1. Assume $x, y, w, z \in \mathbb{R}^n$ and $A \in \mathbb{R}^{n \times n}$. Denote $\langle A, B\rangle = \sum_{i=1}^n \sum_{j=1}^n A(i,j)B(i,j)$ as the inner product in the Euclidean space.

   (a) Prove $x^T Ay = \operatorname{tr}\big(Ayx^T\big)$.
   **Answer**: We have $\operatorname{tr}\big(Ayx^T\big) = \operatorname{tr}\big(x^T Ay\big)$, but $x^T Ay$ is a scalar so $\operatorname{tr}\big(x^T Ay\big) = x^T Ay$. Therefore $x^T Ay = \operatorname{tr}\big(Ayx^T\big)$.

   (b) Prove $\langle xy^T, wz^T\rangle = (x^T w)(y^T z)$.
   **Answer**: We can expand the left hand side as follows:

$$
\begin{aligned}
\langle xy^T, wz^T\rangle &= \langle \begin{pmatrix} x_1y_1 & \cdots & x_1y_n \\ \vdots & \ddots & \vdots \\ x_ny_1 & \cdots & x_ny_n \end{pmatrix}, \begin{pmatrix} w_1z_1 & \cdots & w_1z_n \\ \vdots & \ddots & \vdots \\ w_nz_1 & \cdots & w_nz_n \end{pmatrix} \rangle \\
&= x_1y_1w_1z_1 + x_1y_2w_1y_2 + \cdots + x_1y_nw_1sz_n \\
&\quad + x_2y_1w_2z_1 + x_2y_2w_2z_2 + \cdots + x_2y_nw_2sz_n \\
&\quad \vdots \\
&\quad + x_ny_1w_nz_1 + x_ny_2w_nz_2 + \cdots + x_ny_nw_nz_n \\
&= x_1w_1(y_1z_1 + y_2z_2 + \cdots + y_nz_n) \\
&\quad + x_2w_2(y_1z_1 + y_2z_2 + \cdots + y_nz_n) \\
&\quad \vdots \\
&\quad + x_nw_n(y_1z_1 + y_2z_2 + \cdots + y_nz_n) \\
&= (x_1w_1 + x_2w_2 + \cdots + x_nw_n)(y_1z_1 + y_2z_2 + \cdots + y_nz_n) \\
&= (x^T w)(y^T z)
\end{aligned}
$$

   Therefore the two sides are equal, so $\langle xy^T, wz^T\rangle = (x^T w)(y^T z)$.

2. Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix and $I_n$ be an identity matrix in $\mathbb{R}^{n\times n}$. Let $\lambda_1, \ldots, \lambda_n$ be $n$ eigenvalues of $A$ with corresponding eigenvectors $u_1, u_2, \ldots, u_n$.

   (a) Prove $u_i$ is an eigenvector of $A - 3I_n$ for $i = 1, \ldots, n$.

   **Answer**: By definition of eigenvalue, we have $Au_i = \lambda_i u_i$; we can subtract $3u_i$ from both sides:

$$Au_i - 3u_i = \lambda_i u_i - 3u_i \implies (A - 3I_n)u_i = (\lambda_i - 3)u_i.$$

   Therefore by definition $u_i$ is an eigenvector of $(A - 3I_n)$ with corresponding eigenvalue $\lambda_i - 3$.

   (b) Prove $u_i$ is an eigenvector of $(A - I_n)(A - 3I_n)$ for $i = 1, \ldots, n$.

   **Answer**: From part (a) we have $(A - 3I_n)u_i = (\lambda_i - 3)u_i$, following the same procedure we also have $(A - I_n)u_i = (\lambda_i - 1)u_i$. Then

$$\begin{aligned}(A - I_n)(A - 3I_n)u_i &= (A - I_n)[(A - 3I_n)u_i] \\ &= (A - I_n)(\lambda_i - 3)u_i \\ &= (\lambda_i - 3)[(A - I_n)u_i] \\ &= (\lambda_i - 3)(\lambda_i - 1)u_i.\end{aligned}$$

   Therefore by definition $u_i$ is an eigenvector of $(A - I_n)(A - 3I_n)$ with corresponding eigenvalue $(\lambda_i - 3)(\lambda_i - 1)$.

   (c) Compute $\mathrm{tr}((A - I_n)(A - 3I_n))$.

   **Answer**: By part (b) the eigenvalues of $(A - I_n)(A - 3I_n)$ are $(\lambda_i - 3)(\lambda_i - 1)$ for $i = 1, \ldots, n$. Then since the trace of a square matrix is the sum of its eigenvalues, we have

$$\mathrm{tr}((A - I_n)(A - 3I_n)) = \sum_{i=1}^{n}(\lambda_i - 3)(\lambda_i - 1).$$

   (d) Compute $\det((A + I_n)(A - 3I_n))$.

   **Answer**: Following the same procedure as part (b) we know that the eigenvalues of $(A + I_n)(A - 3I_n)$ are $(\lambda_i - 3)(\lambda_i + 1)$ for $i = 1, \ldots, n$. Then since the determinant of a square matrix is the product of its eigenvalues, we have

$$\det((A + I_n)(A - 3I_n)) = \prod_{i=1}^{n}(\lambda_i - 3)(\lambda_i + 1).$$

3. Denote $e_i$ as the $i$th column of the identity matrix $I_n \in \mathbb{R}^{n\times n}$ for $i = 1, \dots, n$. Let $f(x) = \sum_{i=1}^{10} \|x - e_i\|_2^2$ where $x \in \mathbb{R}^n$.

   (a) Compute $\nabla f(x)$.

   **Answer**:

$$Df(x) = 2\sum_{i=1}^{10}(x - e_i) \implies \nabla f(x) = 2\sum_{i=1}^{10}(x - e_i)^T$$

   (b) Compute $\nabla^2 f(x)$.

   **Answer**:

$$\nabla^2 f(x) = D(\nabla f(x)) = 2\sum_{i=1}^{10} 1 = 20$$

   (c) Show $f(x)$ is a convex function.

   **Answer**: By part (b), the Hessian is positive semidefinite. Therefore $f(x)$ is convex.

   (d) Find the global optimal solution of $\min_x f(x)$.

   **Answer**: Solving $\nabla f(x) = 0$ gives us

$$2\sum_{i=1}^{10}(x - e_i)^T = 0 \implies \sum_{i=1}^{10} x = \sum_{i=1}^{10} e_i \implies x = \frac{1}{10}\sum_{i=1}^{10} e_i = \begin{pmatrix} \frac{1}{10} \\ \vdots \\ \frac{1}{10} \end{pmatrix}.$$

4. Let $f(x) = \frac{1}{2}\|Ax - b\|_2^2 + \frac{1}{2}\|x\|_2^2$ where $A \in \mathbb{R}^{10\times 5}$, $x \in \mathbb{R}^5$ and $b \in \mathbb{R}^{10}$.

   (a) Prove $\nabla f(x)$ is Lipschitz continuous with Lipschitz constant $L = \left\|A^T A\right\| + 1$.

   **Answer**: We have

$$Df(x) = (Ax - b)A + x = A^2x - Ab + x \implies \nabla f(x) = (A^2x - Ab + x)^T$$

   and

$$\nabla^2 f(x) = D^2 f(x) = A^2 + 1 = \left\|A^T A\right\| + 1.$$

   Therefore $\nabla f(x)$ is Lipschitz continuous with Lipschitz constant $L = \left\|A^T A\right\| + 1$.

   (b) Write down the $k$th iteration of the gradient descent method.

   **Answer**: We have

$$\begin{aligned} x_k &= x_{k-1} - \frac{1}{L}\nabla f(x_{k-1}) \\ &= x_{k-1} - \frac{(A^2 x_{k-1} - Ab + x_{k-1})^T}{\|A^T A\| + 1}. \end{aligned}$$

   (c) Write down the $k$th iteration of Newton's method.

   **Answer**: Our search direction is

$$\begin{aligned} d_{k-1}^N &= -\nabla^2 f(x_{k-1})^{-1}\nabla f(x_{k-1}) \\ &= -(A^2 + 1)(A^2 x_{k-1} - Ab + x_{k-1})^T \end{aligned}$$

   So

$$\begin{aligned} x_k &= x_{k-1} + d_{k-1}^N \\ &= x_{k-1} - (A^2 + 1)(A^2 x_{k-1} - Ab + x_{k-1})^T. \end{aligned}$$

5. Judge the following sets are whether convex or not.

   (a) $\Omega = \{(x_1, x_2) : |x_1| + |x_2| \leq 1\}$

   **Answer**: Take $(x_1, x_2), (y_1, y_2) \in \Omega$ and $\alpha \in [0, 1]$, we have

$$|x_1| + |x_2| \leq 1 \implies \alpha(|x_1| + |x_2|) \leq \alpha$$

   and

$$|y_1| + |y_2| \leq 1 \implies (1 - \alpha)(|y_1| + |y_2|) \leq 1 - \alpha.$$

   Adding the two gives us

$$\alpha(|x_1| + |x_2|) + (1 - \alpha)(|y_1| + |y_2|) \leq \alpha + (1 - \alpha)$$

$$\implies \alpha(|x_1| + |x_2|) + (1 - \alpha)(|y_1| + |y_2|) \leq 1.$$

   So $\alpha(x_1, x_2) + (1 - \alpha)(y_1, y_2) \in \Omega$ and therefore $\Omega$ is a convex set.

   (b) $\Omega = \{(x_1, x_2) : x_1^2 + x_2^2 \geq 1\}$

   **Answer**: Take $(x_1, x_2), (y_1, y_2) \in \Omega$ and $\alpha \in [0, 1]$, we have

$$x_1^2 + x_2^2 \geq 1 \implies \alpha(x_1^2 + x_2^2) \geq \alpha$$

   and

$$y_1^2 + y_2^2 \geq 1 \implies (1 - \alpha)(y_1^2 + y_2^2) \geq 1 - \alpha.$$

   Adding the two gives us

$$\alpha(x_1^2 + x_2^2) + (1 - \alpha)(y_1^2 + y_2^2) \geq \alpha + (1 - \alpha)$$

$$\implies \alpha(x_1^2 + x_2^2) + (1 - \alpha)(y_1^2 + y_2^2) \geq 1.$$

   So $\alpha(x_1, x_2) + (1 - \alpha)(y_1, y_2) \in \Omega$ and therefore $\Omega$ is a convex set.

   (c) $\Omega = \{(x_1, x_2) : x_1^2 + x_2^2 \leq 1, x_1 + x_2 \geq 0\}$

   **Answer**: Take $(x_1, x_2), (y_1, y_2) \in \Omega$ and $\alpha \in [0, 1]$, we have

$$x_1^2 + x_2^2 \leq 1 \implies \alpha(x_1^2 + x_2^2) \leq \alpha,$$

$$x_1 + x_2 \geq 0 \implies \alpha(x_1 + x_2) \geq 0$$

   and

$$y_1^2 + y_2^2 \leq 1 \implies (1 - \alpha)(y_1^2 + y_2^2) \leq 1 - \alpha,$$

$$y_1 + y_2 \geq 0 \implies (1 - \alpha)(y_1 + y_2) \geq 0.$$

Adding the two gives us

$$\alpha(x_1^2 + x_2^2) + (1-\alpha)(y_1^2 + y_2^2) \geq \alpha + (1-\alpha) \implies \alpha(x_1^2 + x_2^2) + (1-\alpha)(y_1^2 + y_2^2) \geq 1,$$

$$\alpha(x_1 + x_2) + (1-\alpha)(y_1 + y_2) \geq 0$$

So $\alpha(x_1, x_2) + (1-\alpha)(y_1, y_2) \in \Omega$ and therefore $\Omega$ is a convex set.

6. Consider

$$\min_{x_1, x_2} x_1^2 + 2x_2^2 \text{ s.t. } x_1 + x_2 = 1$$

(a) Write down the KKT conditions.

**Answer**:

$$L(x, \lambda) = x_1^2 + 2x_2^2 - \lambda(x_1 + x_2 - 1) = 0$$

$$\implies \frac{\delta L}{\delta x_1} = 2x_1 - \lambda, \frac{\delta L}{\delta x_2} = 4x_2 - \lambda$$

Then we have

$$\begin{aligned}
&ZG : \nabla L(x^*, \lambda^*) = 0 \implies 2x_1^* - \lambda^* = 0, 4x_2^* - \lambda^* = 0\\
&PF : c(x^*) = 0 \implies x_1^* + x_2^* - 1 = 0\\
&DF : \lambda^* \geq 0\\
&CS : \lambda^* c(x^*) = 0 \implies \lambda(x_1^* + x_2^* - 1) = 0.
\end{aligned}$$

(b) Find the KKT points.

**Answer**: From ZG we have

$$2x_1^* - \lambda^* = 0, 4x_2^* - \lambda^* = 0 \implies 2x_1^* = 4x_2^* = \lambda^* \implies x_1^* = 2x_2^*,$$

then we can substitute into PF which gives us

$$x_1^* + x_2^* - 1 = 0 \implies x_1^* = \frac{2}{3}, x_2^* = \frac{1}{3}.$$