1. (10 points) For \( f(x) = x_1^2 - x_1 x_2 + 20 x_2^2 \), apply the improved version of steepest descent method to find a minimizer with starting point \( x^0 = (1,1) \). Find the explicit iteration formula for \( \{ x^k \} \). Plot the points \( x^1, x^2, \ldots, x^{20} \) in the plane by Matlab. What is the limit of the sequence \( \{ x^k \} \)?

**Solution:**

\[ f(x) = \frac{1}{2} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}^T \begin{bmatrix} 2 & -1 \\ -1 & 40 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \] is a quadratic function.

Improved version of steepest descent method:

\[ p_k = -\nabla f(x_k) = - \begin{bmatrix} 2 & -1 \\ -1 & 40 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \]

\[ x_{k+1} = x_k + \frac{\|p_k\|^2}{p_k^T H p_k} p_k, \]

limit of the sequence \( x^* = -H^{-1}c = (0,0)^T \).

Use the formula, calculate by Matlab, get \( x^1, \ldots, x^{20} \) as follows:

\[
\begin{align*}
x_1 &= [0.9750 \quad 0.0231]; \quad x_2 = [0.0469 \quad 0.0469]; \\
x_3 &= [0.0458 \quad 0.0011]; \\
x_4 &= [0.0022 \quad 0.0022]; \quad x_5 = [0.0021 \quad 0.0001]; \\
x_6 &= 1.0e-003 \times [0.1034 \quad 0.1034]; \\
x_7 &= 1.0e-003 \times [0.1008 \quad 0.0024]; \\
x_8 &= 1.0e-005 \times [0.4852 \quad 0.4852]; \\
x_9 &= 1.0e-005 \times [0.4731 \quad 0.0112]; \\
x_{10} &= 1.0e-006 \times [0.2277 \quad 0.2277]; \\
x_{11} &= 1.0e-006 \times [0.2220 \quad 0.0053];
\end{align*}
\]
x12 = 1.0e-007 *[ 0.1069 0.1069];

x13= 1.0e-007 [ 0.1042 0.0025];
x14 = 1.0e-009 [ 0.5017 0.5017];

x15 = 1.0e-009 [ 0.4891 0.0116];
x16 = 1.0e-010 [ 0.2355 0.2355];

x17 = 1.0e-010*[ 0.2296 0.0054];
x18 = 1.0e-011 * [ 0.1105 0.1105];

x19 = 1.0e-011 * [ 0.1077 0.0026];
x20 = 1.0e-013* [ 0.5187 0.5187];

2. (10 points) For the quadratic function \( f(x) = x_1^2 + 2x_2^2 + 4x_1 + 4x_2 \), apply the improved version of the steepest descent method to find a minimizer with the starting point \( x^1 = (0, 0) \). Use induction to show that the sequence \( x^k \) has the expression

\[
x^k = \left( \frac{2}{3} \right)^{k-1} - 2, \left( -\frac{1}{3} \right)^{k-1} - 1.
\]

Show that the sequence \( x^k \) converges to the unique minimizer of \( f(x) \).

**Solution:**

\[
f(x) = \frac{1}{2} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}^T \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \frac{4}{4} = \frac{1}{2} x^T H x + c^T x.
\]

\( H \) is positive definition, stationary point \( Hx^* = c, x^* = (-2, -1)^T \) is the unique minimizer.

Improved steepest descent method:

\[
p_k = -\nabla f(x_k) = - \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix} x_k - \frac{4}{4} \]

\[
x_{k+1} = x_k + \frac{\|p_k\|^2}{p_k^T H p_k} p_k.
\]

By induction to prove the formula for \( x^k \).

If \( k = 1, x^1 = (2/3^{k-1} - 2, (-1/3)^{k-1} - 1) = (0, 0) \), which is the starting point \( x^1 = (0, 0) \).

Suppose \( k \) is true \( x^k = (2/3^{k-1} - 2, (-1/3)^{k-1} - 1) \).

Next we prove \( p_k = -\nabla f(x_k) = - \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix} x_k - \frac{4}{4} = -4 \begin{bmatrix} \frac{1}{3} \end{bmatrix} \)

\[
\alpha_k = \frac{\|p_k\|^2}{p_k^T H p_k} = \frac{1}{3}.
\]

\[
x^{k+1} = x_k + \frac{\|p_k\|^2}{p_k^T H p_k} p_k = \left( \frac{2}{3} \right)^{k-1} - 2, \left( -\frac{1}{3} \right)^{k-1} - 1 - \frac{4}{3} \left( \frac{1}{3} \right) - \left( -\frac{1}{3} \right)^{k-1} = \left( \frac{2}{3} \right)^{k-1} - \left( -\frac{1}{3} \right)^{k-1} - 1 \]

which is the same as the formula \( x^{k+1} \).
\[ \lim_{k \to \infty} x^k = (-2, -1)^T, \] which is the unique minimizer of problem.

The proof is completed.

3. (10 points) For the quartic function \( f(x) = x_1^4 + x_2^4 + x_1^2 - x_1 x_2 + x_2^2 + x_1 - 2 x_2, \) apply Newton’s method to find a minimizer with starting point \( x^0 = (0, 0). \) Find the explicit iteration formula for \( \{x^k\}. \) Plot the points \( x^0, \cdots, x^5 \) in the plane by Matlab. Is \( x^5 \) a good approximate minimizer? If so, give reasons for justification.

Solution: Newton method formula:

\[
p_k = -[\nabla^2 f(x^k)]^{-1} \nabla f(x^k),
\]

\[
x^{k+1} = x^k + p_k.
\]

\[
\nabla f(x^k) = \begin{bmatrix} 4x_1^3 + 2x_1 - x_2 + 1 \\ 4x_2^3 + 2x_2 - x_1 - 2 \end{bmatrix}, \quad \nabla^2 f(x^k) = \begin{bmatrix} 12x_1^2 + 2 & -1 \\ -1 & 12x_2^2 + 2 \end{bmatrix} \succeq 0.
\]

\[
p_k = -\frac{1}{(12x_1^2+2)(12x_2^2+2)} \begin{bmatrix} 12x_1^2 + 2 & 1 \\ 1 & 12x_2^2 + 2 \end{bmatrix} \begin{bmatrix} 4x_1^3 + 2x_1 - x_2 + 1 \\ 4x_2^3 + 2x_2 - x_1 - 2 \end{bmatrix}.
\]

so \( x^{k+1} = x^k + p_k. \)

\[
x_1 = [0 \ 0.7500]; \ x_2 = [-0.1929 \ 0.6000] \quad x_3 = [-0.2036 \ 0.5603]
\]

\[
x_4 = [-0.2050 \ 0.5556] \quad x_5 = [-0.2051 \ 0.5552]
\]

\( x^5 \) is a good approximate minimizer. Check

\[
\nabla f(x^5) = \begin{bmatrix} -6.2334 \times 10^{-6} \\ 1.4604 \times 10^{-4} \end{bmatrix},
\]

\( x^5 \) is a stationary point, and hessian is positive definite, so it’s a good approximate minimizer.

4. (10 points) Let \( f(x) \) be differentiable in \( \mathbb{R}^n \) and \( p_k \) be a descent direction at point \( x_k, \) set

\[
q(x) = \nabla f(x_k)^T(x - x_k) + \frac{1}{2}(x - x_k)^T B_k(x - x_k).
\]

With \( B_k \) being symmetric positive definite. Show that

\[
\lim_{\alpha \to 0} \frac{f(x_k) - f(x_k + \alpha p_k)}{q(x_k) - q(x_k + \alpha p_k)} = 1.
\]

Proof. \( q(x_k) = 0, q(x_k + \alpha p_k) = \alpha \nabla f(x_k)^T p_k + \frac{1}{2} \alpha^2 p_k^T B_k p_k. \)

Taylor expansion: \( f(x_k + \alpha p_k) = f(x_k) + \alpha \nabla f(x_k)^T p_k + \frac{1}{2} \alpha^2 p_k^T \nabla^2 f(\xi) p_k. \)

So

\[
\lim_{\alpha \to 0} \frac{f(x_k) - f(x_k + \alpha p_k)}{q(x_k) - q(x_k + \alpha p_k)} = \lim_{\alpha \to 0} \frac{\alpha \nabla f(x_k)^T p_k + \frac{1}{2} \alpha^2 p_k^T \nabla^2 f(\xi) p_k}{\alpha \nabla f(x_k)^T p_k + \frac{1}{2} \alpha^2 p_k^T B_k p_k}.
\]
\[
\lim_{\alpha \to 0} f(x_k) - f(x_k + \alpha p_k) = \lim_{\alpha \to 0} q(x_k) - q(x_k + \alpha p_k) = \lim_{\alpha \to 0} \nabla f(x_k)^T p_k + \frac{1}{2} \alpha p_k^T \nabla^2 f(\xi) p_k = 1.
\]

5. Let \( g = \nabla f(x) \). Consider the optimization problem
\[
\min_{p \neq 0} g^T \frac{p}{\|p\|_2}
\]
show that \( p^* = -g \) is a solution to the optimization problem by Cauchy-Schwarz inequality. It shows why \(-\nabla f(x)\) is the steepest descent direction.

**Proof.** By Cauchy-Schwarz inequality,
\[
|g^T \frac{p}{\|p\|_2}| \leq \|g\|^2 \left\| \frac{p}{\|p\|_2} \right\|_2 = \|g\|_2
\]
It implies
\[
g^T \frac{p}{\|p\|_2} \geq -\|g\|_2
\]
When \( p^* = -g \), the equality holds which shows \(-g\) is a global minimizer.

6. Let \( f(x) = 2x_1^2 + 2x_2^2 + 4x_1x_2 + x_1 + x_2 \). Use unmodified steepest descent method \((x_{k+1} = x_k + p_k)\) and modified steepest descent method \((x_{k+1} = x_k + \alpha_k p_k)\) to find minimizer of \( f(x) \) with starting point \( x_0 = (0, 0) \) respectively. Which one is better?

**Sol:** The unmodified one:
\[
\nabla f(x) = \begin{pmatrix} 4x_1 + 4x_2 + 1 \\ 4x_2 + 4x_1 + 1 \end{pmatrix}
\]
Thus \( p_0 = (-1, -1)^T \) and \( x^1 = (-1, -1) \). Then \( x^2 = (6, 6) \), \( x^3 = (-43, -43) \). It can be shown by induction that
\[
x_1^k = x_2^k = \left( -\frac{7}{8} \right)^k - \frac{1}{8}
\]
So it does not converge at all.

The modified one:
We have \( p_0 = (-1, -1) \) and \( \alpha_0 = \frac{1}{8} \). Thus \( x^1 = (-\frac{1}{8}, -\frac{1}{8}) \). It satisfies
\[
\nabla f(x^1) = 0, \nabla^2 f(x^1) \succeq 0
\]
\( f(x) \) is a quadratic function, so \( x^1 \) is a global minimizer.
7. (10 points) A function $G : \mathbb{R}^n \to \mathbb{R}^m$ is called Lipschitz continuous if for any $x, y \in \mathbb{R}^n$, $\|G(x) - G(y)\| \leq M\|x - y\|$ for some constant $M$. Consider $F(x) : \mathbb{R}^n \to \mathbb{R}$, if $F'(x)$ is Lipschitz continuous with Lipschitz constant $M$, prove that the error in the linear model $L_k(x) = F(x^k) + F'(x^k)(x - x^k)$ of $F(x)$ can be bounded as

$$|F(x) - L_k(x)| \leq \frac{1}{2} M\|x - x^k\|^2.$$  

**Proof.** Let $p = x - x_k$. By Taylor expansion,

$$|F(x) - L_k(x)| = \left| \int_0^1 (F'(x_k + hp) - F'(x_k))p \, dh \right|$$  

$$\leq \int_0^1 \left| (F'(x_k + hp) - F'(x_k))p \right| \, dh$$  

$$\leq \int_0^1 \|F'(x_k + hp) - F'(x_k)\|_2 \|p\|_2 \, dh$$  

$$\leq \int_0^1 M\|hp\|_2 \|p\|_2 \, dh$$  

$$= \int_0^1 Mh \|p\|_2^2 \, dh$$  

$$= \frac{1}{2} M \|p\|_2^2 = \frac{1}{2} M\|x - x_k\|^2$$