IE 411: Introduction to Nonlinear Optimization

Fall 2022 - Homework Assignment 2
Due:

Question 1. For each of the following functions, find all the stationary points and classify them according to whether they are saddle points, strict/nonstrict local/global minimum/maximum points:

a. $f(x_1, x_2) = (4x_1^2 - x_2)^2$.

b. $f(x_1, x_2, x_3) = 4x_1^4 - 2x_1^2 + x_2^2 + 2x_2x_3 + 2x_3^2$.

c. $f(x_1, x_2) = 2x_2^3 - 6x_2^2 + 3x_1^2x_2$

d. $f(x_1, x_2) = x_1^4 + 2x_1^2x_2 + x_2^2 - 4x_1^2 - 8x_1 - 8x_2$.

Sol:

a. $\nabla f(x_1, x_2) = \begin{bmatrix} 64x_1^3 - 16x_1x_2 \\ -8x_1^2 + 2x_2 \end{bmatrix}$. For stationary points it suffices to solve the system

$$\begin{aligned} 64x_1^3 - 16x_1x_2 &= 0 \\ -8x_1^2 + 2x_2 &= 0 \end{aligned}$$

For any $x_1 \in \mathbb{R}$, the pair $(x_1, 4x_1^2)$ will solve this system. In order to classify them, we check

$$\nabla^2 f(x_1, x_2) = \begin{bmatrix} 192x_1^2 - 16x_2 & -16x_1 \\ -16x_1 & 2 \end{bmatrix}.$$

Since we know how stationary points act,

$$\nabla^2 f(x_1, 4x_1^2) = \begin{bmatrix} 128x_1^2 & -16x_1 \\ -16x_1 & 2 \end{bmatrix}.$$

This matrix has two eigenvalues $\lambda_1 = 0, \lambda_2 = 128x_1^2 + 2$. For each x_1 we have $\lambda_1 \geq 0, \lambda_2 > 0$, so we cannot talk about saddle, strict/nonstrict local maximum points. But, every pair $(x_1, 4x_1^2)$ is a non-strict local minimum point since $f(x_1, 4x_1^2) = f(x_1 + \epsilon, 4(x_1 + \epsilon)^2)$ and a non-strict global minimum since $f(x_1, x_2) \geq 0$ and $f(x_1, 4x_1^2) = 0$.

b. $\nabla f(x_1, x_2, x_3) = \begin{bmatrix} 16x_1^3 - 4x_1 \\ 2x_2 + 2x_3 \\ 2x_2 + 4x_3 \end{bmatrix}$. For stationary points it suffices to solve the system

$$\begin{aligned} 16x_1^3 - 4x_1 &= 0 \\ 2x_2 + 2x_3 &= 0 \\ 2x_2 + 4x_3 &= 0 \end{aligned}$$

This system has three solutions, $p_1 = (0, 0, 0), p_2 = (\frac{1}{2}, 0, 0)$ and $p_3 = (-\frac{1}{2}, 0, 0)$. In order to classify them, we check

$$\nabla^2 f(x_1, x_2, x_3) = \begin{bmatrix} 48x_1^2 - 4 & 0 & 0 \\ 0 & 2 & 2 \\ 0 & 2 & 4 \end{bmatrix}.$$

This matrix has three eigenvalues $\lambda_1 = 3 + \sqrt{5}$, $\lambda_2 = 3 - \sqrt{5}$ and $\lambda_3 = 48x_1^2 - 4$. For p_1, $\lambda_1, \lambda_2 > 0$ and $\lambda_3 < 0$ so the Hessian is indefinite and the point is a saddle point. For p_2 and p_3, $\lambda_1, \lambda_2, \lambda_3 > 0$ so the Hessian is positive definite and the points are strict local minima. Finally

$$\begin{aligned} f(x_1, x_2, x_3) = 4x_1^4 - 2x_1^2 + (x_2 + x_3)^2 + x_3^2 &\geq 4x_1^4 - 2x_1^2 \\ &\geq 4(x_1^2 - \frac{1}{4})^2 - \frac{1}{4} \geq -\frac{1}{4}. \end{aligned}$$

We also have $f(\frac{1}{2}, 0, 0) = f(-\frac{1}{2}, 0, 0) = -\frac{1}{4}$. Therefore p_2 and p_3 are non-strict global minima.

c. $\nabla f(x_1, x_2) = \begin{bmatrix} 6x_1x_2 \\ 6x_2^2 - 12x_2 + 3x_1^2 \end{bmatrix}$. For stationary points it suffices to solve the system

$$\begin{aligned} 6x_1x_2 &= 0 \\ 6x_2^2 - 12x_2 + 3x_1^2 &= 0 \end{aligned}$$

This system has two solutions, $p_1 = (0, 0)$ and $p_2 = (0, 2)$. In order to classify them, we check

$$
\begin{aligned}
\nabla^2 f(x_1, x_2) &= \begin{bmatrix} 6x_2 & 6x_1 \\ 6x_1 & 12x_2 - 12 \end{bmatrix}, \\
\nabla^2 f(0, 0) &= \begin{bmatrix} 0 & 0 \\ 0 & -12 \end{bmatrix}, \\
\nabla^2 f(0, 2) &= \begin{bmatrix} 12 & 0 \\ 0 & 12 \end{bmatrix}.
\end{aligned}
$$

$(0, 0)$ may be a local max due to the Hessian but for sufficiently small ϵ we have $f(-\epsilon, -\epsilon) < f(0, 0) < f(\epsilon, -\epsilon)$ implies that $(0, 0)$ is a saddle point, while $(0, 2)$ is a strict local minimum. Finally, $\lim\limits_{x_2 \to -\infty} f(0, x_2) = -\infty$ so $(0, 2)$ is not a global minimum.

d. $\nabla f(x_1, x_2) = \begin{bmatrix} 4x_1^3 + 4x_1x_2 - 8x_1 - 8 \\ 2x_1^2 + 2x_2 - 8 \end{bmatrix}$. For stationary points it suffices to solve the system

$$
\begin{aligned}
4x_1^3 + 4x_1x_2 - 8x_1 - 8 = 0 \\
2x_1^2 + 2x_2 - 8 = 0
\end{aligned}
$$

Solving these equations together will give

$$
4x_1^3 + 4x_1(4 - x_1^2) - 8x_1 - 8 = 0 \Rightarrow x_1 = 1, x_2 = 3.
$$

In order to classify this stationary point we check

$$
\begin{aligned}
\nabla^2 f(x_1, x_2) &= \begin{bmatrix} 12x_1^2 + 4x_2 - 8 & 4x_1 \\ 4x_1 & 2 \end{bmatrix} \\
\nabla^2 f(1, 3) &= \begin{bmatrix} 16 & 4 \\ 4 & 2 \end{bmatrix}
\end{aligned}
$$

This matrix is positive definite, hence $(1, 3)$ is a strict local minimum. Then

$$
\begin{aligned}
f(x_1, x_2) &= (x_1^2 + x_2 - 4)^2 + 4(x_1^2 - 2x_1 - 4) \\
&= (x_1^2 + x_2 - 4)^2 + 4(x_1^2 - 2x_1 + 1)^2 - 20 \\
&\geq -20.
\end{aligned}
$$

Hence $(1, 3)$ is a strict global minimum.

Question 2. Generate thirty points $(x_i, y_i), i = 1, \ldots, 30$, by the code:

```
   rand('seed',314);
x=linspace(0,1,30)';
y=2*x.^2-3*x+1+0.05*randn(size(x));
```

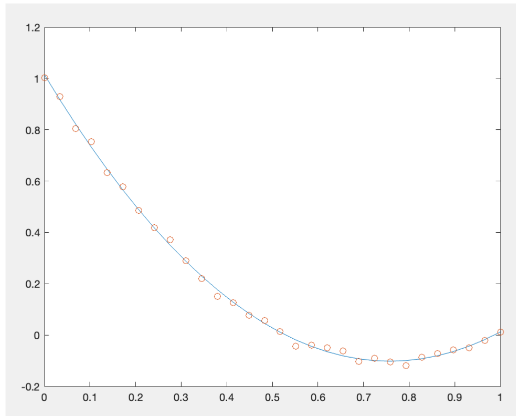
Find the quadratic function $y = ax^2 + bx + c$ that best fits the points in the least square sense. Indicate what are the parameters a, b, c found by the solution and plot the points along with the derived quadratic function. (Print out the Command Window of your MATLAB code together with the plot and attach it to your solution.)

Sol: We look for the best triple (a, b, c) to minimize

$$\left\| \underbrace{\begin{bmatrix} x_1^2 & x_1 & 1 \\ x_2^2 & x_2 & 1 \\ \vdots & \vdots & \vdots \\ x_{30}^2 & x_{30} & 1 \end{bmatrix}}_{X} \begin{bmatrix} a \\ b \\ c \end{bmatrix} - y \right\|_2 .$$

We obtain it by

```
>> rand('seed',314);
>> x=linspace(0,1,30);
>> y=2*x.^2-3*x+1+0.05*rand(size(x));
>> X=[x.^2' x' ones(length(x),1)];
>> abc = inv(X'*X)*X'*y';
>> yls=abc(1)*x.^2+abc(2)*x+abc(3)*ones(1,length(x));
>> plot(x,yls)
>> hold on
>> scatter(x,y)
```



Question 3. Consider the quadratic minimization problem

$$\text{minimize} \quad x^{\mathsf{T}} A x, \quad x \in \mathbb{R}^5,$$

where A is the 5×5 Hilbert matrix defined by $A_{ij} = \frac{1}{i+j-1}$, for $i, j = 1, \ldots, 5$. This matrix can be constructed via the MATLAB command `A=hilb(5)`. Implement and run the following methods and compare the number of iterations required by each of the methods when the initial vector is $x_0 = (1, 2, 3, 4, 5)^{\mathsf{T}}$ to obtain a solution with $\|\nabla f(x)\| \leq 10^{-4}$:

- gradient method with backtracking stepsize rule and parameters $\alpha = 0.5, \beta = 0.5, s = 1$;
- gradient method with backtracking stepsize rule and parameters $\alpha = 0.1, \beta = 0.5, s = 1$;
- gradient method with exact line search.

Sol:

- In order to use gradient method with backtracking stepsize rule and parameters $\alpha = 0.5, \beta = 0.5, s = 1$ we use the following code and it returns 3301 iterations.
- In order to use gradient method with backtracking stepsize rule and parameters $\alpha = 0.1, \beta = 0.5, s = 1$ we changed α in the previous code and it returns 3732 iterations.

```
>> x0=(1:5)';
>> A=hilb(5);
>> x=x0;
>> grad=2*A*x;
>> alpha=0.5;
>> beta=0.5;
>> s=1;
>> fun_val=x'*A*x;
>> iter=0;
>> while(norm(grad)>10^(-4))
iter=iter+1;
t=s;
while(fun_val-(x-t*grad)'*A*(x-t*grad)<alpha*t*norm(grad)^2)
t=beta*t;
end
x=x-t*grad;
fun_val=x'*A*x;
grad=2*A*x;
fprintf('iter_number = %3d norm_grad = %2.6f fun_val = %2.6f\n', iter,norm(grad),fun_val);
end
```

- In order to use gradient method with exact line search we use the following code and it returns 1271 iterations.

```
>> x0=(1:5)';
x=x0;
iter=0;
A=hilb(5);
grad=2*A*x;
while(norm(grad)>10^(-4))
iter=iter+1;
t=norm(grad)^2/(2*grad'*A*grad);
x=x-t*grad;
grad=2*A*x;
fun_val=x'*A*x;
fprintf('iter_number = %3d norm_grad = %2.6f fun_val = %2.6f\n', iter,norm(grad),fun_val);
end
```

Question 4. Show that the following set is not convex:

$$S = \{x \in \mathbb{R}^2 \mid x_1^2 - x_2^2 + x_1 + x_2 \leq 4\}.$$

Sol: Both $p_1 = (2,2)$ and $p_2 = (2,-2)$ belong to S, but $0.5p_1 + 0.5p_2 = (2,0)$ does not belong to S.

Question 5. Show that the conic hull of the set

$$S = \{x \in \mathbb{R}^2 \mid (x_1 - 1)^2 + x_2^2 = 1\}$$

is the set

$$\{x \in \mathbb{R}^2 \mid x_1 > 0\} \cup \{(0,0)^\mathsf{T}\}.$$

(Note that even if the set S is closed its conic hull cone S is not a closed set.)

Sol:

Let $x^1, \ldots, x^k \in S$ and $\lambda_1, \ldots, \lambda_k \geq 0$. If $x_1^i \neq 0$ for some $i = 1, \ldots, k$ and $\lambda_i \neq 0$, then $(\sum_{i=1}^k \lambda_i x^i)_1 > 0$ since $x_1^j, \lambda_j \geq 0$. Otherwise, $\sum_{i=1}^k \lambda_i x^i = (0,0)^\mathsf{T}$. Hence cone$(S) \subseteq \{x \in \mathbb{R}^2 | x_1 > 0\} \cup \{(0,0)^\mathsf{T}\}$.

It is clear that $(0,0)^\mathsf{T} \in \operatorname{cone}(S)$ by definition. Now, let $(a,b) \in \{x \in \mathbb{R}^2 | x_1 > 0\}$ so we have $a^2+b^2 \neq 0$. Then we have the line equation $\frac{b}{a}x_1 = x_2$ and circle equation $(x_1-1)^2 + x_2^2 = 1$. From this we obtain $x = (\frac{2a^2}{a^2+b^2}, \frac{2ab}{a^2+b^2}) \in S$ and $\lambda = \frac{a^2+b^2}{2a}$. Hence $\lambda x \in \operatorname{cone}(S)$ and $\{x \in \mathbb{R}^2 | x_1 > 0\} \cup \{(0,0)^\mathsf{T}\} \subseteq \operatorname{cone}(S)$.

Question 6. Let $a, b \in \mathbb{R}^n$ with $a \neq b$. For what values of μ is the set

$$S_\mu = \{x \in \mathbb{R}^n \mid \|x-a\|_2 \leq \mu \|x-b\|_2\}$$

convex?

Sol:

Since a norm is non-negative, $S_\mu = \emptyset$ for $\mu < 0$ and $\{a\}$ for $\mu = 0$, both convex. So we assume $\mu > 0$. We can write the description of set S_μ equivalently as

$$\begin{aligned} S_\mu &= \{x \in \mathbb{R}^n | \|x-a\|_2^2 \leq \mu^2 \|x-b\|_2^2\} \\ &= \{x \in \mathbb{R}^n | x^\mathsf{T}x - 2a^\mathsf{T}x + a^\mathsf{T}a \leq \mu^2 x^\mathsf{T}x - 2\mu^2 b^\mathsf{T}x + \mu^2 b^\mathsf{T}b\} \\ &= \{x \in \mathbb{R}^n | (1-\mu^2)x^\mathsf{T}x - 2(a+\mu^2 b)^\mathsf{T}x + a^\mathsf{T}a - \mu^2 b^T b \leq 0\} \\ &. \end{aligned}$$

In order to have this set convex we need $\mu^2 \leq 1$ which makes the quadratic function and the level set convex. Hence we need $\mu \leq 1$.