In this homework, we work on the basic concepts of convex optimization and linear algebra. You are free to use software tools to facilitate the calculation. However, you are also encouraged to go through the derivations manually wherever it is feasible because these concepts are fundamental and needed for the study of convex optimization.

All the problems are graded by completion.

1. Convex Optimization (10 pts)
1.1. Given a function \( f_0(x) = x^4 - 2x^3 + 3x^2 - 2x + 4 \), where \( x \in \mathbb{R} \). Solve \( \min_x f_0(x) \) using Kuhn-Tucker conditions. Show your derivation. (2 pts)

Solution. The KT conditions are:
\[
\nabla f_0(x^*) = 0 \\
\nabla^2 f_0(x^*) \geq 0
\]
Solving for \( \nabla f(x^*) = 4(x^*)^3 - 6(x^*)^2 + 2x^* - 2 = 0 \), we get \( x^* = 0.5 \) as a potential local minimizer.

To confirm this, note that \( \nabla^2 f_0(x) = 12(x^2) - 12(x^* + 6 = 3 \geq 0 \). In fact, \( \nabla^2 f_0(x) = 12x^2 - 12x + 6 \geq 0 \) for all \( x \in \mathbb{R} \), meaning that \( f_0 \) is convex and \( x^* = 0.5 \) achieves the globally minimum value given by \( f(x^*) = 0.5^4 - 2(0.5)^3 + 3(0.5)^2 - 2(0.5) + 4 = 3.5625 \).

1.2. Given two functions \( f_0(x) = x^2 - 4x + 4 \), and \( f_1(x) = x + 3 \), where \( x \in \mathbb{R} \). Solve \( \min_x f_0(x) \) subject to \( f_1(x) \leq 0 \). (8 pts)

Solution. The Lagrangian is \( L(x, \lambda) = f_0(x) + \lambda f_1(x) = x^2 - 4x + 4 + \lambda(x + 3) \), where \( \lambda \in \mathbb{R} \geq 0 \) is a Lagrange multiplier. The primal problem is \( \min_{x \in \mathbb{R}} \max_{\lambda \in \mathbb{R} \geq 0} L(x, \lambda) \) and the dual is \( \max_{\lambda \in \mathbb{R} \geq 0} \min_{x \in \mathbb{R}} L(x, \lambda) = \max_{\lambda \in \mathbb{R} \geq 0} g(\lambda) \). Since the primal is harder to solve directly and strong duality holds for this problem, we will focus on solving the dual formulation. To do so, we first solve \( \min_{x \in \mathbb{R}} L(x, \lambda) \) using the KT conditions:
\[
\frac{\partial L}{\partial x} = 2x - 4 + \lambda \\
\frac{\partial^2 L}{\partial x^2} = 2 \geq 0
\]
By setting \( \frac{\partial L}{\partial x} = 0 \), we get that \( x^* = 2 - 0.5\lambda \) is the optimal choice to minimize \( L(x, \lambda) \) for any given choice of \( \lambda \). Substituting this result gives us \( g(\lambda) = \min_{x \in \mathbb{R}} L(x, \lambda) = 5\lambda - 0.25\lambda^2 \).

We then solve \( \max_{\lambda \in \mathbb{R} \geq 0} g(\lambda) \) again using the KT conditions:
\[
\frac{\partial g}{\partial \lambda} = 5 - 0.5\lambda
\]
\[ \frac{\partial^2 g}{\partial \lambda^2} = -0.5 \leq 0 \]

Solving \( \frac{\partial g}{\partial \lambda} = 5 - 0.5\lambda = 0 \) yields \( \lambda^* = 10 \), which in turn means that \( x^* = 2 - 0.5(10) = -3 \) and \( f_0(x^*) = (-3)^2 - 4(3) + 4 = 25 \).

2. Matrix Properties (16 pts)

2.1. Linear System:
Consider the following system of linear equations
\[
\begin{align*}
x_1 + 2x_2 + x_3 &= 1 \\
x_2 + x_3 &= -2 \\
x_1 + x_2 &= -1.
\end{align*}
\]
Write the equations in a matrix form. (2 pts)

Solution. \( A \vec{x} = \vec{b} \), where
\[
A = \begin{bmatrix}
1 & 2 & 1 \\
0 & 1 & 1 \\
1 & 1 & 0
\end{bmatrix},
\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \text{ and } \vec{b} = \begin{bmatrix} 1 \\ -2 \\ -1 \end{bmatrix}.
\]

2.2. For the matrix in problem 2.1, derive its range. What’s the rank of this matrix? (2pts)

Solution. To find the range of matrix \( A \), perform Gaussian Elimination to convert \( A \) to its reduced echelon form \( \tilde{A} \):
\[
\begin{bmatrix}
1 & 2 & 1 \\
0 & 1 & 1 \\
1 & 1 & 0
\end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}
\]
The columns of \( A \) corresponding to the pivot columns of \( \tilde{A} \) form a basis for \( A \)'s range, giving us:
\[
\text{range}(A) = \left\{ x \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + y \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} \mid x,y \in \mathbb{R} \right\}
\]
Since \( \text{range}(A) \) has 2 vectors in its basis, \( \text{rank}(A) = 2 \).

2.3. Derive the nullspace of the matrix in problem 2.1. What’s the relation between the range and nullspace of a matrix? (2pts)

Solution. The nullspace of \( A \) consists of the set of solutions to the homogeneous equation \( A \vec{x} = \vec{0} \), which we can find by augmenting \( A \) with a column of zeros and row reducing:
\[
\begin{bmatrix}
1 & 2 & 1 & 0 \\
0 & 1 & 1 & 0 \\
1 & 1 & 0 & 0
\end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}
\]
Translating the reduced echelon matrix back to a system of equations, we have

\[
\begin{align*}
  x_1 - x_3 &= 0 \\
  x_2 + x_3 &= 0
\end{align*}
\]

equivalent to

\[
\begin{align*}
  x_1 &= x_3 \\
  x_2 &= -x_3
\end{align*}
\]

Thus, any \( \vec{x} \) satisfying \( A\vec{x} = \vec{0} \), must have the form

\[
\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = x_3 \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}
\]

and the nullspace of \( A \) is given by \( \text{Nul}(A) = \{ x \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} | x \in \mathbb{R} \} \).

You may have noticed that the dimensionality of \( A \)'s nullspace (1) and its range (2) add up to the number of columns in \( A \). This observation holds true for any arbitrary matrix and is known as the **rank-nullity theorem**.

2.4. Derive the trace and determinant of the matrix in problem 2.1. Write the eigenvalues and eigenvectors. (3pts)

**Solution.**

- Trace given by sum of diagonal elements, so \( \text{tr}(A) = 1 + 1 + 0 = 2 \)
- \( \text{det}(A) = 0 \)
- Eigenvalues given by roots of \( \text{det}(A-\lambda I) = \lambda - 2\lambda^2 - \lambda^3 \Rightarrow \lambda_1 = 0, \lambda_2 = 1+\sqrt{2}, \lambda_3 = 1-\sqrt{2} \)
- Eigenvector(s) \( \vec{v}_i \) corresponding to \( \lambda_i \) given by basis vector(s) of \( \text{Nul}(A-\lambda_i I) \), giving us:

\[
\vec{v}_1 = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}, \vec{v}_2 = \begin{bmatrix} 1 + \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 1 \end{bmatrix}, \vec{v}_3 = \begin{bmatrix} 1 - \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \\ 1 \end{bmatrix}
\]

2.5. Prove the following properties. Note that \( \text{det} \) means determinant operator and \( \text{tr} \) trace operator. (3 pts)

- For \( A, B \in \mathbb{R}^{n \times n} \), \( \text{det}(AB) = \text{det}(A) \times \text{det}(B) \).
- For \( A \in \mathbb{R}^{m \times n}, B \in \mathbb{R}^{n \times m} \), \( \text{tr}(AB) = \text{tr}(BA) \).
- For \( A \in \mathbb{R}^{n \times n} \), \( \text{det}(A) = \prod_{i=1}^{n} \lambda_i \), and \( \text{tr}(A) = \sum_{i=1}^{n} \lambda_i \), where \( \lambda_i, i = 1, \ldots, n \) are the eigenvalues of \( A \).

**Solution.**

**Distributivity of Determinant:**
If \( A \) is not invertible, then \( AB \) is not invertible and we have \( \text{det}(AB) = \text{det}(A)\text{det}(B) = 0 \). If \( A \) is invertible, then \( A \) can be row reduced to the identity matrix \( I \) by a finite number of elementary row operations \( E_1, E_2, \ldots, E_n \), i.e.

\[
A = E_n E_{n-1} \ldots E_1 I
\]

(1)
Right-multiplying boths sides of the above by $B$, we have

$$AB = E_n E_{n-1} \ldots E_1 B$$  \hspace{1cm} (2)

Taking the determinant of both sides for both (1) and (2), we have

\[
det(A) = det(E_n E_{n-1} \ldots E_1) \\
det(AB) = det(E_n E_{n-1} \ldots E_1 B)
\]

If $E$ is an elementary row operation, we have $det(EA) = det(E)det(A)$. So,

\[
det(E_n E_{n-1} \ldots E_1 B) = det(E_n)det(E_{n-1} \ldots E_1)det(B) \\
= det(A)det(B)
\]

**Commutativity of Trace:**

\[
tr(AB) = \sum_{i=1}^{m} (AB)_{ii} = \sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij}B_{ji} \\
= \sum_{j=1}^{n} \sum_{i=1}^{m} B_{ji}A_{ij} = \sum_{j=1}^{n} (BA)_{jj} = tr(BA)
\]

**Determinant as Product of Eigenvalues:**
Let $\lambda_1, \lambda_2, \ldots, \lambda_n$ be eigenvalues of $A$. By definition, $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the roots of the characteristic polynomial of $A$:

$$p_A(t) = det(A - tI) = (\lambda_1 - t)(\lambda_2 - t) \ldots (\lambda_n - t)$$

Then, for $t = 0$ we have:

$$p_A(0) = det(A) = \lambda_1 \lambda_2 \ldots \lambda_n = \prod_{i=1}^{n} \lambda_i$$

**Trace as Sum of Eigenvalues:**
Any matrix $A \in \mathbb{R}^n$ can be transformed into Jordan canonical form $J$ by a similarity transformation $T$:

$$A = TJT^{-1},$$

where $J$ is upper-triangular with diagonal entries corresponding to the eigenvalues of $A$. Note that $tr(J) = \sum_{i=1}^{n} \lambda_i$. In conjunction with the associativity of matrix multiplication and the commutativity of trace, this lets us write:

$$tr(A) = tr(TJT^{-1}) = tr((JT^{-1})T) = tr(J(T^{-1}T)) = tr(JT) = tr(J) = \sum_{i=1}^{n} \lambda_i$$
2.6. Use the following example of matrices $A$ and $B$ to illustrate the equations in problem 2.5. (4 pts)

$$A = \begin{bmatrix} 2 & 1 & 3 \\ 1 & 1 & -2 \\ 3 & 2 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 2 & 1 \\ -2 & -1 & 2 \\ 3 & -2 & 3 \end{bmatrix}.$$

Solution. First, we compute:

$$AB = \begin{bmatrix} 2 & 1 & 3 \\ 1 & 1 & -2 \\ 3 & 2 & 1 \end{bmatrix} \times \begin{bmatrix} 1 & 2 & 1 \\ -2 & -1 & 2 \\ 3 & -2 & 3 \end{bmatrix} = \begin{bmatrix} 9 & -3 & 13 \\ -7 & 5 & -3 \\ 2 & 2 & 10 \end{bmatrix}$$

$$BA = \begin{bmatrix} 1 & 2 & 1 \\ -2 & -1 & 2 \\ 3 & -2 & 3 \end{bmatrix} \times \begin{bmatrix} 2 & 1 & 3 \\ 1 & 1 & -2 \\ 3 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 7 & 5 & 0 \\ 7 & 1 & -2 \\ 13 & 7 & 16 \end{bmatrix}$$

$$\det(A - \lambda I) = -\lambda^3 + 4\lambda^2 + \lambda \implies \lambda = 0, 2 + \sqrt{5}, 2 - \sqrt{5}$$

- $\det(A) = 0$ and $\det(B) = 32$ while $\det(AB) = 0$, so $\det(AB) = 0 = 0 \cdot 32 = \det(A)\det(B)$
- $\text{tr}(AB) = 9 + 5 + 10 = 24 = 7 + 1 + 16 = \text{tr}(BA)$
- $\det(A) = (0)(2 + \sqrt{5})(2 - \sqrt{5}) = \prod_{i=1}^{3} \lambda_i$
- $\text{tr}(A) = 2 + 1 + 1 = 4 = (0) + (2 + \sqrt{5}) + (2 - \sqrt{5}) = \sum_{i=1}^{3} \lambda_i$

3. Matrix Operations (24 pts)

**Gradient:** consider a function $f : \mathbb{R}^n \to \mathbb{R}$ that takes a vector $x \in \mathbb{R}^n$ and returns a real value. Then the gradient of $f$ (w.r.t. $x$) is the vector of partial derivatives, defined as

$$\nabla_x f(x) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \frac{\partial f(x)}{\partial x_2} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{bmatrix}.$$

**Hessian:** consider a function $f : \mathbb{R}^n \to \mathbb{R}$ that takes a vector $x \in \mathbb{R}^n$ and returns a real value. Then the Hessian matrix of $f$ (w.r.t. $x$) is the $n \times n$ matrix of partial derivatives, defined as

$$\nabla_x^2 f(x) = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_2^2} & \cdots & \frac{\partial^2 f(x)}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \frac{\partial^2 f(x)}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{bmatrix}.$$

3.1. Write the gradient and Hessian matrix for the linear function $f(x) = b^T x$, where
where \( x \in \mathbb{R}^n \) and vector \( b \in \mathbb{R}^n \). (2 pts)

**Solution.**

\[
 f(\bar{x}) = \bar{b}^\top \bar{x} = \sum_{i=1}^n b_i x_i
\]

Gradient:

\[
 \nabla_x f(x) = \begin{bmatrix}
    \frac{\partial f(x)}{\partial x_1} \\
    \frac{\partial f(x)}{\partial x_2} \\
    \vdots \\
    \frac{\partial f(x)}{\partial x_n}
\end{bmatrix} = \begin{bmatrix}
    b_1 \\
    b_2 \\
    \vdots \\
    b_n
\end{bmatrix} = \bar{b}'
\]

Hessian:

\[
 \nabla_x^2 f(x) = \begin{bmatrix}
    \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\
    \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_2^2} & \cdots & \frac{\partial^2 f(x)}{\partial x_2 \partial x_n} \\
    \vdots & \vdots & \ddots & \vdots \\
    \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \frac{\partial^2 f(x)}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_n^2}
\end{bmatrix} = \nabla_x^2 f(x) = \begin{bmatrix}
    0 & 0 & \cdots & 0 \\
    0 & 0 & \cdots & 0 \\
    \vdots & \vdots & \ddots & \vdots \\
    0 & 0 & \cdots & 0
\end{bmatrix}
\]

3.2. Write the gradient and Hessian matrix of the quadratic function

\[
 f(x) = x^T Ax + 2b^T x + c,
\]

where \( x \in \mathbb{R}^n \), matrix \( A \in \mathbb{R}^{n \times n} \), vector \( b \in \mathbb{R}^n \), and \( c \in \mathbb{R} \). (2 pts)

**Solution.**

Note that \( A \) is not necessarily symmetric.

\[
 f(\bar{x}) = \bar{x}^\top A \bar{x} + 2\bar{b}^\top \bar{x} + c = \sum_{i=1}^n \sum_{j=1}^n x_i a_{ij} x_j + 2 \sum_{i=1}^n b_i x_i + c
\]

Gradient:

\[
 \nabla_x f(x) = \begin{bmatrix}
    \sum_{i=1}^n a_{i1} x_i + \sum_{j=1}^n a_{1j} x_j + 2b_1 \\
    \sum_{i=1}^n a_{i2} x_i + \sum_{j=1}^n a_{2j} x_j + 2b_2 \\
    \vdots \\
    \sum_{i=1}^n a_{in} x_i + \sum_{j=1}^n a_{nj} x_j + 2b_n
\end{bmatrix} = \begin{bmatrix}
    \sum_{i=1}^n (a_{i1} + a_{1i}) x_i + 2b_1 \\
    \sum_{i=1}^n (a_{i2} + a_{2i}) x_i + 2b_2 \\
    \vdots \\
    \sum_{i=1}^n (a_{in} + a_{ni}) x_i + 2b_n
\end{bmatrix} = (A + A^\top) \bar{x} + 2\bar{b}
\]

Hessian:

\[
 \nabla_x^2 f(x) = \begin{bmatrix}
    \frac{\partial^2 f(x)}{\partial x_1 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\
    \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_2^2} & \cdots & \frac{\partial^2 f(x)}{\partial x_2 \partial x_n} \\
    \vdots & \vdots & \ddots & \vdots \\
    \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \frac{\partial^2 f(x)}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_n^2}
\end{bmatrix} = \begin{bmatrix}
    2a_{11} & a_{12} + a_{21} & \cdots & a_{1n} + a_{n1} \\
    a_{21} + a_{12} & 2a_{22} & \cdots & a_{2n} + a_{n2} \\
    \vdots & \vdots & \ddots & \vdots \\
    a_{n1} + a_{1n} & a_{n2} + a_{2n} & \cdots & 2a_{nn}
\end{bmatrix} = A + A^\top
\]
3.3. Given matrix $A \in \mathbb{R}^{m \times n}$ where $m < n$ and $\text{rank}(A) = m$, and vector $b \in \mathbb{R}^m$, find a solution $x \in \mathbb{R}^n$ such that $Ax = b$. (3 pts)

**Solution.** Since $\text{rank}(A) = m$ and $m < n$, the system $A\vec{x} = \vec{b}$ has infinitely many solutions. One particularly interesting solution is the one with minimum $\ell_2$ norm. We can find it by solving the following constrained optimization problem:

$$\min_{\vec{x} \in \mathbb{R}^n} f(\vec{x}) = \vec{x}^\top \vec{x}$$

s.t. $A\vec{x} = \vec{b}$

The Lagrangian is $L(\vec{x}, \vec{\lambda}) = \vec{x}^\top \vec{x} + \vec{\lambda}^\top (A\vec{x} - \vec{b})$. The first order conditions can then be solved:

$$\frac{\partial L}{\partial \vec{x}} = 2\vec{x} + A^\top \vec{\lambda} = \vec{0} \implies \vec{x} = -\frac{1}{2} A^\top \vec{\lambda} \quad (1)$$

$$\frac{\partial L}{\partial \vec{\lambda}} = A\vec{x} - \vec{b} = \vec{0} \quad (2)$$

Plugging (1) into (2) yields:

$$-\frac{1}{2} AA^\top \vec{\lambda} - \vec{b} = \vec{0} \implies AA^\top \vec{\lambda} = -2\vec{b} \implies \vec{\lambda} = -2(AA^\top)^{-1} \vec{b} \quad (3)$$

Plugging (3) back into (1), our solution is:

$$\vec{x} = A^\top (AA^\top)^{-1} \vec{b}$$

3.4. Given a nonsingular matrix $M = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$, where elements $A, B, C, D \in \mathbb{R}^{2 \times 2}$, write an analytic solution of $M^{-1}$.

a. Assume that matrix $A$ is not singular. (2 pts)

b. Assume that matrix $D$ is not singular. (2 pts)

c. Assume that both matrices $A$ and $D$ are singular. (4 pts)

**Solution.** Leveraging the fact that $A^{-1}$ exists, we can use block-matrix multiplication to reduce $M$ into something resembling row echelon form:

$$\begin{bmatrix} I & \frac{0}{-CA^{-1}} \\ -CA^{-1} & I \end{bmatrix} \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} I & -A^{-1}B \\ 0 & I \end{bmatrix} = \begin{bmatrix} A & 0 \\ 0 & D - CA^{-1}B \end{bmatrix}$$

We can then rewrite this as the following factorization:

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} I & \frac{0}{CA^{-1}} \\ CA^{-1} & I \end{bmatrix} \begin{bmatrix} A & 0 \\ 0 & D - CA^{-1}B \end{bmatrix} \begin{bmatrix} I & A^{-1}B \\ 0 & I \end{bmatrix}$$
Since $M$ is invertible and each matrix in its factorization is square, they must also be invertible, letting us compute:

$$M^{-1} = \left( \begin{bmatrix} I & 0 \\ CA^{-1} & I \end{bmatrix} \begin{bmatrix} A & 0 \\ 0 & D - CA^{-1}B \end{bmatrix} \begin{bmatrix} I & A^{-1}B \\ 0 & I \end{bmatrix} \right)^{-1}$$

$$= \begin{bmatrix} I & A^{-1}B \\ 0 & I \end{bmatrix}^{-1} \begin{bmatrix} A & 0 \\ 0 & D - CA^{-1}B \end{bmatrix}^{-1} \begin{bmatrix} I & 0 \\ CA^{-1} & I \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} I & -A^{-1}B \\ 0 & I \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & (D - CA^{-1}B)^{-1} \end{bmatrix} \begin{bmatrix} I & 0 \\ -CA^{-1} & I \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1} & -A^{-1}B(D - CA^{-1}B)^{-1} \\ -(D - CA^{-1}B)^{-1}CA^{-1} & (D - CA^{-1}B)^{-1} \end{bmatrix}^{-1}$$

You may have noticed the submatrix $(D - CA^{-1}B)$ arise several times throughout this computation. It’s known as the Schur Complement of the submatrix $A$ and is a helpful tool for computing block matrix inversions/factorizations.

(b.) Very similar to part (a), except our initial reduction is given by

$$\begin{bmatrix} I & -BD^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} I & 0 \\ -D^{-1}C & I \end{bmatrix} = \begin{bmatrix} A - BD^{-1}C & 0 \\ 0 & D \end{bmatrix},$$

which ultimately gives us:

$$M^{-1} = \begin{bmatrix} (A - BD^{-1}C)^{-1} & -(A - BD^{-1}C)^{-1}BD^{-1} \\ -D^{-1}C(A - BD^{-1}C)^{-1} & D^{-1} + D^{-1}C(A - BD^{-1}C)^{-1}BD^{-1} \end{bmatrix}$$

Here, the submatrix $(A - BD^{-1}C)$ is known as the Schur Complement of the submatrix $D$.

(c.) Since $M$ is nonsingular, so are $M^\top$ and $M^\top M$, so we can rewrite $M^{-1} = (M^\top M)^{-1}M^\top$. Note that $M^\top M$ is positive definite, meaning that its principle diagonal blocks are also positive definite (and thus nonsingular), so we can use the Schur Complement approach from parts (a) and (b) to compute $(M^\top M)^{-1}$ and subsequently $M^{-1} = (M^\top M)^{-1}M^\top$.

3.5. Given a nonsingular matrix

$$A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix},$$

write the analytic solution of $A^{-1}$. (4 pts)

**Solution.** We can obtain each entry of $A^{-1}$ using the formula $(A^{-1})_{ij} = \frac{1}{\text{det}(A)}(-1)^{i+j}\text{det}(M_{ji})$, where $M_{ji}$ is the matrix remaining after deleting row $j$, column $i$ from $A$.

Feel free to use computational tools instead of carrying out this computation by hand, but either way your final answer should be:

$$A^{-1} = \frac{1}{aei + bfg + cdh - afh - bdi - ceg} \begin{bmatrix} ei - fh & ch - bi & bf - ce \\ fg - di & ai - cg & cd - af \\ dh - eg & bg - ah & ae - bd \end{bmatrix}$$

8
3.6. Assume that matrix $A = [a_{i,j}] \in \mathbb{R}^{n \times n}$ is not singular. Derive the analytical form of the derivative of $f$ over matrix $A$ (i.e. $u_{i,j} = \nabla A f$, where $u_{i,j} = \partial f / \partial a_{i,j}$) for function $f = tr A^{-1}$. (3 pts)

**Solution.** From equation (63) in The Matrix Cookbook, we have:

$$\frac{\partial \text{Tr}(AX^{-1}B)}{\partial X} = -(X^{-1}BAX^{-1})^\top$$

Substituting $A = B = I$ and $X = A$, our solution is $\frac{\partial \text{Tr}(A^{-1})}{\partial X} = -(A^{-1}A^{-1})^\top = -(A^{-2})^\top$.

3.7. Use expression (4) in subproblem 3.5. to validate your solution of subproblem 3.6. (2 pts)

**Solution.** From 3.5, we have:

$$A^{-1} = \frac{1}{ael + bfg + cdh - afh - bdi - ceg} \begin{bmatrix} ei - fh & ch - bi & bf - ce \\
fg - di & ai - cg & cd - af \\
dh - eg & bg - ah & ae - bd \end{bmatrix}$$

Substituting this into our solution to 3.6, we get:


On the other hand, $\text{tr}(A^{-1}) = \frac{(ei + af + ac) - (fh + cg + bd)}{ael + bfg + cdh - afh - bdi - ceg}$, and computing $\frac{\partial \text{Tr}(A^{-1})}{\partial A}$ will yield the same result as above.