LSA University of Michigan

MATH562: Continuous Optimisation

Homework 2

Name: Keegan Gyoery UM-ID: 31799451

1. a) Let $f(x) = x^2$. Clearly, $\nabla f(x) = 2x$, so $\nabla f(0) = 0$. Obviously, $x = 0$ satisfies the first order necessary condition. Considering the Hessian (H) of $f(x)$, we have $Hf(x) = 2$, so $Hf(0) = 2$. For any $h \in \mathbb{R}$, where $h \neq 0$, $h^T(2)h = 2h^2 > 0$. Thus, $Hf(0)$ is positive definite and so also positive semi-definite. Again, $x = 0$ satisfies the second order necessary condition. Clearly, $x = 0$ satisfies both second order sufficient conditions.

   b) Let $f(x) = x^3$. Clearly, $\nabla f(x) = 3x^2$, so $\nabla f(0) = 0$. Obviously, $x = 0$ satisfies the first order necessary condition. Considering the Hessian (H) of $f(x)$, we have $Hf(x) = 6x$, so $Hf(0) = 0$. For any $h \in \mathbb{R}$, where $h \neq 0$, $h^T(0)h = 0 \geq 0$. Thus, $Hf(0)$ is positive semi-definite. Again, $x = 0$ satisfies the second order necessary condition. Clearly, $x = 0$ satisfies the 2nd second order sufficient condition, but not the 1st as $Hf(0)$ is not positive definite.

   c) Let $f(x) = x^4$. Clearly, $\nabla f(x) = 4x^3$, so $\nabla f(0) = 0$. Obviously, $x = 0$ satisfies the first order necessary condition. Considering the Hessian (H) of $f(x)$, we have $Hf(x) = 12x^2$, so $Hf(0) = 0$. For any $h \in \mathbb{R}$, where $h \neq 0$, $h^T(0)h = 0 \geq 0$. Thus, $Hf(0)$ is positive semi-definite. Again, $x = 0$ satisfies the second order necessary condition. Clearly, $x = 0$ satisfies the 2nd second order sufficient condition, but not the 1st as $Hf(0)$ is not positive definite.

2. Consider the problem

$$\min f(\mathbf{x}) = x_1^2 + \frac{1}{x_1} + x_2 + \frac{1}{x_2}.$$

   a) The gradient and Hessian for the function $f(x)$ are

$$\nabla f(\mathbf{x}) = \left[2x_1 - \frac{1}{x_1^2},\, 1 - \frac{1}{x_2^2}\right]^T,$$
$$Hf(\mathbf{x}) = \begin{bmatrix} 2 + \frac{2}{x_1^3} & 0 \\ 0 & \frac{2}{x_2^3} \end{bmatrix}.$$

   b) The first order necessary condition for a point, $\bar{\mathbf{x}}$, is that $\nabla f(\bar{\mathbf{x}}) = \mathbf{0}$. The second order necessary condition for a point, $\bar{\mathbf{x}}$, is again that $\nabla f(\bar{\mathbf{x}}) = \mathbf{0}$, and that $Hf(\bar{\mathbf{x}})$ is positive semi-definite. Setting the gradient to be $\mathbf{0}$,

$$\begin{aligned} \nabla f(\bar{\mathbf{x}}) &= \mathbf{0} \\ \therefore \left[2x_1 - \frac{1}{x_1^2},\, 1 - \frac{1}{x_2^2}\right]^T &= (0,0)^T \\ \therefore 2x_1 &= \frac{1}{x_1^2} \ldots (A), \text{ and} \\ 1 &= \frac{1}{x_2^2} \ldots (B). \\ (A) \Rightarrow x_1 &= \frac{1}{\sqrt[3]{2}}, \\ (B) \Rightarrow x_2 &= \pm 1. \end{aligned}$$

Thus our candidates for the point $\bar{\mathbf{x}}$ are $\left(\frac{1}{\sqrt[3]{2}}, 1\right)$, and $\left(\frac{1}{\sqrt[3]{2}}, -1\right)$. Clearly, both points satisfy the first order necessary condition. In order to show they satisfy the second order necessary condition, we must show that the Hessians at the two points are positive semi-definite. Examining the Hessian at the first point,

$$Hf\left(\frac{1}{\sqrt[3]{2}}, 1\right) = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}.$$

Consider $\mathbf{h} = (h_1, h_2) \neq \mathbf{0}$, and thus,

$$\begin{aligned} \mathbf{h}^T Hf\left(\frac{1}{\sqrt[3]{2}}, 1\right)\mathbf{h} &= \mathbf{h}^T \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix} \mathbf{h} \\ &= 3h_1^2 + 2h_1^2 \\ &> 0 \,\forall \mathbf{h} \end{aligned}$$

Thus, $Hf\left(\frac{1}{\sqrt[3]{2}}, 1\right)$ is positive semi-definite, so $\left(\frac{1}{\sqrt[3]{2}}, 1\right)$ satisfies the second order necessary condition as well. Examining the Hessian at the second point,

$$Hf\left(\frac{1}{\sqrt[3]{2}}, -1\right) = \begin{bmatrix} 3 & 0 \\ 0 & -2 \end{bmatrix}.$$

Consider $\mathbf{h} = (h_1, h_2) \neq \mathbf{0}$, and thus,

$$\begin{aligned} \mathbf{h}^T Hf\left(\frac{1}{\sqrt[3]{2}}, 1\right)\mathbf{h} &= \mathbf{h}^T \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} \mathbf{h} \\ &= 3h_1^2 - 2h_1^2 \\ &\not> 0 \,\forall \mathbf{h} \end{aligned}$$

Thus, $Hf\left(\frac{1}{\sqrt[3]{2}}, -1\right)$ is not positive semi-definite, so $\left(\frac{1}{\sqrt[3]{2}}, -1\right)$ does not satisfy the second order necessary condition as well.

c) $Hf\left(\frac{1}{\sqrt[3]{2}}, 1\right)$ is also positive definite, and thus satisfies both the second order sufficient conditions, and thus $\left(\frac{1}{\sqrt[3]{2}}, 1\right)$ is a local minimum of $f(\mathbf{x})$.

3. Consider $S = \{\mathbf{x} : x_1 + x_2 + x_3 \leq 1, x_1 \geq 0, x_2 \geq 0, x_3 \geq 0\}$, and the point, $\hat{\mathbf{y}} = (1, 2, 3)$. To find the point $\bar{\mathbf{x}} \in S$ that minimises the distance to $\hat{\mathbf{y}}$, the problem can be formulated as

$$\begin{aligned} \min f(\mathbf{x}) &= \|\hat{\mathbf{y}} - \mathbf{x}\|_2, \text{ subject to} \\ x_1 + x_2 + x_3 &\leq 1, \\ x_1 &\geq 0, \\ x_2 &\geq 0, \\ x_3 &\geq 0. \end{aligned}$$

4. Consider the problem

$$\min f(\mathbf{x}) = 5(x_1x_2 - 3)^2 + 4x_1x_2 + 2(x_1x_2 - 1)^3.$$

a) The gradient and Hessian of $f(\mathbf{x})$ are

$$\nabla f(\mathbf{x}) = \left[10x_2(x_1x_2 - 3) + 4x_2 + 6x_2(x_1x_2 - 1)^2,\ 10x_1(x_1x_2 - 3) + 4x_1 + 6x_1(x_1x_2 - 1)^2\right]^T,$$
$$Hf(\mathbf{x}) = \begin{bmatrix} 10x_2^2 + 12x_2^2(x_1x_2 - 1) & A \\ A & 10x_1^2 + 12x_1^2(x_1x_2 - 1) \end{bmatrix},$$

where $A = 20x_1x_2 - 26 + 6(x_1x_2 - 1)^2 + 12x_1x_2(x_1x_2 - 1)$. Evaluating the gradient and Hessian at $\mathbf{x} = (1, 1)$ gives

$$\nabla f(1, 1) = (-16, -16)^T,$$
$$Hf(1, 1) = \begin{bmatrix} 10 & -6 \\ -6 & 10 \end{bmatrix}.$$

b) Given $\mathbf{x} = (1, 0)$, and $\mathbf{d} = (1, 1)$,

$$\begin{aligned}
a(\theta) &= f(\mathbf{x} + \theta\mathbf{d}) \\
&= f(1 + \theta, \theta) \\
&= 5((1 + \theta)\theta - 3)^2 + 4(1 + \theta)\theta + 2((1 + \theta)\theta - 1)^3 \\
&= 5\left(\theta^2 + \theta - 3\right)^2 + 4\theta^2 + 4\theta + 2\left(\theta^2 + \theta - 1\right)^3 \\
\therefore a'(\theta) &= 10(2\theta + 1)\left(\theta^2 + \theta - 3\right) + 8\theta + 4 + 6(2\theta + 1)\left(\theta^2 + \theta - 1\right)^2 \\
\therefore a(0) &= 43 \text{ and } a'(0) = -20
\end{aligned}$$

c) Starting with $\mathbf{x}^0 = (1, 0)$, and applying steepest descent method, we have

$$\begin{aligned}
\nabla f(\mathbf{x}^0) &= (0, -20)^T, \\
\therefore \mathbf{d}^0 &= (0, 20), \\
\therefore \mathbf{x}^0 + \theta\mathbf{d}^0 &= (1, 20\theta). \\
\therefore f\left(\mathbf{x}^0 + \theta\mathbf{d}^0\right) &= f(1, 20\theta) \\
&= 5(20\theta - 3)^2 + 80\theta + 2(20\theta - 1)^3. \\
\frac{df\left(\mathbf{x}^0 + \theta\mathbf{d}^0\right)}{d\theta} &= 200(20\theta - 3) + 80 + 120(20\theta - 1)^2 \\
&= 48000\theta^2 - 800\theta - 400. \\
\frac{df\left(\mathbf{x}^0 + \theta\mathbf{d}^0\right)}{d\theta} &= 0, \\
\therefore 400\theta^2 - 2\theta - 1 &= 0 \\
\therefore \theta &= \frac{2 \pm \sqrt{1604}}{800} \\
\therefore \theta^0 &= \frac{2 + \sqrt{1604}}{800} \\
&\approx 0.05256246. \\
\therefore \mathbf{x}^1 &= \mathbf{x}^0 + \theta^0\mathbf{d}^0 \\
\therefore \mathbf{x}^1 &= (1, 1.05124922)
\end{aligned}$$

5. Consider the problem

$$\min f(\mathbf{x}) = (x_1 - 2)^2 + (x_1 - 2x_2)^3.$$

The gradient of $f(x)$ is

$$\nabla f(\mathbf{x}) = \left[2(x_1 - 2) + 3(x_1 - 2x_2)^2,\ -6(x_1 - 2x_2)^2\right]^T$$

Starting with $\mathbf{x}^0 = (0, 0)$, and applying the first step of steepest descent, we have,

$$\begin{aligned}
\nabla f(\mathbf{x}^0) &= (0, -4)^T, \\
\therefore \mathbf{d}^0 &= (0, 4), \\
\therefore \mathbf{x}^0 + \theta\mathbf{d}^0 &= (4\theta, 0). \\
\therefore f\left(\mathbf{x}^0 + \theta\mathbf{d}^0\right) &= f(4\theta, 0) \\
&= (4\theta - 2)^2 + (4\theta)^3 \\
&= 64\theta^3 + 16\theta^2 - 16\theta + 4. \\
\frac{df\left(\mathbf{x}^0 + \theta\mathbf{d}^0\right)}{d\theta} &= 192\theta^2 + 32\theta - 16. \\
\frac{df\left(\mathbf{x}^0 + \theta\mathbf{d}^0\right)}{d\theta} &= 0, \\
\therefore 12\theta^2 + 2\theta - 1 &= 0 \\
\therefore \theta &= \frac{-2 \pm \sqrt{52}}{24} \\
\therefore \theta^0 &= \frac{-22 + \sqrt{52}}{24} \\
&\approx 0.217129273. \\
\therefore \mathbf{x}^1 &= \mathbf{x}^0 + \theta^0\mathbf{d}^0 \\
\therefore \mathbf{x}^1 &= (0.868517091, 0) \\
&= (b, 0).
\end{aligned}$$

Now, applying the second step of steepest descent, we have,

$$\begin{aligned}
\nabla f(\mathbf{x}^1) &= (0, -4.525931633)^T, \\
\therefore \mathbf{d}^1 &= (0, 4.525931633) \\
&= (0, a), \\
\therefore \mathbf{x}^1 + \theta \mathbf{d}^1 &= (b, a\theta). \\
\therefore f\left(\mathbf{x}^1 + \theta \mathbf{d}^1\right) &= f(b, a\theta) \\
&= (b-2)^2 + (b - 2a\theta)^3 \\
\frac{df\left(\mathbf{x}^0 + \theta \mathbf{d}^0\right)}{d\theta} &= -6a(b - 2a\theta)^2, \\
\frac{df\left(\mathbf{x}^0 + \theta \mathbf{d}^0\right)}{d\theta} &= 0, \\
-6a(b - 2a\theta)^2 &= 0 \\
\therefore \theta^1 &= \frac{b}{2a} \\
&= 0.095948984. \\
\therefore \mathbf{x}^2 &= \mathbf{x}^1 + \theta^1 \mathbf{d}^1 \\
\therefore \mathbf{x}^2 &= (b, 0) + \frac{b}{2a}(0, a) \\
&= \left(b, \frac{b}{2}\right) \\
&= (0.868517091, 0.4342585455).
\end{aligned}$$