1. Let $A \in \mathbb{R}^{m\times n}$ and consider the system $Ax = b$. Prove the following:

   i. If b lies in the column space of A, then the system has at least one solution. **(5 points)**

   Solution: If b lies in the column space of A, then we can write b as the linear combination of the columns of A, namely,
   $$b = a_1 x_1 \cdots + a_n x_n$$
   where a_i are the columns of A and $x_1, \dots, x_n \in \mathbb{R}$. That is,
   $$b = Ax$$
   where $x = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}^\top$. The solution to the system is then given by x. □

   ii. Further, if the columns of A are linearly independent, then this solution is unique. **(5 points)**

   Solution: Suppose there exists more than one solution to the system. Consider two such solutions, x and x^*. Then,
   $$b = Ax = Ax^*$$
   whereby
   $$A(x - x^*) = 0.$$
   Since the columns of A are linearly independent,
   $$\text{rank } A = n \quad \Longrightarrow \quad \mathcal{N}(A) = \{0\}.$$
   Therefore, we must have
   $$x = x^*.$$
   □

iii. $\mathcal{N}(A^T A) = \mathcal{N}(A)$ **(5 points)**

Solution: $(\supseteq)$: Let $x \in \mathcal{N}(A)$. Then,

$$A^\top Ax = A^\top 0 = 0.$$

Therefore, $\mathcal{N}(A) \subseteq \mathcal{N}(A^\top A)$.

$(\subseteq)$: Let $x \in \mathcal{N}(A^\top A)$. Then,

$$\|Ax\|^2 = x^\top A^\top Ax = x^\top 0 = 0.$$

Hence, we must have

$$Ax = 0 \quad \Longrightarrow \quad x \in \mathcal{N}(A).$$

$\square$

iv. Given

$$\Omega = \{x \in \mathbb{R}^n : Ax = b\},$$

$d \in \mathbb{R}^n$ is a feasible direction at $x \in \Omega$ if and only if d is in the null space of $A^\top A$. **(5 points)**

Solution: By iii., it suffices to prove the statement for $\mathcal{N}(A)$.

$(\Rightarrow)$ Let d be a feasible direction. Then, $\exists\ \alpha > 0$ such that

$$x + \alpha d \in \Omega.$$

That is, $A(x + \alpha d) = b \quad \Longrightarrow \quad \alpha Ad = 0 \quad \Longrightarrow \quad d \in \mathcal{N}(A)$.

$(\Leftarrow)$ Let $d \in \mathcal{N}(A)$. Then, for any $\alpha \neq 0$, we have

$$A(x + \alpha d) = Ax + 0 = b$$

whereby $x + \alpha d \in \Omega$. Hence, d is a feasible direction.

2. **(Generalized inverse)**

   i. Show that for $A \in \mathbb{R}^{m\times n}, m \geq n, \text{rank } A = n$, **(5 points)**

$$A^\dagger = (A^\top A)^{-1} A^\top.$$

Solution: First,

$$AA^\dagger A = A\left[(A^\top A)^{-1}A^\top\right]A = A(A^\top A)^{-1}A^\top A = A.$$

We choose $U = (A^\top A)^{-1}$ and $V = A(A^\top A)^{-1}(A^\top A)^{-1}A^\top$ whereby

$$UA^\top = (A^\top A)^{-1}A^\top = A^\dagger$$

and

$$A^\top V = A^\top A(A^\top A)^{-1}(A^\top A)^{-1}A^\top = (A^\top A)^{-1}A^\top = A^\dagger$$

which completes the proof. □

   ii. $A \in \mathbb{R}^{m\times n}, m \leq n, \text{rank } A = m$, **(5 points)**

$$A^\dagger = A^\top (AA^\top)^{-1}.$$

Solution: First,

$$AA^\dagger A = A\left[A^\top(AA^\top)^{-1}\right]A = AA^\top(AA^\top)^{-1}A = A.$$

We choose $U = A^\top(AA^\top)^{-1}(AA^\top)^{-1}A$ and $V = (AA^\top)^{-1}$ whereby

$$UA^\top = A^\top(AA^\top)^{-1}(AA^\top)^{-1}AA^\top = A^\top(AA^\top)^{-1} = A^\dagger$$

and

$$A^\top V = A^\top(AA^\top)^{-1} = A^\dagger$$

which completes the proof. □

iii. Let $A \in \mathbb{R}^{3\times 4}$ be given by

$$A = \begin{bmatrix} 2 & 1 & -2 & 5 \\ 1 & 0 & -3 & 2 \\ 3 & -1 & -13 & 5 \end{bmatrix}.$$

Compute the full-rank factorization $A = BC$ where $B \in \mathbb{R}^{3\times r}$, $C \in \mathbb{R}^{r\times 4}$, and $r = \operatorname{rank} A$.

Solution: We have

$$\operatorname{rank} A = 2$$

and

$$B = \begin{bmatrix} 2 & 1 \\ 1 & 0 \\ 3 & -1 \end{bmatrix} \quad \text{and} \quad C = \begin{bmatrix} 1 & 0 & -3 & 2 \\ 0 & 1 & 4 & 1 \end{bmatrix}.$$

iv. Compute $A^\dagger$. **(5 points)**

Solution: We have

$$A^\dagger = C^\dagger B^\dagger$$

where

$$B^\dagger = (B^\top B)^{-1} B^\top = \frac{1}{27} \begin{bmatrix} 5 & 2 & 5 \\ 16 & 1 & -11 \end{bmatrix}$$

and

$$C^\dagger = C^\top (CC^\top)^{-1} = \frac{1}{76} \begin{bmatrix} 9 & 5 \\ 5 & 7 \\ -7 & 13 \\ 23 & 17 \end{bmatrix}$$

whereby

$$A^\dagger = \frac{1}{2052} \begin{bmatrix} 125 & 23 & -10 \\ 137 & 17 & -52 \\ 173 & -1 & -178 \\ 387 & 63 & -72 \end{bmatrix}.$$

3. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be defined as

$$f(x) = x^\top \begin{bmatrix} 1 & 2 \\ 4 & 7 \end{bmatrix} x + x^\top \begin{bmatrix} 3 \\ 5 \end{bmatrix} + 6.$$

i. Find the gradient and Hessian of f at the point $[1 \quad 1]^\top$. **(5 points)**

Solution: The derivative of $x^\top Ax + b^\top x + c$ is given by

$$x^\top(A + A^\top) + b^\top$$

whereby

$$\nabla f(x) = \begin{bmatrix} 2 & 6 \\ 6 & 14 \end{bmatrix} x + \begin{bmatrix} 3 \\ 5 \end{bmatrix} \quad \implies \quad \nabla f(\begin{bmatrix} 1 & 1 \end{bmatrix}^\top) = \begin{bmatrix} 8 \\ 20 \end{bmatrix} + \begin{bmatrix} 3 \\ 5 \end{bmatrix} = \begin{bmatrix} 11 \\ 25 \end{bmatrix}$$

and

$$F(x) = A + A^\top = \begin{bmatrix} 2 & 6 \\ 6 & 14 \end{bmatrix} \quad \forall x \in \mathbb{R}^2.$$

ii. Find the directional derivative of f at $[1 \quad 1]^\top$ with respect to a unit vector in the direction of maximal rate of increase. **(5 points)**

Solution: The unit vector in the direction of maximal rate of increase at the point x is given by

$$d = \frac{\nabla f(x)}{\|\nabla f(x)\|}$$

which for $x = \begin{bmatrix} 1 & 1 \end{bmatrix}^\top$ is

$$d = \frac{1}{\sqrt{746}} \begin{bmatrix} 11 \\ 25 \end{bmatrix}.$$

The directional derivative of f at $[1 \quad 1]^\top$ with respect to d is given by

$$d^\top \nabla f([1 \quad 1]^\top) = \frac{746}{\sqrt{746}} = \sqrt{746}.$$

4. Show that for the steepest descent algorithm,

   i. If $\nabla f(x) \neq 0$, then $f(x^{(k+1)}) < f(x^{(k)})$. **(5 points)**

   Solution: We have $\phi_k(\alpha_k) \leq \phi_k(\alpha)$. By chain rule,

$$\phi_k'(0) = \frac{d\phi_k}{d\alpha}(0) = \nabla f(x^{(k)} - 0\nabla f(x^{(k)}))^\top(-\nabla f(x^{(k)})) = -\|\nabla f(x^{(k)})\|^2 < 0.$$

   Then, $\exists\ \overline{\alpha}$ such that $\phi_k(0) > \phi_k(\alpha)$ for all $\alpha \in (0, \overline{\alpha}]$.

   Therefore,

$$f(x^{(k+1)}) = \phi_k(\alpha_k) \ \leq \phi_k(\overline{\alpha}) \ < \phi_k(0) = f(x^{(k)}).$$

$\square$

   ii. For each k, the vectors $x^{(k+1)} - x^{(k)}$ and $x^{(k+2)} - x^{(k+1)}$ are orthogonal. **(5 points)**

   Solution: We have

$$\left\langle x^{(k+1)} - x^{(k)}, x^{(k+2)} - x^{(k+1)} \right\rangle = \alpha_k \alpha_{k+1} \left\langle \nabla f(x^{(k)}), \nabla f(x^{(k+1)}) \right\rangle$$

   By chain rule,

$$\begin{aligned} 0 = \phi_k'(\alpha_k) = \frac{d\phi_k}{d\alpha}(\alpha_k) &= \nabla f(x^{(k)} - \alpha_k \nabla f(x^{(k)}))^\top(-\nabla f(x^{(k)})) \\ &= -\left\langle \nabla f(x^{(k)}), \nabla f(x^{(k+1)}) \right\rangle. \end{aligned}$$

$\square$

5. Using Newton's method, find the minimizer of

$$f(x) = \frac{1}{2}x^2 - \sin x$$

starting with $x^{(0)} = 0.5$. You may stop the iteration when $|x^{(k+1)} - x^{(k)}| < 10^{-5}$.
Caution – Calculator users may note that x is in radians when you calculate $\sin x$. **(10 points)**

Solution: We compute

$$f'(x) = x - \cos x, \qquad f''(x) = 1 + \sin x.$$

Hence,

$$\begin{aligned} x^{(1)} &= 0.5 - \frac{0.5 - \cos 0.5}{1 + \sin 0.5} \\ &= 0.5 - \frac{-0.3775}{1.479} \\ &= 0.7552. \end{aligned}$$

Proceeding in a similar manner, we obtain

$$\begin{aligned} x^{(2)} &= x^{(1)} - \frac{f'(x^{(1)})}{f''(x^{(1)})} = x^{(1)} - \frac{0.02710}{1.685} &&= 0.7391, \\ x^{(3)} &= x^{(2)} - \frac{f'(x^{(2)})}{f''(x^{(2)})} = x^{(2)} - \frac{9.461 \times 10^{-5}}{1.673} &&= 0.7390, \\ x^{(4)} &= x^{(3)} - \frac{f'(x^{(3)})}{f''(x^{(3)})} = x^{(3)} - \frac{1.17 \times 10^{-9}}{1.673} &&= 0.7390. \end{aligned}$$

Note that $|x^{(4)} - x^{(3)}| < 10^{-5}$. Therefore, we stop iterating and output $x^* = x^{(4)} = 0.7390$.

6. Let $A \in \mathbb{R}^{m\times n}, m < n, \text{rank A} = m$. Let $Q \in \mathbb{R}^{n\times n}, Q = Q^\top > 0$.

   i. Show that $AQ^{-1}A^\top$ is invertible. *Hint:* Write $Q^{-1} = Q^{-\frac{1}{2}}Q^{-\frac{1}{2}}$ **(3 points)**

   Solution: Since $Q = Q^\top > 0$, we can write

$$AQ^{-1}A^\top = AQ^{-\frac{1}{2}}Q^{-\frac{1}{2}}A^\top = BB^\top$$

   where $B = AQ^{-\frac{1}{2}}$. Since $BB^\top \in \mathbb{R}^{m\times m}$ and

$$\text{rank } BB^\top = \text{rank } B = \text{rank } AQ^{-\frac{1}{2}} = \text{rank } A = m,$$

   we conclude that it is invertible. □

   ii. Using Lagrange conditions, solve

$$\min \frac{1}{2}x^\top Qx \text{ subject to } Ax = b.$$

   Verify that the x^* you obtain is indeed a strict local minimizer. **(7 points)**

   Solution: The Lagrangian function is given by

$$\mathcal{L}(x, \lambda) = \frac{1}{2}x^\top Qx + \lambda^\top(b - Ax).$$

   Applying the Lagrange condition yields

$$D_x\mathcal{L}(x^*, \lambda) = x^{*\top}Q - \lambda^{*\top}A = 0^\top$$

   rewriting which yields

$$x^* = Q^{-1}A^\top\lambda \qquad \Longrightarrow \qquad Ax^* = AQ^{-1}A^\top\lambda.$$

   Since $Ax^* = b$ and since $AQ^{-1}A^\top$ is invertible (from i.),

$$\lambda = (AQ^{-1}A^\top)^{-1}b$$

   whereby

$$x^* = Q^{-1}A^\top(AQ^{-1}A^\top)^{-1}b.$$

   The sufficient condition for x^* to be the minimum is that the Hessian L of $\mathcal{L}$ satisfies

$$y^\top L(x^*, \lambda)y > 0$$

   for all $y \in T(x^*)$. This holds true since the Hessian $L(x^*, \lambda) = Q > 0$. □