-1} = T^*$ and writing this equation in terms of an arbitrary element of the diagonal of T (say ρ) we have $\frac{1}{\rho} = \rho^*$ or $|\rho|^2 = 1$. Thus the eigenvalues of A lie on the unit circle.

Another way to see this is to consider the Rayleigh quotient ρ for an eigenvector x with eigenvalue ρ. Since A is unitary and x is the eigenvector we have that

$$\rho^* = \frac{x^*A^*x}{x^*x} = \frac{(x^*A^{-1}x}{x^*x} = \frac{x^*\left(\frac{1}{\rho}x\right)}{x^*x} = \frac{1}{\rho}\,.$$

Again we have $\rho\rho^* = 1$ or that $|\rho|^2 = 1$.

Exercise 4.4.21 (normal matrices)

Part (a): We will show this for just the case where A is Hermitian since all the other cases are the same. In that case $A^* = A$ so that $AA^* = AA = A^2$. The other product, A^*A, is equal to A^2 also.

Part (b): If $B = U^*AU$ then we have

$$BB^* = U^*AU(U^*A^*U) = U^*AA^*U = U^*A^*AU\,,$$

since A is normal. Continuing we have

$$U^*A^*AU = U^*A^*UU^*AU = B^*B\,,$$

showing that B is normal.

Part (c): Partition the original $n \times n$ upper triangular matrix T in the following way

$$T = \left[\begin{array}{c|c} t_{11} & t_r \\ \hline 0 & \hat{T} \end{array}\right] .$$

Here $\hat{T}$ is a $n-1 \times n-1$ upper triangular matrix, t_{11} is a scalar, and t_r is a *row* vector. Then with this we have

$$T^* = \left[\begin{array}{c|c} t_{11}^* & 0 \\ \hline {t_r}^* & \hat{T}^* \end{array}\right] .$$

The normal products TT^* and T^*T are

$$TT^* = \left[\begin{array}{c|c} t_{11} & t_r \\ \hline 0 & \hat{T} \end{array}\right] \left[\begin{array}{c|c} t_{11}^* & 0 \\ \hline {t_r}^* & \hat{T}^* \end{array}\right] = \left[\begin{array}{c|c} |t_{11}|^2 + t_r {t_r}^* & t_r \hat{T}^* \\ \hline \hat{T} {t_r}^* & \hat{T}\hat{T}^* \end{array}\right]$$

$$T^*T = \left[\begin{array}{c|c} t_{11}^* & 0 \\ \hline {t_r}^* & \hat{T}^* \end{array}\right] \left[\begin{array}{c|c} t_{11} & t_r \\ \hline 0 & \hat{T} \end{array}\right] = \left[\begin{array}{c|c} |t_{11}|^2 & {t_{11}}^* t_r \\ \hline t_{11} {t_r}^* & \hat{T}^*\hat{T} \end{array}\right]$$

Since we assume that $TT^* = T^*T$ equating the components of the above we get

$$\begin{aligned} |t_{11}|^2 + t_r t_r^* &= |t_{11}|^2 \\ t_r \hat{T}^* &= {t_{11}}^* t_r \\ \hat{T} t_r^* &= t_{11} t_r^* \\ \hat{T}\hat{T}^* &= \hat{T}^*\hat{T} . \end{aligned}$$

By induction $\hat{T}$ is normal, upper triangular and therefore diagonal. From the $(1,1)$ element or the first equation above we must have

$$t_r t_r^* = 0 .$$

But $t_r t_r^*$ is the sum of the normed squared of the elements of t_r and this is only equal to zero if each element is. Thus $t_r = 0$. This solution makes the other two equations satisfied.

Part (d): Let A be a diagonal matrix then we have

$$AA^* = \begin{bmatrix} |a_{11}|^2 & & & 0 \\ & |a_{22}|^2 & & \\ & & \ddots & \\ 0 & & & |a_{nn}|^2 \end{bmatrix} ,$$

and A^*A is the same so A is normal.

Part (e): Assume that A is unitary similar to a diagonal matrix say D. Since D is normal by Part (d) and every matrix unitary similar to a normal matrix is normal we have that A is normal.

Assume that A is normal thus $AA^* = A^*A$. By Schur's Theorem there exists a unitary matrix U such that $U^*AU = T$ where T upper triangular. Now A is normal and thus T is normal by Part (b). By Part (c) when T is upper triangular and normal we have that T is diagonal. Thus A is unitary similar to a diagonal matrix.

Exercise 4.4.22 (A has n orthonormal eigenvectors then A is normal)

Assume that A has n orthonormal eigenvectors. Form the matrix U with these n eigenvector as columns. Then $AU = UD$ since the eigenvectors are orthonormal we can create these such that $D = U^*AU$ or A and D are unitary similar. As a diagonal matrix is normal then by Exercise 4.4.21 Part (b) A is also normal.

Exercise 4.4.23 (properties of diagonal matrices)

Part (a): D is Hermitian so that $D^* = D$. In components this means that $d_{ii}^* = d_{ii}$ or the diagonal elements must be real. These are also the eigenvalues of the matrix D.

Part (b): Let e_i be a vector of all zeros with a 1 in the ith location. Then $e_i^* D e_i = d_{ii} \geq 0$ so D is positive semi-definite.

Part (c): This is the same as the previous part where we have $e_i^* D e_i = d_{ii} > 0$.

Part (d): For $D^* = -D$ or in component form we have $d_{ii}^* = -d_{ii}$ or that d_{ii} is imaginary thus the eigenvalues of D are imaginary.

Part (e): The matrix D is such that $D^* = D^{-1}$ in component form we have $d_{ii}^* = \frac{1}{d_{ii}}$ or $|d_{ii}|^2 = 1$.

Exercise 4.4.24 (properties of normal matrices)

Part (a): A is a normal matrix then its unitary similar to the diagonal matrix so that $U^*AU = D$ thus $D^* = U^*A^*U = U^*AU = D$ so D is Hermitian. By Exercise 4.4.23 (a) the eigenvalues of D are real. As A is similar to D it has the same (real) eigenvalues.

Part (b-e): Following the steps as in Part (a) to show A unitary similar to a diagonal matrix, that has the same properties as A. Then we use Exercise 4.4.23 to show properties of the eigenvalues of D. Then A has the same eigenvalues as D (due to the fact that they are similarity related) and thus the eigenvalues of A satisfy the desired properties.

Exercise 4.4.25 (Rayleigh quotient of normal matrices approximate eigenvalues)

Since the book has shown that every normal matrix has n orthonormal eigenvectors (the main requirement needed for Exercise 4.4.13) we can follow the steps in that proof to show the required steps.

Exercise 4.4.26 (defective matrices and a nearby simple matrix)

Let A be defective. By Schur's Theorem $A = U^* T U$ with T upper triangular with diagonal elements of T the eigenvalues of A. Consider a matrix A_ϵ created from another (as yet) unknown upper triangular matrix T_ϵ as $A_\epsilon = U^* T_\epsilon U$. The matrix U is the same Hermitian matrix created in Schur's Theorem. Then we have

$$||A - A_\epsilon||_2 = ||U^*(T - T_\epsilon)U||_2 = ||T - T_\epsilon||_2 .$$

We now pick T_ϵ such that it equals T for all off diagonal elements. Then the matrix $T - T_\epsilon$ is a diagonal matrix say D. The two norm of this diagonal matrix is the square root of the largest eigenvector of $D^* D$. This is the same as the absolute value of the diagonal elements. Thus we have

$$||T - T_\epsilon||_2 = \max_i(|T_{ii} - T_{\epsilon ii}|) .$$

We can now select the elements to use as the diagonal elements of T_ϵ in such a way that they are all distinct and that the maximum above is as small as possible. Thus if we pick T_ϵ to be as above T_ϵ will be simple and so will A_ϵ and A_ϵ has the requested distance to the matrix A.

Exercise 4.4.28 (properties of skew-symmetric matrices)

Part (a): Assume that A is n by n skew-symmetric. As the matrix A is real the characteristic equation $\det(A - \lambda I)$ is a nth order *real* polynomial and any nonzero complex roots that it might have must come in complex conjugate pairs. Since A is skew-symmetric (also skew-Hermitian) it must have its eigenvalues on the imaginary axis. Since n is odd it cannot have all of its roots be complex since they must come in pairs and we would need to have an even polynomial order. Thus $\lambda = 0$ must be a root of the characteristic equation and our matrix A must be singular.

Part (b): These matrices look like $\begin{bmatrix} 0 & b \\ 0 & -b \end{bmatrix}$ is a 2 by 2 skew-symmetric matrix.

Part (c): T will be block diagonal with each block a 2×2 block like $\begin{bmatrix} 0 & b \\ -b & 0 \end{bmatrix}$ or a 1×1 block of zero.

Exercise 4.4.29 (properties of the trace function)

Part (a): The trace function is a linear operator and this is just an expression for that fact.

Part (b): We want to show that $\text{trace}(CD) = \text{trace}(DC)$ which we do by considering the

left-hand-side and showing that it equals the right-hand-side. We have

$$\begin{aligned}\text{trace}(CD) &= \sum_{i=1}^{n}(CD)_{ii} = \sum_{i=1}^{n}\sum_{k=1}^{n} C_{ik}D_{ki} \\ &= \sum_{k=1}^{n}\sum_{i=1}^{n} C_{ik}D_{ki} = \sum_{k=1}^{n}(DC)_{kk} \\ &= \text{trace}(DC)\,.\end{aligned}$$

Part (c): Note that $||B||_F^2 = \sum_{i=1}^{n}\sum_{j=1}^{n}|b_{ij}|^2$. Now consider the (i,j)th element of B^*B given by

$$(B^*B)_{ij} = \sum_{k=1}^{n}(B^*)_{ik}B_{kj} = \sum_{k=1}^{n}(B_{ki})^*B_{kj}\,.$$

Thus

$$\text{trace}(B^*B) = \sum_{i=1}^{n}\sum_{k=1}^{n}(B_{ki})^*B_{ki} = \sum_{i=1}^{n}\sum_{k=1}^{n}|B_{ki}|^2 = ||B||_F^2\,.$$

Using the result from Part (b) we have $\text{trace}(B^*B) = \text{trace}(BB^*)$ and thus this also equals $||B||_F$.

Exercise 4.4.30 (normal and block triangular)

Part (a): We assume that T is normal thus $TT^* = T^*T$. When we write out the left-hand-side and right-hand-side of this expression we get

$$\begin{aligned}TT^* &= \begin{bmatrix} T_{11} & T_{12} \\ 0 & T_{22}\end{bmatrix}\begin{bmatrix} T_{11}^* & 0 \\ T_{12}^* & T_{22}^*\end{bmatrix} = \begin{bmatrix} T_{11}T_{11}^* + T_{12}T_{12}^* & T_{12}T_{22}^* \\ T_{22}T_{12}^* & T_{22}T_{22}^*\end{bmatrix} \\ T^*T &= \begin{bmatrix} T_{11}^* & 0 \\ T_{12}^* & T_{22}^*\end{bmatrix}\begin{bmatrix} T_{11} & T_{12} \\ 0 & T_{22}\end{bmatrix} = \begin{bmatrix} T_{11}T_{11}^* & T_{11}^*T_{12} \\ T_{12}^*T_{11} & T_{12}T_{12}^* + T_{22}^*T_{12}\end{bmatrix}\,.\end{aligned}$$

Since T is normal we have the above two expressions are equal. If we consider the $(1,1)$ component of the above we get that we must have

$$T_{11}T_{11}^* + T_{12}T_{12}^* = T_{11}^*T_{11}\,.$$

Canceling the common terms and taking the trace of the remaining expression we get

$$\text{trace}(T_{12}T_{12}^*) = 0\,.$$

Using the trace is equal to the Frobenius norm result proven earlier we have that the above is equal to

$$||T_{12}||_F^2 = \text{trace}(T_{12}T_{12}^*) = 0\,,$$

or

$$\sum_i\sum_j |(T_{12})_{ij}|^2 = 0\,,$$

which means that echo component of T_{12} must be zero or that the matrix T_{12} is zero. This means that T is block diagonal as we were to show.

Exercise 4.4.31 (2×2 normal matrices)

Part (a): We find

$$AA^* = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} a & c \\ b & d \end{bmatrix} = \begin{bmatrix} a^2+b^2 & ac+bd \\ ac+db & c^2+d^2 \end{bmatrix}$$
$$A^*A = \begin{bmatrix} a & c \\ b & d \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a^2+c^2 & ab+cd \\ ab+cd & b^2+d^2 \end{bmatrix}.$$

To be normal when we equating the $(1,1)$ component of $AA^* = A^*A$ we get

$$a^2+b^2 = a^2+c^2 \quad \Rightarrow \quad b^2 = c^2\,,$$

which means that $b = \pm c$. Equating the $(1,2)$ component of $AA^* = A^*A$ we get

$$ac+bd = ab+cd\,.$$

If we have $b = c$ then this last equation is (changing all c's to b's) gives

$$ab+bd = ab+cd \quad \text{or} \quad b = c\,,$$

the same condition as before. Thus in this case we get $A = \begin{bmatrix} a & b \\ b & d \end{bmatrix}$ and we have a symmetric matrix. If $b = -c$ then the second equation above becomes (changing all c's into $-b$'s) gives

$$-ab+bd = ab-bd \quad \text{or} \quad -a+d = a-d \quad \text{or} \quad a = d\,.$$

Thus we have $A = \begin{bmatrix} a & b \\ -b & a \end{bmatrix}$.

Part (b): For A of the given form the eigenvalues are given by $|A-\lambda I| = 0$ or

$$\begin{vmatrix} a-\lambda & b \\ -b & a-\lambda \end{vmatrix} = 0\,.$$

This gives $(a-\lambda)^2+b^2 = 0$ or since b is real we get $\lambda = a \pm bi$.

Part (c): By Ex. 4.4.30 T is block diagonal. The elements of the blocks are either scalars for the real eigenvalues or 2×2 blocks of the form $\begin{bmatrix} a & b \\ -b & a \end{bmatrix}$ for the complex eigenvalues.

Exercise 4.4.32 (more properies of the trace)

Part (a): A and B are similar so $B = S^{-1}AS$. Then

$$\text{trace}(B) = \text{trace}(S^{-1}AS) = \text{trace}(ASS^{-1}) = \text{trace}(A)\,.$$

Part (b): By Schur's Theorem for all matrices A there exists an unitary matrix U such that $U^*AU = T$ with T upper triangular with the eigenvalues of A on the diagonal of T. Then using that result we have

$$\text{trace}(U^*AU) = \text{trace}(A) = \text{trace}(T) = \sum_{i=1}^{n} \lambda_i\,.$$

Exercise 4.4.33 (the determinant of a matrix)

By the complex Schur's Theorem for all matrices A there exists an unitary matrix U such that $U^*AU = T$ with T upper triangular with the eigenvalues of A on the diagonal of T. Then we have since $U^* = U^{-1}$ that

$$|U^{-1}AU| = |T| = \prod_{i=1}^{n} \lambda_i \, .$$

The left-hand-side is $|U|^{-1}|A||U| = |A|$ and we get

$$|A| = \prod_{i=1}^{n} \lambda_i \, ,$$

as we were to show.