MA 576 Optimization for Data Science
Homework 3

Nicolas Jorquera

Problem 1

The objective of this exercise is to classify the stationary points locally and globally and observe the effect of the restrictions. Later in the course we will use a more systematic way by the Lagrangian and Karush-Kuhn-Tucker conditions.

1. Here we will use the **Second Derivative Test**; which states: which states a local maximum exists if $f'(c) = 0$ and $f''(c) < 0$ and a local minimum exists if $f'(c) = 0$ and $f''(c) > 0$

$$\begin{aligned} \textbf{minimize } & 4x^4 - x^3 - 4x^2 + 1 \\ s.t. \quad & x \in [-1, 1]. \end{aligned}$$

First; we must find the first and second derivative of $f(x)$:

$$f'(x) = 16x^3 - 3x^2 - 8x f''(x) = 48x^2 - 6x - 8$$

To find the stationary points, we need to find all points x for which $f'(x) = 0$

$$\begin{aligned} 0 &= 16x^3 - 3x^2 - 8x \\ 0 &= x(16x^2 - 6x - 8) \end{aligned}$$

Therefore the stationary point exist at $x = 0$, $\frac{3+\sqrt{521}}{32} \approx 0.81$, $\frac{3-\sqrt{521}}{32} \approx -0.62$. Now we can try identify the stationary point by using the second derivative test:

$$\begin{aligned} f''(0) &= 48(0)^2 - 6(0) - 8 = -8 \\ f''(0.81) &= 48(0.81)^2 - 6(0.81) - 8 = 18.6 \\ f''(-0.67) &= 48(-0.67)^2 - 6(-0.67) - 8 = 18.7 \end{aligned}$$

Because $f''(0) < 0$ then we know there is a local maxima exists. Because $f''(-0.67) > f''(0.81) > 0$ we can conclude that a minimum exists here. However in order to minimize the function we need to find the value of the function at the critical points; as well as the boundary points:

$$\begin{aligned} f(0) = 4(0)^4 - (0)^3 - 4(0)^2 + 1 &= 1 & & \\ f(0.81) &= -0.41 & f(1) &= 0 \\ f(-0.67) &= 0.29 & f(-1) &= 2 \end{aligned}$$

Therefore we can conclude that **in order to minimize the function $f(x)$ we set x = 0.81.**

2.

$$\begin{aligned} \textbf{minimize } & (x-a)^2 + 1 \quad a \in \mathbf{R} \\ s.t. \quad & x \in [-1, 1]. \end{aligned}$$

Because we have a function being squared, we know $(x - a)^2 \geq 0$ because square functions are always positive. Therefore we can also conclude that $(x - a)^2 + 1 \geq 1$. Since the minimum of this inequality is 1, we can conclude this can only happen when $x = a$. However, given the constraints $x \in [-1, 1]$; we know that $(1 - a)^2 + 1 < (-1 - a)^2 + 1$; and can conclude that **in order to minimize the function $f(x)$ we set x = $(1 - a)^2 + 1$.**

3.

$$\textbf{minimize } \|(x-c)\|^2 + 1$$
$$s.t. \quad x \in [0,1]^2,$$

where $x \in \mathbf{R}^2, c = (2, 1/2)^T$ and $[0,1]^2 = (x_1, x_2) \in \mathbf{R}^2 : 0 \le x_1 \le 1, 0 \le x_2 \le 1$.

Here we are given various constraints to the equation, but there are come similarities to the equation above as both are square functions. However, here we are dealing with a vector $[x_1, x_2]$ since $x \in \mathbf{R}^2$. By plugging this function we can see that the contour lines of this function are circular in nature. To find the vector that minimizes this function let's solve the function:

$$\|(x-c)\|^2 + 1 = (x_1 - 2)^2 + (x_2 - \frac{1}{2})^2 + 1$$

Here we will solve this function similar to the problem above. By solving $(x_2 - \frac{1}{2})^2 + 1 \ge 0$ and $(x_1 - 2)^2 \ge 0$ which gives us a minimum at $(2, \frac{-1}{2})$. We can also solve $(x_2 - \frac{1}{2})^2 \ge 0$ and $(x_1 - 2)^2 + 1 \ge 0$ which gives us a minimum at $(1, \frac{1}{2})$. **Given the constraints to the function $f(x)$ we set x** $=(1, \frac{1}{2})$ **to find the minimum** since $x \in [-1, 1]$.

Problem 2

Given the experimental data *Solution for this problem is attached as a Python File*

1. Find the least-squares line that fit this data.
   Here; we will use different linear square regression methods that python has.

2. By inspecting the residuals, can you determine if the previous model is suitable?
   Here; we can determine the suitability of the function by seeing if the residuals have any linear correlation; which would mean that our data is not linearly independent.

3. Can you find a better model than 1.? Use residual plot to measure "betterness".

Problem 3

Rewrite each problem minimizes $f_x(x)$ as a LS problem. Specify matrix A and vector b, and then solve the problem.

Therefore in order to solve this **Least Square Problem we use the Normal Equation**, since we know the set of least-squares solutions of $Ax = b$ coincides with the set of solutions of the normal equations: $A^T Ax = A^T b$. The solutions would be $x = (A^T A)^{-1} A^T b$. If the columns of A form an orthogonal basis for $Col(A)$; then the columns of A ensure that the least square solution is unique; given by $Ax = b$ (since all columns of A are independent).

1. $f(x) = (2x_1 - x_2 + 1)^2 + (x_2 - x_3)^2 + (x_3 - 1)^2$

$$\begin{aligned} f(x) &= \|(2x_1 - x_2 + 1, x_2 - x_3, x_3)\|^2 \\ &= \|(2x_1 - x_2, x_2 - x_3, x_3) - (-1, 0, 1)\|^2 \\ &= \| \begin{pmatrix} 2 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} - \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} \| \end{aligned}$$

Therefore we can minimize $f(x)$ by solving $f(x) = \|Ax - b\|^2$. Because A is an upper triangular matrix with positive diagonal entries we can solve the equation below using MATLAB.

$$X = A^{-1}b = \frac{1}{2} \begin{pmatrix} 1 & 0 & 0 \\ 1 & 2 & 0 \\ 1 & 2 & 2 \end{pmatrix}^T \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$$

By plugging in (0,1,1) into f(x) we can find the minimum = 0.

2. $f(x) = (1-x_1)^2 + \sum_{j=1}^{3}(x_j - x_{j+1})^2$
   By expanding the summation we get: $f(x) = (1-x_1)^2 + (x_1-x_2)^2 + (x_2-x_3)^2 + (x_2-x_3)^2$.
   We can represent this function in the same way we did above:

$$\begin{aligned} f(x) &= \|(-x_1, x_1 - x_2, x_2 - x_3, x_3 - x_4) - (-1,0,0,0)\|^2 \\ &= \| \begin{pmatrix} -1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} - \begin{pmatrix} -1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \| \end{aligned}$$

   From here, we can solve it the same way as we did above, plugging in terms into a matrix calculator.

$$X = A^{-1}b = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

   By plugging in (1,1,1,1) into f(x) we can find the minimum = 0.

Problem 4

Consider the LS problem with

$$A = \begin{pmatrix} 1 & -1 & 0 \\ 1 & 1 & 2 \\ 1 & -1 & 0 \\ 1 & 1 & 2 \end{pmatrix} \text{ and } b = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \end{pmatrix}$$

1. Solve the LS problem:

   Solving this LS problem using the Normal Equation we find:

$$\begin{aligned} x &= (A^T A)^{-1} A^T b \\ &= (\begin{pmatrix} 1 & -1 & 0 \\ 1 & 1 & 2 \\ 1 & -1 & 0 \\ 1 & 1 & 2 \end{pmatrix}^T \begin{pmatrix} 1 & -1 & 0 \\ 1 & 1 & 2 \\ 1 & -1 & 0 \\ 1 & 1 & 2 \end{pmatrix})^{-1} \begin{pmatrix} 1 & -1 & 0 \\ 1 & 1 & 2 \\ 1 & -1 & 0 \\ 1 & 1 & 2 \end{pmatrix}^T \begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \end{pmatrix} \end{aligned}$$

   Because the determinant of $A^T A$ is 0 we know that the inverse does not exist and the equation above cannot be solved. Therefore instead we can use QR decomposition to rewrite A; since any matrix can be represented as $A = QR$ where A is orthonormal and R is an upper triangular matrix. Instead we will use QR decomposition and find that for $A = QR$

$$R = \begin{pmatrix} -2 & 0 & -2 \\ 0 & -2 & -2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

   Using householder transformation we can find a symmetric matrix that is similar to A:

$$Q = \begin{pmatrix} -1 & -1 & -1 & -1 \\ -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

   Therefore the QR factorization gives us the solution $A = QR$. In order to get the LS problem we can solve $Ax = b$ since Q is orthonormal. Therefore:

$$QRx = b \quad \rightarrow \quad x = R^{-1}Q^T b$$

2. Write down the matrix that represents the orthogonal projection onto col(A).

   The equation also does not give a solution since R is not inevitable. Upon further reflection I will try to reduce the A into 2 basis vectors, since the columns of A are not linearly independent. After Row Echelon we find that matrix A^TA is

$$A^TA = \begin{pmatrix} 4 & 0 & 4 \\ 0 & 4 & 4 \\ 4 & 4 & 8 \end{pmatrix} \quad \rightarrow \quad \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

   Therefore we can insert the first two columns into $B\cdot$ **pseudo-inverse** such that : $B = \begin{pmatrix} 1 & -1 \\ 1 & 1 \\ 1 & -1 \\ 1 & 1 \end{pmatrix}$:

$$B(B^TB)^{-1} = \frac{1}{2}\begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}$$

Problem 5

Write the following function as a quadratic one $(x^TQx + 2b^Tx + c)$.

1. $f = (2x - y)^2 + (y - z)^2 + (z - 1)^2$
   To solve the problems above we'll expand the function, in order to rewrite a quadratic function:

$$\begin{aligned} f(x) &= (2x - y)^2 + (y - z)^2 + (z - 1)^2 \\ &= 4x^2 - 4xy + y^2 + 2y^2 - 2yz + z^2 + z^2 - 2z + 1 \\ &= x^T \begin{pmatrix} 4 & -2 & 0 \\ -2 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix} x + 2\begin{pmatrix} 0 \\ 0 \\ -2 \end{pmatrix}^T x + 1 \end{aligned}$$

2. $f = x^2 + 16xy + 4yz + y^2$
   This function written as a quadratic function is:

$$\begin{aligned} f(x) &= x^2 + 16xy + 4yz + y^2 \\ &= x^T \begin{pmatrix} 1 & 8 & 0 \\ 8 & 1 & 2 \\ 0 & 2 & 0 \end{pmatrix} x + 2\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}^T x + 0 \end{aligned}$$

Problem 6

Let $A \in \mathbf{R}^{mxn}$, $b \in \mathbf{R}^m$, $D \in \mathbf{R}^{pxn}$, and $\lambda > 0$. Consider the regularized least squares problem

$$\min_{x\in\mathbf{R}^b} \|Ax - b\|^2 + \lambda\|Dx\|^2$$

Show that the problem has a unique solution iff

$$\mathbf{null}(A) \cap \mathbf{null}(D) = 0$$

where the the null space of a linear map T, denoted by null(T), is the set of vectors x such that $Tx = 0$. A synonym for null space is kernel.

Here we are trying to minimize the regularized least square function above. In order to solve this we will start by assuming the opposite; given some matrix T, we are trying to minimize the function $\|Ax - b\|^2 + \lambda\|Tx\|^2$ and null(A) $\cup$ null(T) $\neq 0$. Using the normal equation we find that

$(A^T A + \lambda)(T^T Tx) = A^T b$. Since we know the null of any matrix can be solved using $Tx = 0$; we can let $M = A^T A + \lambda T^T T$; and we can find:

$$(A^T A + \lambda T^T T)x = 0$$

If we solve this we find that

$$\|Ax\|^2 = 0 \qquad\qquad \|Tx\|^2 = 0$$

Therefore $x \in$ **null(A) and** $x \in$ **null(T) when x = 0**. However, we can see that if we substitute x with $x+c$ the solution for the null space would still exist and would still be viable under the constraints of the problem. However because this goes against the assumptions stated above, we can conclude that the unique solution exists iff null(A) $\cup$ null(B) $= 0$.

Problem 7

Let $A \in \mathbf{R}^{mxn}$, $m \leq n$, rank$(A)= m$, and $c \in \mathbf{R}^n$, $b \in \mathbf{R}^m$. Show that the problem

$$\textbf{minimize } \|(x-c)\|^2$$
$$s.t. \quad Ax = b$$

has a unique solution and find the expression for x^*.

As mentioned in Problem 1, we know the constraint is $\|(x-c)\|^2 \geq 0$. However here we are also constrained by $Ax = b$; which allows us to solve for x: $x = A^{-1}b$. By plugging this into our constrain problem we get:

$$\textbf{minimize } \|(A^{-1}b-c)\|^2$$
$$s.t. \quad Ax = b$$

which can be rewritten as $A^T(AA^T)^{-1}b - c$. **Knowing this we can solve for any input** x^* **such that the minimum is** $x^* = A^T(AA^T)^{-1}b - c$. Because this solution is not dependent on x^* we can conclude that it is a unique solution given our constraints.