Assignment 2 of MATP4820

Xinshi Wang MATP 4820 RIN:661975305

Problem 1

Consider the function

$$f(x_1, x_2) = \frac{1}{4}\left(2x_1 + x_2^2\right)^2.$$

1. Give the gradient of f at the point $(-1, 2)$;
   $\nabla f(x_1, x_2) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{bmatrix} = \begin{bmatrix} \frac{1}{2}(2x_1 + x_2^2) \cdot 2 \\ \frac{1}{2}(2x_1 + x_2^2) \cdot 2x_2 \end{bmatrix}$
   $\nabla f(-1, 2) = \begin{bmatrix} \frac{1}{2}(-2+4) \cdot 2 \\ \frac{1}{2}(-2+4) \cdot 4 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$

2. Show that $\mathbf{p} = (-1, -1)$ is a descent direction, i.e., $\langle \mathbf{p}, \nabla f(-1, 2)\rangle < 0$, where $\langle \cdot\,, \cdot \rangle$ denotes vector inner product.
   $\langle \mathbf{p}, \nabla f(-1, 2)\rangle = \begin{bmatrix} -1 & -1 \end{bmatrix} \begin{bmatrix} 2 \\ 4 \end{bmatrix} = -6 < 0$

3. Let $c_1 = 0.1$. Does $\alpha = 1$ satisfy the Armijo's condition?
   We need to verify that

   $$f(\mathbf{x} + \alpha\mathbf{p}) \leq f(\mathbf{x}) + c_1\alpha\langle \mathbf{p}, \nabla f(\mathbf{x})\rangle$$

   We have $f(\mathbf{x} + \alpha\mathbf{p}) = f(\begin{bmatrix} -1 \\ 2 \end{bmatrix} + 1 \times \begin{bmatrix} -1 \\ -1 \end{bmatrix}) = f(-2, 1) = \frac{1}{4} \times (-4+1)^2 = \frac{9}{4}$
   $f(\mathbf{x}) + c_1\alpha\langle \mathbf{p}, \nabla f(\mathbf{x})\rangle = f(-1, 2) + 0.1 \times 1 \times -6 = \frac{1}{4}(-2+4)^2 - 0.6 = 0.4$
   Since $\frac{9}{4} \geq 0.4$, this does not satisfy the Armijo's condition.

Problem 2

Consider the function: $f(x_1, x_2) = x_1^2 - 2x_1x_2 + x_2^2 - 4x_1 - 2x_2$. At the point $\mathbf{x}^{(0)} = (-1, 1)$,

1. give the search direction vector $\mathbf{p}$ for the steepest descent method.
   $\nabla f(x_1, x_2) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 2x_1 - 2x_2 - 4 \\ -2x_1 + 2x_2 - 2 \end{bmatrix}$
   $\nabla f(-1, -1) = \begin{bmatrix} -2 - 2 - 4 \\ 2 + 2 - 2 \end{bmatrix} = \begin{bmatrix} -8 \\ 2 \end{bmatrix}$
   $\mathbf{p} = -\nabla f(x_1, x_2) = \begin{bmatrix} 8 \\ -2 \end{bmatrix}$

2. Let $c_1 = 0.25$, determine $\bar{\alpha} > 0$ such that the Armijo's condition holds for any $\alpha \in (0, \bar{\alpha}]$.
   We need to verify that

$$f(\mathbf{x} + \alpha\mathbf{p}) \leq f(\mathbf{x}) + c_1\alpha\langle \mathbf{p}, \nabla f(\mathbf{x})\rangle$$

   After Simplification, we have:

$$f(x) = (x_1 - x_2)^2 - 4x_1 - 2x_2$$

$$\begin{aligned} f(\mathbf{x} + \bar{\alpha}\mathbf{p}^{(0)}) &= f(8\bar{\alpha} - 1, 1 - \bar{\alpha}) \\ &= (8\bar{\alpha} - 1 - 1 + 2\bar{\alpha})^2 - 32\bar{\alpha} + 4 - 2 + 4\bar{\alpha} \\ &= 100\bar{\alpha}^2 - 40\bar{\alpha} + 4 - 28\bar{\alpha} + 2 \\ &= 100\bar{\alpha}^2 - 68\bar{\alpha} + 6 \end{aligned}$$

$$\begin{aligned} f(\mathbf{x}) + c_1\alpha\langle \mathbf{p}, \nabla f(\mathbf{x})\rangle &= (1 + 2 + 1 + 4 - 2) + 0.25\bar{\alpha}(-64 - 4) \\ &= 6 + 0.25\bar{\alpha}(-68) \\ &= 6 - 17\bar{\alpha} \end{aligned}$$

   Plugging both sides into the inequality, we have:

$$\begin{aligned} 100\bar{\alpha}^2 - 68\bar{\alpha} + 6 &\leq 6 - 17 \\ 100\bar{\alpha}^2 &\leq 51\bar{\alpha} \\ 100\bar{\alpha} &\leq 51 \\ \bar{\alpha} &\leq \frac{51}{100} \end{aligned}$$

3. Let $c_1 = 0.25$ and $c_2 = 0.5$, determine the values of $\underline{\alpha}$ and $\bar{\alpha}$ such that the Wolfe's conditions hold for any $\alpha \in [\underline{\alpha}, \bar{\alpha}]$.
   To find the lowerbound $\underline{\alpha}$, we make sure it satisfies the second condition of Wolfe's condition.

$$\langle \nabla f(\mathbf{x} + \underline{\alpha}\mathbf{p}), \mathbf{p} \rangle \geq c_2 \langle \mathbf{p}, \nabla f(\mathbf{x}) \rangle$$

   We have that:

$$\begin{aligned}
\langle \nabla f(\mathbf{x} + \underline{\alpha}\mathbf{p}), \mathbf{p} \rangle &= \langle \nabla f(\begin{bmatrix} -1 \\ 1 \end{bmatrix} + \underline{\alpha} \begin{bmatrix} 8 \\ -2 \end{bmatrix}), \begin{bmatrix} 8 \\ -2 \end{bmatrix} \rangle \\
&= \langle \nabla f(8\underline{\alpha} - 1, 1 - 2\underline{\alpha}), \begin{bmatrix} 8 \\ -2 \end{bmatrix} \rangle \\
&= \begin{bmatrix} 20\underline{\alpha} - 8 & -20\underline{\alpha} + 2 \end{bmatrix} \begin{bmatrix} 8 \\ -2 \end{bmatrix} \\
&= 160\underline{\alpha} - 64 + 40\underline{\alpha} - 4 \\
&= 200\underline{\alpha} - 68
\end{aligned}$$

$$\begin{aligned}
c_2 \langle \mathbf{p}, \nabla f(\mathbf{x}) \rangle &= 0.5 \times -68 \\
&= -34
\end{aligned}$$

   Plugging values into the inequality gives us:

$$\begin{aligned}
200\underline{\alpha} - 68 &\geq -34 \\
\underline{\alpha} &\geq \frac{17}{100}
\end{aligned}$$

   Therefore the boundary is $[0.17, 0.51]$

Problem 3

Consider the unconstrained quadratic minimization problem:

$$\underset{\mathbf{x}}{\text{minimize}}\, f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\top \mathbf{A}\mathbf{x} - \mathbf{b}^\top \mathbf{x} \tag{QuadMin}$$

where $\mathbf{A} \in \mathbb{R}^{n\times n}$ is a symmetric positive definite matrix, and $\mathbf{b} \in \mathbb{R}^n$.

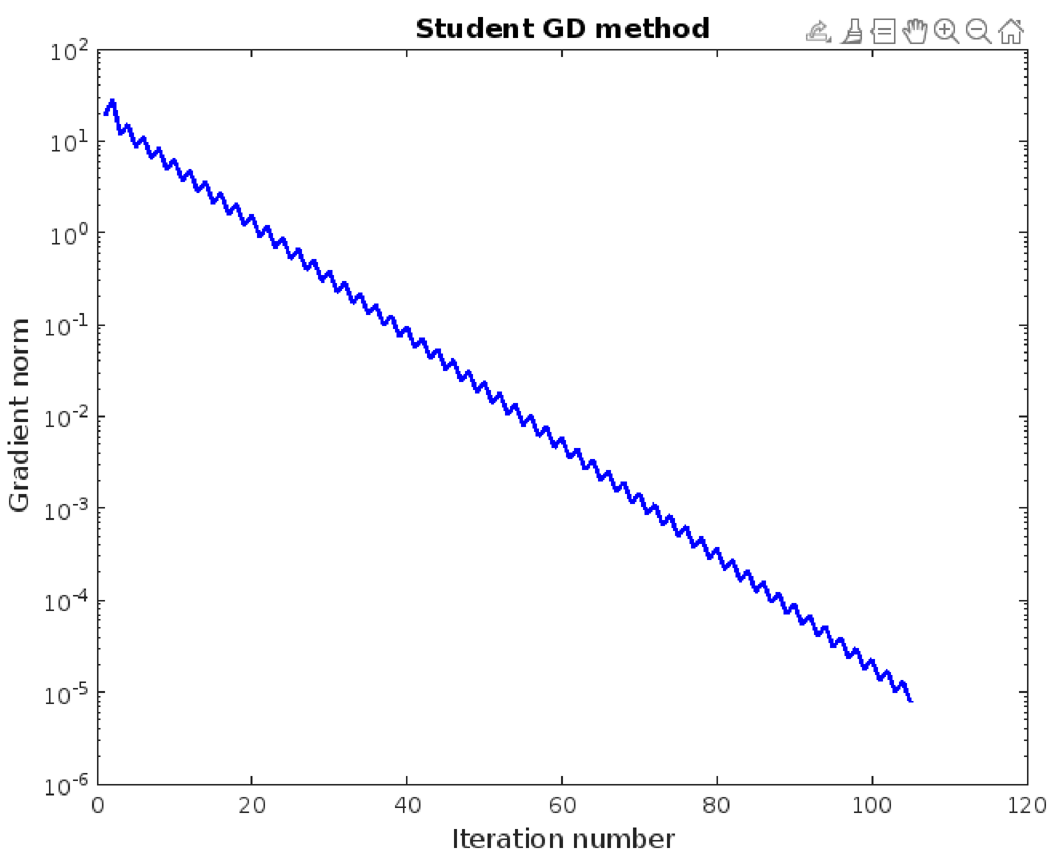
1. Let $\mathbf{A} = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 3 \\ -3 \end{bmatrix}$. Set the initial vector $\mathbf{x}^0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$. Do by hand two iterations of steepest gradient descent with exact line search.

$$\begin{aligned}
f(\mathbf{x}) &= \frac{1}{2}\mathbf{x}^\top \mathbf{A}\mathbf{x} - \mathbf{b}^\top \mathbf{x} \\
&= \frac{1}{2}(2x_1^2 - 2x_1x_2 + 2x_2^2) - 3x_1 + 3x_2 \\
&= x_1^2 - x_1x_2 + x_2^2 - 3x_1 + 3x_2 \\
\nabla f(\mathbf{x}) &= Ax - b \\
\nabla f(\mathbf{x}^{(0)}) &= A\begin{bmatrix} 0 \\ 0 \end{bmatrix} - \begin{bmatrix} 3 \\ -3 \end{bmatrix} = \begin{bmatrix} -3 \\ 3 \end{bmatrix} \\
p^{(0)} &= -\nabla f(x^{(0)}) = \begin{bmatrix} 3 \\ -3 \end{bmatrix} \\
\alpha_0 &= \underset{\alpha>0}{\arg\min}\, f(x^{(0)} + \alpha p^{(0)}) \\
&= \underset{\alpha>0}{\arg\min}\, f(3\alpha, -3\alpha) = \underset{\alpha>0}{\arg\min}\, 9\alpha^2 + 9\alpha^2 + 9\alpha^2 - 9\alpha - 9\alpha \\
54\alpha - 18 &= 0 \\
\alpha &= \frac{1}{3} \\
x^{(1)} &= x^{(0)} + \frac{1}{3}\begin{bmatrix} 3 \\ -3 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \\
\nabla f(x^{(1)}) &= \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}\begin{bmatrix} 1 \\ -1 \end{bmatrix} - \begin{bmatrix} 3 \\ -3 \end{bmatrix} \\
&= \begin{bmatrix} 3 \\ -3 \end{bmatrix} - \begin{bmatrix} 3 \\ -3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\
p^{(0)} &= \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \text{ thus } \alpha \text{ could be any value} \\
x^{(2)} &= x^{(1)} = \begin{bmatrix} 3 \\ -3 \end{bmatrix}
\end{aligned}$$

2. Use the instructor's provided file `quadMin_gd.m` to write a Matlab function `quadMin_gd` with input $\mathbf{A}$, $\mathbf{b}$, initial vector $\mathbf{x}0$, and tolerance `tol`, and with stopping condition $\|\nabla f(\mathbf{x}^k)\| \leq \text{tol}$. Also test your function by running the provided test file `test_gd.m` and compare to the instructor's function. Print your code and the results you get. [**Do your best to optimize the efficiency of your code. You will lose 10% marks if your solver takes more than double of the instructor's solver, and lose 20% if your solver takes more than triple of the instructor's solver.**]

```
function [x, hist_res] = quadMin_gd(A,b,x0,tol)
% steepest gradient method for solving
% min_x 0.5*x'*A*x - b'*x
% get the size of the problem
n = length(b);
x = x0;
% compute gradient of the objective
grad = A*x-b;
% evaluate the norm of gradient
res = norm(grad);
% save the value of res, i.e., the norm of grad
hist_res = res;
while res > tol
    % compute the stepsize alpha by exact line search
    alpha = - grad' * grad / (grad' * A * grad);
    % update x
    x = x + alpha * grad;
    % compute gradient of the objective
    grad = A*x-b;
    % evaluate the norm of gradient
    res = norm(grad);
    % save the value of res
    hist_res = [hist_res; res];
end
end
```

```
>> test_gd
Results by student code
Total running time is 0.0066
Final objective value is -281.0643
Results by Instructor code
Total running time is 0.1119
Final objective value is -281.0643
>>
```



Problem 4

Let $\mathbf{A} = \begin{bmatrix} 2 & -1 \\ -1 & 3 \end{bmatrix}$.

1. Given $\mathbf{p}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, find a vector $\mathbf{p}_2$ with unit 2-norm, i.e., $\|\mathbf{p}_2\|_2 = 1$, and $\mathbf{A}$-conjugate to $\mathbf{p}_1$.

$$p_2 = \begin{bmatrix} x \\ y \end{bmatrix}$$
$$\langle AP_1, P_2 \rangle = 0$$
$$\langle \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} x \\ y \end{bmatrix} \rangle = 0$$
$$x + 2y = 0$$

Since $||p_2||_2 = 1$, we have $\sqrt{x^2 + y^2} = 1 \implies x^2 + y^2 = 1$.
Plugging $x = -2y$ into the equation gives us $4y^2 + y^2 = 1$, thus $y^2 = \frac{1}{5}$, $y = \pm\frac{\sqrt{5}}{5}$.
Thus $\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -\frac{2\sqrt{5}}{5} \\ \frac{\sqrt{5}}{5} \end{bmatrix}$ or $\begin{bmatrix} \frac{2\sqrt{5}}{5} \\ -\frac{\sqrt{5}}{5} \end{bmatrix}$

2. Give two vectors $\mathbf{p}_1$ and $\mathbf{p}_2$ with unit 2-norm such that they are $\mathbf{A}$-conjugate and also orthogonal to each other.

$$p_1 = \begin{bmatrix} x_1 \\ y_2 \end{bmatrix} p_2 = \begin{bmatrix} x_2 \\ y_2 \end{bmatrix}$$
$$\langle AP_1, P_2 \rangle = 0$$
$$2x_1x_2 - y_1x_2 - x_1y_2 + 3y_1y_2 = 0$$

Since p_1 and p_2 are orthogonal to each other, we have $p_1 \cdot p_2 = 0 \implies x_1x_2 + y_1y_2 = 0$.
We also have $x_1^2 + y_1^2 = 1$ and $x_2^2 + y_2^2 = 1$.

A pair of vectors that would satisfy these conditions is:
$P_1 = \begin{bmatrix} -\sqrt{\frac{1}{2} - \frac{1}{2\sqrt{5}}} \\ \frac{1}{20}(6\sqrt{10(5-\sqrt{5})} - \sqrt{10}(5-\sqrt{5})^{3/2}) \end{bmatrix}$, $P_2 = \begin{bmatrix} -\sqrt{\frac{1}{10}(5+\sqrt{5})} \\ \frac{1}{20}(\sqrt{10(5+\sqrt{5})} - 5\sqrt{2(5+\sqrt{5})}) \end{bmatrix}$