Problem 1:

*Find the global minima (in \( \mathbb{R}^2 \)) of the function*

\[
f(x, y) = 2x^2 + y^2 - 2xy - 2x^3 + x^4.
\]  

(1)

*State the general results you are using.*

**Solution:**

We compute \( \nabla f \) and \( \nabla^2 f \):

\[
\nabla f(x, y) = (4x - 2y - 6x^2 + 4x^3, 2y - 2x),
\]

\[
\nabla^2 f(x, y) = \begin{bmatrix}
4 - 12x + 12x^2 & -2 \\
-2 & 2
\end{bmatrix}.
\]

(2)

The candidate points will be solutions of

\[
y = x,
\]

\[4x - 6x^2 + 4x^3 = 2y,
\]

which are easily seen to be

\[
x(1) = (0, 0)',
\]

\[
x(2) = \left(\frac{1}{2}, \frac{1}{2}\right)',
\]

\[
x(3) = (1, 1)'
\]

(3)

Now,

\[
\nabla^2 f(0, 0) = \begin{bmatrix}
4 & -2 \\
-2 & 2
\end{bmatrix} > 0,
\]

\[
\nabla^2 f\left(\frac{1}{2}, \frac{1}{2}\right) = \begin{bmatrix}
1 & -2 \\
-2 & 2
\end{bmatrix}, \text{ indefinite,}
\]

\[
\nabla^2 f(1, 1) = \begin{bmatrix}
4 & -2 \\
-2 & 2
\end{bmatrix} > 0.
\]

(4)

Thus, only \((0, 0)'\) and \((1, 1)'\) are minima, and both are strict since the Hessian is positive definite. The point in the middle is a saddle point. The function values in both minima are equal to 0. Both are *global* minima because \( f(x, y) \approx 2x^2 + y^4 \to \infty \) when \( \|(x, y)\| \to \infty \). The simplest here is then just to check the function values for the three possible candidates.

The problem may finally also be solved by the following *trick*:

\[
f(x, y) = 2x^2 + y^2 - 2xy - 2x^3 + x^4 = (x - y)^2 + (x - x^2)^2.
\]

(5)

Thus, the global minimum is 0, which is obtained for \( x = x^2, y = x \), or \((0, 0)'\) and \((1, 1)'\).
Problem 2
(a) Define a convex set and a convex function defined on a convex set.
(b) Show that all solutions of a linear equation, \( \Omega = \{ x : Ax = b \} \) makes up a convex and closed set.
(c) Assume that \( \Omega = \{ x : Ax = b \} \neq \emptyset \) and \( b \neq 0 \). Show that the problem
\[
\min_{x \in \Omega} \| x \|
\]  
(i) for sure have solutions, and (ii) the solution is unique by verifying that the function \( f(x) = \| x \| \) is strictly convex on \( \Omega \).
(Hint: Use that if \( y \) and \( z \) are non-zero and non-parallel vectors in \( \mathbb{R}^n \), then \( k y + z < k y + k z \)).

Solution:
(a) See Foundation Note 1.
(b) The set \( \Omega \) is convex since \( \theta x + (1 - \theta) y \in \Omega \) if \( x, y \in \Omega \). It is also closed, since \( Ax = b \) if \( Ax_n = b \) and \( x_n \to x \). This may alternatively also be proved by observing that \( \Omega \) is an intersection between hyperplanes ( \( a_i'x = b_i \) ).
(c) Since \( \Omega \neq \emptyset \), we know there is at least one feasible solution, say \( x_0 \), and we may then look for the minimum of \( \| x \| \) on the closed and bounded set
\[
\Omega \cap \{ x : \| x \| \leq \| x_0 \| \}.
\]

Since \( x \to \| x \| \) is continuous, we have a minimum (Foundation Note 1).
For (ii), we observe that \( \Omega \) does not contain any parallel vectors when \( b \neq 0 \): If \( Ax = b \), then \( A(\alpha x) = ab \). The hint then shows that \( f(x) = \| x \| \) is strictly convex on \( \Omega \): If \( x, y \in \Omega \) and \( x \neq y \), then, as long as \( 0 < \theta < 1 \),
\[
\| \theta x + (1 - \theta) y \| < \| \theta x \| + (1 - \theta) \| y \| = \| x \| + (1 - \theta) \| y \|.
\]  
(8)
If the function is strictly convex and a solution exists, it is unique (Foundation Note 1).

Problem 3:
What is a Trust Region Method and how is the size of the trust region adjusted during the iterations?

Solution:
The function \( f \) is approximated by a quadratic function,
\[
f(x_k + p) \approx f_k + \nabla f_k p + \frac{1}{2} p' B p = m_k(p),
\]  
(9)
and we solve
\[
x_{k+1} = \arg \min_{p \in D} m_k(p),
\]  
(10)
where typically \( D \) is a ball with diameter \( \Delta \) centred at \( x_k \). The size of \( D \) is determined from the behaviour of
\[
\rho = \frac{f(x_k) - f(x_{k+1})}{m(x_k) - m(x_{k+1})}.
\]  
(11)
If \( \rho \approx 1 \), \( \Delta \) is increased, say \( \Delta := 2\Delta \); if \( \rho \ll 1 \), \( \Delta \) is decreased, say \( \Delta := \Delta / 2 \).
Problem 4:

Consider the problem

\[
\min_{x \in \mathbb{R}^2} f(x),
\]

\[
f(x) = x_1^2 + 4x_2^2 - 4x_1 - 8x_2.
\]  

(a) We start the Conjugate Gradient and the Steepest Descent methods at \(x_0 = 0\) and use \(d_0 = -\nabla f(0)\) as the first search direction for both. Find \(x_1\). Do the methods converge in a finite number of steps?

(b) Find a vector parallel to the next search direction of the Conjugate Gradient method.

**Solution:**

(a) The problem may be reformulated as

\[
\min_{(x_1, x_2) \in \mathbb{R}^2} \left\{ \frac{1}{2} x' \begin{bmatrix} 2 & 0 \\ 0 & 8 \end{bmatrix} x - \begin{bmatrix} 4 \\ 8 \end{bmatrix} x \right\},
\]  

and

\[
\nabla f(x) = \begin{bmatrix} 2 & 0 \\ 0 & 8 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - \begin{bmatrix} 4 \\ 8 \end{bmatrix}.
\]  

The first search direction is \(-\nabla f(0) = (4 \ 8)'\), and we end up at \(x_1 = \alpha_0 (1 \ 2)'\) where

\[
\alpha_0 = \arg \min_{\alpha \geq 0} \left\{ \alpha^2 + 4(2\alpha)^2 - 4\alpha - 8(2\alpha) \right\} = \arg \min_{\alpha \geq 0} \left\{ 17\alpha^2 - 20\alpha \right\} = \frac{10}{17}.
\]  

Thus, \(x_1 = \frac{10}{17} (1 \ 2)' = (\frac{10}{17} \ \frac{20}{17})'\).

Alternatively, the same answer is obtained by requiring

\[
0 = \nabla f(x) \cdot \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \left( \begin{bmatrix} 2 & 0 \\ 0 & 8 \end{bmatrix} \alpha - \begin{bmatrix} 4 \\ 8 \end{bmatrix} \right)' \begin{bmatrix} 1 \\ 2 \end{bmatrix} = (2\alpha - 4) \cdot 1 + (16\alpha - 8) \cdot 2 = 34\alpha - 20.
\]  

The Conjugate Gradient method converges in 2 steps, whereas Steepest Descent will use infinitely many steps (The only time SD converges in a finite number of steps is when the search direction happens to be an eigenvector of \(A\), here \(e_1\) or \(e_2\). This will never occur in this case).

(b) We need only to find a vector \(d\) such that

\[
d' \begin{bmatrix} 2 & 0 \\ 0 & 8 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = 2d_1 + 16d_2 = 0,
\]  

so \(d = (8 \ (-1))'\) will do.

As a test, we should now have \(x_1 + \alpha d = x_{sol} = (2 \ 1)'\) for a suitable \(\alpha\), and this works:

\[
\begin{bmatrix} 10/17 \\ 20/17 \end{bmatrix} + \frac{3}{17} \begin{bmatrix} 8 \\ -1 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}.
\]
Problem 5:

Solve the optimization problem

\[
\min f(x) = \frac{1}{2} \|Ax - b\|^2
\]  

where

\[
A = \begin{bmatrix}
1 & 1 & 1 \\
2 & 1 & -1 \\
1 & 2 & -2
\end{bmatrix}, \quad b = \begin{bmatrix}
1 \\
1 \\
0
\end{bmatrix}
\]

Solution:

This is a least square problem, which we solve by writing

\[
\frac{1}{2} \|Ax - b\|^2 = \frac{1}{2} (Ax - b)'(Ax - b) = \frac{1}{2} x' A'Ax - (A'b)'x + \|b\|^2;
\]

and hence the solution is found from

\[
A'Ax - A'b = 0,
\]

the normal equations. We have

\[
A'A = \begin{bmatrix}
6 & 5 \\
5 & 6
\end{bmatrix}, \quad A'b = \begin{bmatrix}
3 \\
2
\end{bmatrix}
\]

and

\[
x_{\text{sol}} = \begin{bmatrix}
\frac{8}{11} \\
\frac{3}{11}
\end{bmatrix}.
\]

Problem 6:

(a) State the Karush-Kuhn-Tucker (KKT) Theorem.

(b) Consider the following problem in \( \mathbb{R}^3 \):

\[
\min \{2x_1 + 2x_2 + 6z\}
\]

when

\[
x_1 + x_2 + x_3 = 1, \\
x_1, x_2, x_3 \geq 0.
\]

First solve the problem (e.g., graphically) after eliminating \( x_3 \). Then show how the statements in KKT theorem are at the solution.

Solution:

(a) This is found in N&W and the note about the KKT theorem. We write down the equations:

\[
\nabla f(x^*) = \sum_{i \in \mathcal{A}} \lambda_i \nabla c_i,
\]

\[
\lambda_i c_i(x) = 0, \quad i \in \mathcal{E} \cup \mathcal{I},
\]

\[
\lambda_i \geq 0, \quad i \in \mathcal{I}.
\]
(b) Because of the equality constraint we always have $x_3 = 1 - x_1 - x_2$, and since $2x_1 + 2x_2 + 6x_3 = 2x_1 + 2x_2 + 6(1 - x_1 - x_2) = 6 - 4x_2 - 4x_1$, the problem may be reduced to

$$
\begin{align*}
\text{min} & \{6 - 4(x_1 + x_2)\}, \\
x_1 + x_2 & \leq 1, \\
x_1, x_2 & \geq 0.
\end{align*}
$$

(28)

The solution is obviously given by $x_1 + x_2 = 1$, and $x_1, x_2 \geq 0$. Thus, the solution of the original problem is

$$
x^*_1 = \theta, \ x^*_2 = 1 - \theta, \ x^*_3 = 0, \ 0 \leq \theta \leq 1.
$$

(29)

Then

$$
\nabla f(x^*) = (2 \ 2 \ 6) = \lambda_1 (1 \ 1 \ 1) + \lambda_2 (1 \ 0 \ 0) + \lambda_3 (0 \ 1 \ 0) + \lambda_4 (0 \ 0 \ 1)
= 2 (1 \ 1 \ 1) + 4 (0 \ 0 \ 1).
$$

(30)

Thus $\lambda_1 = 2$, $\lambda_2 = \lambda_3 = 0$, and $\lambda_4 = 4 > 0$. Note that also $\lambda_i c_i(x) = 0$, $i = 1, \ldots, 4$ is satisfied.