NOTE: Additional comments based on your answers have been added in this version.

No aids were permitted.

Problem 1:

(a) Consider a differentiable function, \( f(x) \), where \( f: \mathbb{R}^n \to \mathbb{R} \). Write down the definition of the tangent plane, \( T_{x_0} \), of \( f \) in a point \( x_0 \). Assume that \( f \) has the property that \( f(x) \geq T_{x_0}(x) \) for all points \( x_0 \) and \( x \). Show that the function is convex (Hint: Consider \( x_1 \) and \( x_2 \) in \( \mathbb{R}^n \), and the tangent plane at a point \( x_0 \) between \( x_1 \) and \( x_2 \)).

Let

\[
    f(x,y) = x^2 + y^2 - 2yx - 2y + 2x + 5, \quad (x,y) \in \mathbb{R}^2.
\]

(b) Write down the gradient and the Hessian of \( f \) in Eqn. 1 and determine all unconstrained minima. Is \( f \) strictly convex?

Solution:

(a) The tangent plane is the start of the Taylor expansion, and is defined

\[
    T_{x_0}(x) = f(x_0) + \nabla f(x_0)(x - x_0), \quad x \in \mathbb{R}^n.
\]

We have to show that for two arbitrary points \( x_1 \) and \( x_2 \), and \( x_\theta = \theta x_1 + (1 - \theta) x_2, \quad \theta \in (0,1) \), we have

\[
    f(x_\theta) \leq \theta f(x_1) + (1 - \theta) f(x_2).
\]

For a linear function (like the tangent plane!) we will have equality in Eqn. 3. Thus,

\[
    f(x_\theta) = T_{x_\theta}(x_\theta) = \theta T_{x_1}(x_1) + (1 - \theta) T_{x_2}(x_2) \leq \theta f(x_1) + (1 - \theta) f(x_2),
\]

according to the assumption. There are several equivalent ways to see this. Another variant is given in the note Basic Tools. It is not acceptable to refer to the general result in the same note, which says that at differentiable function is convex if its values are above all tangent planes (!).

(b) The gradient of \( f \) is

\[
    \nabla f = (2x - 2y + 2, 2y - 2x - 2),
\]

and the Hessian

\[
    F = \begin{pmatrix}
        2 & -2 \\
        -2 & 2
    \end{pmatrix}.
\]

The gradient is zero along the line

\[
    y = x + 1.
\]

The Hessian has eigenvalues 0 and 4, and is therefore positive semi-definite in all points. Thus, \( f \) is convex, and the set of (global) minima is

\[
    \Gamma = \{(x,y); y = x + 1\}.
\]

Since \( f \) is constant along \( \Gamma \), it is not strictly convex (A quadratic function is never strictly convex when \( \nabla^2 f \) is only positive semidefinite. Why?).
Problem 2:

(a) Explain how the search directions are chosen for the conjugate gradient applied to the quadratic model problem

\[ \min_{x \in \mathbb{R}^n} \left\{ \frac{1}{2} x' A x - b' x \right\}, \quad A > 0. \]  

(b) How many iterations are (theoretically) needed for the steepest descent, conjugate gradient, and Newton methods for the quadratic model problem.

Solution:

(a) For the Conjugate Gradient (CG) the new search direction, \( p_{k+1} \), is a linear combination of the old search direction and the negative gradient direction, \( g_k = Ax_k - b \), in the current point \( x_k \),

\[ p_{k+1} = -g_k + \beta_k p_k. \]  

The parameter \( \beta_k \) is chosen so that \( p_{k+1} A p_k = 0 \), which leads to

\[ \beta_k = \frac{g_k' A p_k}{p_k' A p_k}. \]  

This is considered to be a complete answer, and additional details are not necessary.

(b) Note that the question is about the quadratic model problem, not a more general function.

Both the Steepest Descent (SD) and the CG methods converge in one iteration if the start point happens to be an eigenvector for \( A \). Otherwise, SD uses infinitely many iterations (This answer has been fully accepted, but I don’t recall a proof. May be some of you can prove or disprove it?).

CG uses as many iterations as there are different eigenvalues (maximum \( n \)). Newtons method, which is based on

\[ x_{k+1} = x_k - \nabla^2 f (x_k)^{-1} \nabla f (x_k)', \]  

always solves the quadratic problem in only one iteration, since

\[ x_1 = x_0 - A^{-1} (Ax_0 - b) = A^{-1} b = x^*. \]

Problem 3:

When it is easy to compute first and second derivatives of a one-dimensional function (that is, \( x \in \mathbb{R} \) and \( f(x) \in \mathbb{R} \)), it is possible to combine a trust region algorithm with Newton’s method for finding the minimum. Outline an algorithm for this.

Solution:

For this problem, it is important to read text, which states that the function is a one-dimensional function of a one-dimensional variable.

We start at a point \( x_k \) and have now an interval on the real line as the trust region,

\[ D_k = [x_k - \Delta_k, x_k + \Delta_k]. \]  

The quadratic approximation is the simple parabola

\[ m_k (x) = f (x_k) + f' (x_k) (x - x_k) + \frac{1}{2} f'' (x_k) (x - x_k)^2, \]
and solving

\[ x_{k+1} = \arg \min_{x \in D_k} m_k(x) \]  

(16)

means finding the minimum of the parabola within the interval \( D_k \). Solving \( \frac{dm_k(x)}{dx} = 0 \) is not sufficient unless the solution really is within the interval.

The rest of the algorithm is as before, starting by considering the actual vs. the predicted decrease

\[ \rho = \frac{f(x_k) - f(x_{k+1})}{f(x_k) - m_k(x_{k+1})}. \]  

(17)

We increase \( \Delta \) if \( \rho > \beta \) and shrink \( \Delta \) if \( \rho < \alpha \), \( 0 < \alpha < \beta < 1 \). It may be reasonable to let \( x_{k+1} = x_k \) when \( \rho \) is very small, at least when \( \rho < 0 \! \)!

**Problem 4:**

Consider the non-linear least square problem

\[ \min_{x \in \mathbb{R}^n} f(x) = \min_{x \in \mathbb{R}^n} \left\{ \frac{1}{2} \|h(x)\|^2 \right\} \]  

(18)

where \( h(x) \in \mathbb{R}^m \). It is easy to show that

\[ \nabla f(x)^t = J'(x)h(x), \]  

(19)

\[ \nabla^2 f(x) = J'(x)J(x) + \sum_{i=1}^{m} h_i(x) \nabla^2 h_i(x), \]  

(20)

where \( J = \{ \partial h_i / \partial x_j \} \). What is the idea behind the Gauss-Newton and Levenberg-Marquardt methods? (Do not discuss the algorithms in detail)

**Solution:**

Both methods use \( J'(x)J(x) \) as an approximation for \( \nabla^2 f(x) \).

Gauss-Newton is a line search method based on Newton’s formula for the search direction,

\[ p_{k+1} = - \left( J'(x_k)J(x_k) \right)^{-1} J'(x_k)h(x_k). \]  

(21)

(This actually solves the linear least square problem \( \min_p \|h(x_k) + J(x_k)p\|^2_2 \).

The Levenberg-Marquardt method is a trust region method where

\[ m_k(p) = f(x_k) + h'(x_k)J(x_k)p + \frac{1}{2}p^t J'(x_k)J(x_k)p. \]  

(22)

is used as the quadratic approximation.

**Problem 5:**

Consider the inequality constrained optimization problem

\[ \min_{x \in \Omega} f(x), \]  

(23)

\[ \Omega = \{ x \mid c_i(x) \geq 0, \ i \in \mathcal{I} \}, \]  

(24)

where the objective function \( f(x) \) and \(-c_i(x)\) are convex functions for all \( i \in \mathcal{I} \).
(a) Assume that $x^*$ is a KKT-point, which is defined by

$$\nabla \mathcal{L}(x^*, \lambda^*) = 0,$$

$$\lambda_i^* \cdot c_i(x^*) = 0, \quad i \in I,$$

$$\lambda_i^* \geq 0, \quad i \in I,$$

$$x^* \in \Omega,$$  \hfill (25)

where $\mathcal{L}(x, \lambda) = f(x) - \sum_{i \in I} \lambda_i c_i(x)$.

Prove that $x^*$ is a global minimum (Hint: Apply that $\mathcal{L}(x, \lambda^*)$ will be convex).

(b) Let

$$f(x, y) = x + 2y,$$

$$c_1(x, y) = y \geq 0,$$

$$c_2(x, y) = 2 - (x - 2)^2 - y^2 \geq 0,$$

$$c_3(x, y) = 1 - x^2 - y^2 \geq 0.$$  \hfill (26)

and explain why this is a problem of the form above. Make a sketch and find the solution graphically. Show that the solution is a KKT-point and a regular point (the LICQ-condition holds).

Solution:

(a) We observe that $\mathcal{L}(x, \lambda^*) = f(x) + \sum_{i \in I} \lambda_i^* (-c_i(x))$ will be convex since $\lambda_i^* \geq 0$. The result i Problem 1a is actually an equivalence (see the note about Basic Tools). Let $x$ be an arbitrary point in $\Omega$, and convince yourself about all steps in the following chain:

$$f(x) \geq \mathcal{L}(x, \lambda^*) \geq \mathcal{L}(x^*, \lambda^*) + \nabla \mathcal{L}(x^*, \lambda^*)(x - x^*) = \mathcal{L}(x^*, \lambda^*) = f(x^*).$$  \hfill (27)

(b) First of all, $f$ and $c_1$ are linear, and hence convex regardless of signs. Moreover, both $-c_2$ and $-c_3$ have positive definite Hessians.

This problem had an error in the expression for $c_2$, which should (of course) have been

$$c_2(x, y) = 4 - (x - 2)^2 - y^2 \geq 0.$$  \hfill (28)

Fortunately, the solution is almost identical under this change (Note: Since we did not see this when we made the questions, the few of you that thought that the radius was 2 instead of $\sqrt{2}$ have all got full score).

The (modified) domain in the $xy$-plane and $\nabla f$ is sketched in Fig. 1. From the direction of the gradient of $f$, it is obvious that the solution is $x^* = (2 - \sqrt{2}, 0)$. All constraints are fulfilled, $c_1$ and $c_2$ are active, and

$$\nabla c_1(x^*) = j,$$

$$\nabla c_2(x^*) = 2\sqrt{2}i.$$  \hfill (29)

Thus, $\nabla c_1(x^*)$ and $\nabla c_2(x^*)$ are linearly independent (in fact orthogonal) and the point is regular,

$$\nabla f(x^*) = i + 2j = \frac{1}{2\sqrt{2}} \nabla c_2(x^*) + 2\nabla c_1(x^*),$$  \hfill (30)

and

$$\lambda^* = \begin{pmatrix} 2 \\ 1 \\ \sqrt{2} \end{pmatrix} > 0.$$  \hfill (31)
Figure 1: Feasible set $\Omega$ and the gradient of the objective.