Applied Linear Algebra in Data Analaysis
Introduction to Optimization

Marks: 47

1. **(Feasible sets)** Find and sketch the feasible set for the following optimization problem to minimize $f(\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^2$, subject to the following constraints. [**Marks: 5**]

   (a) $\mathbf{h}(\mathbf{x}) = \begin{bmatrix} 3x_1 + 4x_2 - 5 = 0 \end{bmatrix}$

   (b) $\mathbf{h}(\mathbf{x}) = \begin{bmatrix} x_1 - x_2 - 2 = 0 \end{bmatrix}^\top$ and $\mathbf{g}(\mathbf{x}) = \begin{bmatrix} x_1 + x_2 + 1 \leq 0 \\ x_1 + x_2 - 2 \geq 0 \end{bmatrix}$

   (c) $\mathbf{g}(\mathbf{x}) = \begin{bmatrix} x_1 + x_2 \geq 0 \\ x_1 - x_2 \geq 0 \\ \mathbf{x}^\top \mathbf{x} \leq 1.0 \\ \mathbf{x}^\top \mathbf{x} \geq 0.5 \end{bmatrix}$

   (d) $\mathbf{g}(\mathbf{x}) = \begin{bmatrix} x_1^2 - x_2 \geq 0 \\ x_1^2 + x_2 + 2 \leq 0 \end{bmatrix}$

   (e) $\mathbf{g}(\mathbf{x}) = \begin{bmatrix} -\mathbf{x}^\top \mathbf{x} \leq 0 \end{bmatrix}$

2. **(Vector derivatives)** Demonstrate that the following gradients and Hessians with respect to the vector $\mathbf{x} \in \mathbb{R}^n$ are true. [**Marks: 5**]

   (a) $f(\mathbf{x}) = \mathbf{c}^\top \mathbf{x} \longrightarrow \nabla_{\mathbf{x}} f(\mathbf{x}) = \mathbf{c}^\top$ and $\mathbf{H}_f(\mathbf{x}) = \mathbf{0}$

   (b) $f(\mathbf{x}) = \mathbf{x}^\top \mathbf{A}\mathbf{x} \longrightarrow \nabla_{\mathbf{x}} f(\mathbf{x}) = 2\mathbf{x}^\top \mathbf{A}$ and $\mathbf{H}_f(\mathbf{x}) = 2\mathbf{A}$, where $\mathbf{A}$ is a symmetric matrix.

   (c) $f(\mathbf{x}) = \mathbf{x}^\top \mathbf{C}\mathbf{x} \longrightarrow \nabla_{\mathbf{x}} f(\mathbf{x}) = \mathbf{x}^\top \left(\mathbf{C} + \mathbf{C}^\top\right)$ and $\mathbf{H}_f(\mathbf{x}) = \mathbf{C} + \mathbf{C}^\top$, where $\mathbf{C}$ is need not be symmetric.

   (d) $f(\mathbf{x}) = \mathbf{x}^\top \mathbf{A}\mathbf{x} + \mathbf{b}^\top \mathbf{x} + c \longrightarrow \nabla_{\mathbf{x}} f(\mathbf{x}) = 2\mathbf{x}^\top \mathbf{A} + \mathbf{b}^\top$ and $\mathbf{H}_f(\mathbf{x}) = 2\mathbf{A}$, where $\mathbf{A}$ is symmetric.

   (e) $\mathbf{f}(\mathbf{x}) = \mathbf{A}\mathbf{x} \longrightarrow \nabla_{\mathbf{x}} \mathbf{f}(\mathbf{x}) = \mathbf{A}$.

3. Does the function $f(\mathbf{x}) = \mathbf{c}^\top \mathbf{x}$, where $\mathbf{c} \in \mathbb{R}^n$ have a minimum? Explain your answer. [**Marks: 1**]

4. Does the function $f(\mathbf{x}) = \mathbf{x}^\top \mathbf{A}\mathbf{x} + \mathbf{b}^\top \mathbf{x} + c$ have a minimum? If so, where is the minimum and explain the conditions under which the function have a minimum. Explain you answer. [**Marks: 2**]

5. **(Gradient descent)** Consider function $f : \mathbb{R} \to \mathbb{R}$ such that $0 < f''(x) \leq L$ for all $x \in \mathbb{R}$. Show that the following gradient descent algorithm converges to the minimum of the function $f(x)$.

$$x_{k+1} = \mathbf{x}_k - \alpha_k f'(x_k), \; 0 < \alpha_k < \frac{2}{L}, \; k \in \{1, 2, \cdots\}$$

This result can be extended to the case of $f : \mathbb{R}^n \rightarrow \mathbb{R}$, where $0 < \|\mathbf{H}_f(\mathbf{x})\|_2 \leq L$ for all $\mathbf{x} \in \mathbb{R}^n$. Then, show that the following gradient descent algorithm converges to the minimum of the function $f(\mathbf{x})$.**[Marks: 2]**

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha_k \nabla f(\mathbf{x}_k), \; 0 < \alpha_k < \frac{2}{L}, \; k \in \{1, 2, \cdots\}$$

6. **[Programming]** The Rosenbrock's function is the given by the following,

$$f(\mathbf{x}) = 100\left(x_2 - x_1^2\right)^2 + (1 - x_1)^2$$

Write a Python/MATLAB program to minimize the Rosenbrock's function using to following algorithms and terminate the search when the norm of the gradient is less than 10^{-3}. **[Marks: 8]**

(a) Gradient descent with a fixed step size $\alpha = 0.001$.

(b) Gradient descent with a inexact line search.

(c) Newton's method.

(d) Levenberg-Marquardt method with a $\lambda = 0.1$.

Assume $\mathbf{x}_1 = \begin{bmatrix} -2 & 2 \end{bmatrix}^\top$ as the initial guess. Plot the trajectory of the $\mathbf{x}$ for the four differnet algorithms in different colors along with the contour of the Rosenbrock's function. How long did the four methods take to reach the termination condition? **[Marks: 2]**

7. **(Perceptron)** A perceptron is a simple model of a neuron that takes a set of inputs $\{x_1, x_2, \ldots, x_n\}$ and produces an output y based on a set of weights $\{w_1, w_2, \ldots, w_n\}$ and a bias w_0. The output of the perceptron y is given by the following equation,

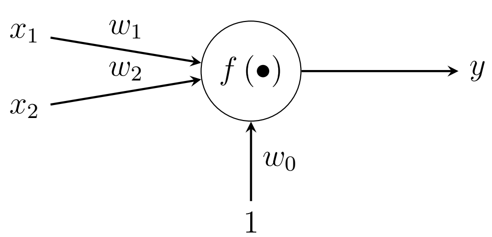
$$y = f\left(\sum_{i=1}^{n} w_i x_i + w_0\right)$$

The function $f(\bullet)$ is called the activation function of the perceptron. One of the most common activation functions is the *Sigmoid* function, which is defined as follows,

$$f(z) = \frac{1}{1 + \exp(-z)}$$

The following figure shows a simple perceptron with two inputs $\mathbf{x} = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \in \mathbb{R}^2$, and a weight vector $\mathbf{w} = \begin{bmatrix} w_0 & w_1 & w_2 \end{bmatrix} \in \mathbb{R}^3$. The output of this perceptron is given by,

$$y = \frac{1}{1 + \exp\left(-\mathbf{w}^\top \begin{bmatrix} 1 \\ \mathbf{x} \end{bmatrix}\right)}$$



We wish to fit the perceptron to a set of data $\{(\mathbf{x}_l, y_l)\}_{l=1}^m$, where $\mathbf{x}_l \in \mathbb{R}^4$ and $y_l \in \{0, 1\}$. The perceptron is trained by minimizing the following loss function,

$$l(\mathbf{w}) = \sum_{l=1}^{m} \left(y_l - \frac{1}{1 + \exp\left(-\mathbf{w}^\top \tilde{\mathbf{x}}_i\right)} \right)^2$$

where, $\tilde{\mathbf{x}}_i = \begin{bmatrix} 1 \\ \mathbf{x}_l \end{bmatrix}$.

The optimization problem can be formualted as the following,

$$\mathbf{w}^* = \arg\min_{\mathbf{w}} l(\mathbf{w})$$

We can solve this using a gradient descent algorithm, starting with a random guess for the weight vector $\mathbf{w}_1$ and update the weight vector using the following update rule,

$$\mathbf{w}_{k+1} = \mathbf{w}_k - \alpha_k \nabla l(\mathbf{w}_k)$$

where, α_k is the step size.

Find the expression for the gradient of the loss function $\nabla l(\mathbf{w})$. **[Marks: 4]**

8. Consider the following optimization problem,

$$\begin{aligned} \min_{\mathbf{x}} \quad & \frac{1}{2}\mathbf{x}^\top \mathbf{Q}\mathbf{x} \\ \text{s.t.} \quad & \mathbf{x}^\top \mathbf{P}\mathbf{x} = 1 \end{aligned}$$

Find the expression for the minimizer of this problem $\mathbf{x}^*$ and the minimum value of the objective function $\frac{1}{2}\mathbf{x}^\top \mathbf{Q}\mathbf{x}$. **[Marks: 2]**

9. Find the mimizer and maximizer of the following optimization problem $\mathbf{x} \in \mathbb{R}^3$,

$$\begin{aligned} \min_{\mathbf{x}} \quad & \left(\mathbf{a}^\top \mathbf{x}\right)\left(\mathbf{b}^\top \mathbf{x}\right) \\ \text{s.t.} \quad & x_1 + x_2 = 0 \\ & x_2 + x_3 = 0 \end{aligned}$$

where, $\mathbf{a} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$. **[Marks: 2]**

10. Consider the following optimization problem,

$$\begin{aligned} \min_{\mathbf{x}} \quad & (x_1 - a)^2 + (x_2 - b)^2 \\ \text{s.t.} \quad & x_1^2 + x_2^2 \le 1 \end{aligned}$$

where, $a, b \in \mathbb{R}$ are constant such that $a^2 + b^2 \ge 1$.

Let $\mathbf{x}^* = \begin{bmatrix} x_1^* & x_2^* \end{bmatrix}^\top$ be the minimizer of the above optimization problem. **[Marks: 10]**

(a) Use the first order necessary conditions for the unconstrained optimization problem to show that $(x_1^*)^2 + (x_2^*)^2 = 1$.

(b) Use the KKT theorem to show that $\mathbf{x}^*$ is unique and has the form $\mathbf{x}^* = \alpha \begin{bmatrix} a \\ b \end{bmatrix}$, where $\alpha \in \mathbb{R}$ is a positive constant.

(c) Find the expression for α in terms of a and b.

(d) Can you explain the solution to this problem geometrically.

(e) Does the above analysis hold when the constraint $a^2 + b^2 < 1$? Explain you answer. If it does not, what would be the solution $\mathbf{x}^*$ in this case?

11. Consider the following optimization problem,

$$\begin{aligned} \min_{\mathbf{x}} \ & \mathbf{c}^\top \mathbf{x} \\ \text{s.t.} \ & \mathbf{a}^\top \mathbf{x} \leq b \end{aligned}$$

where, the non-zero vectors $\mathbf{a}, \mathbf{c} \in \mathbb{R}^n$ and $b \in \mathbb{R}$ are constants. [**Marks: 6**]

(a) Under what conditions does the above optimization problem have a solution?

(b) When the above optimization problem has a solution, is the solution unique? If yes, find the unique mimizer $\mathbf{x}^*$, else find the set of all minimizers.

(c) Can you explain these results geometrically?