MATH520 Homework 1

Fernando

January 30, 2024

Exercise 6.3

Show that if $\boldsymbol{x}^*$ is a global minimizer of f over Ω, and $\boldsymbol{x}^* \in \Omega' \subset \Omega$, then $\boldsymbol{x}^*$ is a global minimizer of f over Ω'.

Solution:

By definition: $f(\boldsymbol{x}^*) \leq f(\boldsymbol{x})$ for all $\boldsymbol{x} \in \Omega$; in particular, it is true for all $\boldsymbol{x} \in \Omega'$ (because $\Omega' \subset \Omega$). Then again by definition $\boldsymbol{x}^*$ is a global minimizer of f over Ω'.

Exercise 6.8

Consider the following function $f : \mathbb{R}^2 \to \mathbb{R}$:

$$f(\boldsymbol{x}) = \boldsymbol{x}^T \begin{bmatrix} 1 & 2 \\ 4 & 7 \end{bmatrix} \boldsymbol{x} + \boldsymbol{x}^T \begin{bmatrix} 3 \\ 5 \end{bmatrix} + 6.$$

a. Find the gradient and Hessian of f at the point $[1, 1]^T$.
b. Find the directional derivative of f at $[1, 1]^T$ with respect to a unit vector in the direction of maximal rate of increase.
c. Find a point that satisfies the FONC (interior case) for f. Does this point satisfy the SONC (for a minimizer)?

Solution:

Part a. Let

$$A = \begin{bmatrix} 1 & 2 \\ 4 & 7 \end{bmatrix},$$

then using the identity (found in the book) $D(\boldsymbol{x}^T A \boldsymbol{x}) = \boldsymbol{x}^T (A + A^T)$ (when A is squared) we get

$$\begin{aligned} \nabla f(\boldsymbol{x}) &= \left[\boldsymbol{x}^T (A + A^T)\right]^T + [3, 5]^T \\ &= \begin{bmatrix} 2 & 6 \\ 6 & 14 \end{bmatrix} \boldsymbol{x} + \begin{bmatrix} 3 \\ 5 \end{bmatrix} \\ &= [2x_1 + 6x_2 + 3, 6x_1 + 14x_2 + 5]^T. \end{aligned}$$

Then at $[1,1]^T$ we get

$$\nabla f([1,1]^T) = [11,25]^T.$$

As for the Hessian, using the previuos calculation we get:

$$\begin{bmatrix} \partial_{x_1}\partial_{x_1} f & \partial_{x_2}\partial_{x_1} f \\ \partial_{x_1}\partial_{x_2} f & \partial_{x_2}\partial_{x_2} f \end{bmatrix} = \begin{bmatrix} 2 & 6 \\ 6 & 14 \end{bmatrix}.$$

This is the Hessian at any point, in particular at $[1,1]^T$.

Part b. Because this function is smooth we can compute the directional derivative as

$$\frac{[11,25]^T \cdot [11,25]^T}{||[11,25]^T||} = ||[11,25]^T|| = \sqrt{11^2+25^2}.$$

Part c. For this we need to solve the system

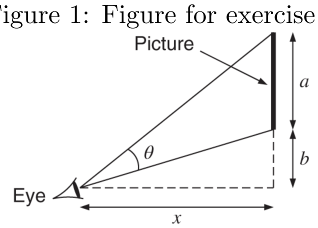
$$\begin{cases} 2x_1 + 6x_2 + 3 = 0 \\ 6x_1 + 14x_2 + 5 = 0 \end{cases}$$

Which gives us $x_1 = 3/2$ and $x_2 = -1$ so the point $[3/2,-1]^T$ satisfies the FONC for f. Using Sylvester's criterion we get that the matrix is indefinite $(\Delta_1 = 2, \Delta_2 = -8)$ so it does not satisfy the SONC.

Exercise 6.19

An art collector stands at a distance of x feet from the wall, there a piece of art (picture) of height a feet is hung, b feet above his eyes, as shown in figure 1. Find the distance from the wall for which the angle θ subtended by the eye to the picture is maximized.

Figure 1: Figure for exercise 6.19



Solution:

Let θ_1 be the angle on the eye vertex of the biggest triangle and θ_2 be the angle on the eye vertex of the lower triangle. Then $\theta = \theta_1 - \theta_2$. As suggested by the book we can optimize $\tan(\theta)$ instead of θ because tangent is increasing and $\theta \in [0, \pi/2]$ (because of the physical limitations of the problem). Using the second suggestion from the book we have to optimize

$$(\tan(\theta_1) - \tan(\theta_2))/(1 + \tan(\theta_1)\tan(\theta_2)) = \frac{a}{x + b(a+b)/x}.$$

Taking the derivative we get

$$\frac{ab(a+b) - ax^2}{(b(a+b) + x^2)^2}$$

So the candidate points are $x = \sqrt{b(a+b)}$. Notice that the negative is not considered because it doesn't make sense. Also it is a maximum because we can see that physically the problem creates a function that is concave.

Exercise 6.29

Line fitting. Let $[x_1, y_1]^T, \ldots, [x_n, y_n]^T$, $n \geq 2$, be points on the $\mathbb{R}^2$ plane (each $x_i, y_i \in \mathbb{R}$). We wish to find the straight line of the "best fit" through these points ("best" in the sense that the average squared error is minimized); that is, we wish to find $a, b \in \mathbb{R}$ to minimize

$$f(a, b) = \frac{1}{n}\sum_{i=1}^{n}(ax_i + b - y_i)^2.$$

a. Let

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$
$$\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} y_i.$$
$$\overline{X^2} = \frac{1}{n}\sum_{i=1}^{n} x_i^2.$$
$$\overline{Y^2} = \frac{1}{n}\sum_{i=1}^{n} y_i^2.$$
$$\overline{XY} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i.$$

Show that $f(a, b)$ can be written in the form $\boldsymbol{z}^T Q\boldsymbol{z} - 2\boldsymbol{c}^T\boldsymbol{z} + d$, where $\boldsymbol{z} = [a, b]^T$, $Q = Q^T \in \mathbb{R}^{2\times 2}$, $\boldsymbol{c} \in \mathbb{R}^2$ and $d \in \mathbb{R}$, and find expressions for $\boldsymbol{Q}, \boldsymbol{c}$, and d in terms of $\overline{X}, \overline{Y}, \overline{X^2}, \overline{Y^2}$, and $\overline{XY}$.
b. Assume that the $x_i, i = 1, \ldots, n$, are not all equal. Find the parameters a^*

and b^* for the line of best fit in terms of $\overline{X}, \overline{Y}, \overline{X^2}, \overline{Y^2}$ and $\overline{XY}$. Show that the point $[a^*, b^*]^T$ is the only local minimizer of f.
c. Show that if a^* and b^* are the parameters of the line of best fit, then $\overline{Y} = a^*\overline{X} + b^*$ (and hence once we have computed a^*, we can compute b^* using the formula $b^* = \overline{Y} - a^*\overline{X}$).

Solution:

a. We begin by noticing that

$$(ax_i + b - y_i)^2 = a^2x_i^2 + b^2 + y_i^2 + 2bax_i - 2ax_iy_i - 2by_i.$$

Taking a sum over all i and multiplying by $\frac{1}{n}$ we get that

$$f(a, b) = a^2\overline{X^2} + \overline{Y^2} - 2b\overline{Y} + 2ab\overline{X} - 2a\overline{XY} + b^2.$$

If we compute the quadratic form (remembering that Q is symmetric) we get

$$\boldsymbol{z}^T Q\boldsymbol{z} - 2\boldsymbol{c}^T\boldsymbol{z} + d = a^2q_{11} + 2abq_{12} + b^2q_{22} - 2ac_1 - 2bc_2 + d$$

Then by comparison we can set

$$Q = \begin{bmatrix} \overline{X^2} & \overline{X} \\ \overline{X} & 1 \end{bmatrix}$$
$$\boldsymbol{c} = [\overline{XY}, \overline{Y}]^T$$
$$d = \overline{Y^2}$$

b. First

$$\nabla f(z)^T = 2z^TQ - 2\boldsymbol{c}^T.$$

Setting this equal to 0 we obtain:

$$(z^*)^TQ = \boldsymbol{c}^T, \tag{1}$$

so

$$(z^*)^T = \boldsymbol{c}^TQ^{-1},$$

or equivalently (remembering that $(A^{-1})^T = (A^T)^{-1}$)

$$z^* = Q^{-1}\boldsymbol{c}.$$

To verify that this is a minimizer we compute

$$D^2f(z) = 2Q = 2\begin{bmatrix} \overline{X^2} & \overline{X} \\ \overline{X} & 1 \end{bmatrix}$$

It is enough to prove that this matrix is positive definite. Using Sylvester's criterion we get $\Delta_1 = \overline{X^2} > 0$ (x_i's are not equal so at most one is 0, we are assuming $i > 1$). Then $\Delta_2 = \overline{X^2} - \overline{X}^2$, which according to the book is $(1/n)\sum_{i=1}^n(x_i - \overline{X})^2$, so it is a strictly positive number because again all the x_i's are different, Hence $\Delta_2 > 0$ and by Sylvester's criterion the matrix is positive definite. Notice that the point is unique because there is only one point satisfying the FONC.

c. For this part we can use (1), which says:

$$[a^*, b^*] \begin{bmatrix} \overline{X^2} & \overline{X} \\ \overline{X} & 1 \end{bmatrix} = [\overline{XY}, \overline{Y}].$$

If we focus on the second component this yields the desired result, namely

$$\overline{Y} = a^* \overline{X} + b^*.$$