LSA University of Michigan

MATH562: Continuous Optimisation
Homework 5

Name: Keegan Gyoery UM-ID: 31799451

1. Consider the function $f(\mathbf{x}) = (x_1 - 3)^2 + 3(x_2 - 2)^2$.

   a) Let $\mathbf{x}^0 = (4, 1)^T$. Firstly, the gradient of $f(\mathbf{x})$ is,
$$\nabla f(\mathbf{x}) = \begin{bmatrix} 2(x_1 - 3) \\ 6(x_2 - 2) \end{bmatrix}.$$
   Using the given $\mathbf{x}^0$, the gradient of $f(\mathbf{x})$ at $\mathbf{x}^0$ is
$$\nabla f(\mathbf{x}^0) = \begin{bmatrix} 2 \\ -6 \end{bmatrix}.$$
   Thus, the steepest descent direction is
$$\mathbf{d}^0 = \begin{bmatrix} -2 \\ 6 \end{bmatrix},$$
   and so
$$\mathbf{x}^0 + \theta \mathbf{d}^0 = \begin{bmatrix} 4 \\ 1 \end{bmatrix} + \theta \begin{bmatrix} -2 \\ 6 \end{bmatrix} = \begin{bmatrix} 4 - 2\theta \\ 1 + 6\theta \end{bmatrix}.$$
   Thus, considering $f\left(\mathbf{x}^0 + \theta \mathbf{d}^0\right)$, we have
$$\begin{aligned} f\left(\mathbf{x}^0 + \theta \mathbf{d}^0\right) &= (4 - 2\theta - 3)^2 + 3(1 + 6\theta - 2)^2 \\ &= (1 - 2\theta)^2 + 3(6\theta - 1)^2 \\ &= 1 - 4\theta + 4\theta^2 + 108\theta^2 - 36\theta + 3 \\ &= 112\theta^2 - 40\theta + 4. \end{aligned}$$
   Setting the derivative equal to 0,
$$\begin{aligned} \frac{df\left(\mathbf{x}^0 + \theta \mathbf{d}^0\right)}{d\theta} &= 0 \\ \therefore 224\theta - 40 &= 0 \\ \therefore \theta_0 &= \frac{5}{23}. \end{aligned}$$
   Thus, we get the vector $\mathbf{x}^1$ as
$$\begin{aligned} \mathbf{x}^1 &= \mathbf{x}^0 + \theta_0 \mathbf{d}^0 \\ \therefore \mathbf{x}^1 &= \begin{bmatrix} 4 \\ 1 \end{bmatrix} + \frac{5}{23} \begin{bmatrix} -2 \\ 6 \end{bmatrix} = \begin{bmatrix} 4 - 10/23 \\ 1 + 30/23 \end{bmatrix} = \frac{1}{23} \begin{bmatrix} 82 \\ 53 \end{bmatrix}. \end{aligned}$$

   b) To show that $\bar{\mathbf{d}} = (1, 1)^T$ is a descent direction, we need to show that the dot product of the gradient of f at $\mathbf{x}^0$ and the descent direction $\bar{\mathbf{d}}$ is less than 0. Considering the dot product, we have

$$\begin{aligned}
\nabla f\left(\mathbf{x}^0\right)^T \bar{\mathbf{d}} &= [2\ \ 6]\begin{bmatrix}1\\1\end{bmatrix}\\
&= 2-6\\
\therefore \nabla f\left(\mathbf{x}^0\right)^T \bar{\mathbf{d}} &= -4.
\end{aligned}$$

Applying the descent method, with an exact line search, we first calculate

$$\mathbf{x}^0 + \theta\bar{\mathbf{d}} = \begin{bmatrix}4\\1\end{bmatrix} + \theta\begin{bmatrix}1\\1\end{bmatrix} = \begin{bmatrix}4+\theta\\1+\theta\end{bmatrix}.$$

Thus, considering $f\left(\mathbf{x}^0 + \theta\bar{\mathbf{d}}\right)$, we have

$$\begin{aligned}
f\left(\mathbf{x}^0 + \theta\bar{\mathbf{d}}\right) &= (4+\theta-3)^2 + 3(1+\theta-2)^2\\
&= (1+\theta)^2 + 3(\theta-1)^2\\
&= 1 + 2\theta + \theta^2 + 3\theta^2 - 6\theta + 3\\
&= 4\theta^2 - 4\theta + 4.
\end{aligned}$$

Setting the derivative equal to 0,

$$\begin{aligned}
\frac{df\left(\mathbf{x}^0 + \theta\mathbf{d}^0\right)}{d\theta} &= 0\\
\therefore 8\theta - 4 &= 0\\
\therefore \theta_0 &= \frac{1}{2}.
\end{aligned}$$

Thus, we get the vector $\mathbf{x}^1$ as

$$\begin{aligned}
\mathbf{x}^1 &= \mathbf{x}^0 + \theta_0\bar{\mathbf{d}}\\
\therefore \mathbf{x}^1 &= \begin{bmatrix}4\\1\end{bmatrix} + \frac{1}{2}\begin{bmatrix}1\\1\end{bmatrix} = \begin{bmatrix}4+1/2\\1+1/2\end{bmatrix} = \frac{1}{2}\begin{bmatrix}9\\3\end{bmatrix}.
\end{aligned}$$

c) Setting $a(\theta) = f\left(\mathbf{x}^0 + \theta\bar{\mathbf{d}}\right)$, we get the derivative of $a(\theta)$ as,

$$\begin{aligned}
a(\theta) &= f\left(\mathbf{x}^0 + \theta\bar{\mathbf{d}}\right)\\
&= 4\theta^2 - 4\theta + 4\\
\therefore a'(\theta) &= 8\theta - 4\\
\therefore a(0) &= 4\\
\therefore a'(0) &= -4.
\end{aligned}$$

Using $\rho = \frac{1}{4}$ and $\sigma = \frac{3}{4}$, and applying Armijo's first condition, we have

$$\begin{aligned}
a(0) + \rho\theta a'(0) &\geq a(\theta)\\
\therefore 4 + \frac{1}{4}\theta(-4) &\geq 4\theta^2 - 4\theta + 4\\
\therefore 4\theta^2 - 3\theta &\leq 0\\
\therefore 0 \leq \theta &\leq \frac{3}{4} \ldots (1).
\end{aligned}$$

Applying Armijo's second condition, we have

$$\begin{aligned}
a'(\theta) &\geq \sigma a'(0) \\
\therefore 8\theta - 4 &\geq \frac{3}{4}(-4) \\
\therefore 8\theta - 1 &\geq 0 \\
\therefore \theta &\geq \frac{1}{8} \ldots (2).
\end{aligned}$$

Combining (1) and (2), we have the range $\frac{1}{8} \leq \theta \leq \frac{3}{4}$ for Armijo's stopping conditions. The length of this interval is $\frac{5}{8}$. So for three uniformly distributed points in the interval $[0,1]$, the probability that one point is not in Armijo's interval is $\frac{3}{8}$, and so the probability that all three points are not in Armijo's interval is $\frac{27}{512}$.

2. Consider the function $a(\theta) = 1 - \theta e^{-\theta}$.

   a) We get the derivative of $a(\theta)$ as,

$$\begin{aligned}
a(\theta) &= 1 - \theta e^{-\theta} \\
\therefore a'(\theta) &= \theta e^{-\theta} - e^{-\theta} \\
\therefore a(0) &= 1 \\
\therefore a'(0) &= -1.
\end{aligned}$$

   Using $\rho = \frac{1}{4}$ and $\sigma = \frac{3}{4}$, and applying Armijo's first condition, we have

$$\begin{aligned}
a(0) + \rho\theta a'(0) &\geq a(\theta) \\
\therefore 1 + \frac{1}{4}\theta(-1) &\geq 1 - \theta e^{-\theta} \\
\therefore \theta\left(e^{-\theta} - \frac{1}{4}\right) &\geq 0 \\
\therefore 0 \leq \theta &\leq \ln 4 \ldots (1).
\end{aligned}$$

   Applying Armijo's second condition, we have

$$\begin{aligned}
a'(\theta) &\geq \sigma a'(0) \\
\therefore \theta e^{-\theta} - e^{-\theta} &\geq \frac{3}{4}(-1) \\
\therefore e^{-\theta}(\theta - 1) &\geq -\frac{3}{4} \\
\therefore \theta &\geq 0.139 \ldots (2).
\end{aligned}$$

   Combining (1) and (2), we have the range $0.139 \leq \theta \leq \ln 4$ for Armijo's stopping conditions.

   b) Applying the bisection method to the interval found above, we set $\alpha_0 = 0$, which violates condition 2, and $\beta_0 = 1$. Take $\gamma_0 = \frac{\alpha_0 + \beta_0}{2} = 0.5$. $\gamma_0 = 0.5$ satisfies conditions 1 and 2, so $\theta = \gamma_0 = 0.5$ is an acceptable value.

3. Consider the function $a(\theta) = \theta^4 - 4\theta^3 + \theta^2 - 10\theta + 12$. We get the derivative of $a(\theta)$ as,

$$\begin{aligned} a(\theta) &= \theta^4 - 4\theta^3 + \theta^2 - 10\theta + 12 \\ \therefore a'(\theta) &= 4\theta^3 - 12\theta^2 + 2\theta - 10 \\ \therefore a(0) &= 12 \\ \therefore a'(0) &= -10. \end{aligned}$$

Using $\rho = \dfrac{1}{4}$ and $\sigma = \dfrac{3}{4}$, and applying Armijo's first condition, we have

$$\begin{aligned} a(0) + \rho\theta a'(0) &\geq a(\theta) \\ \therefore 12 + \frac{1}{4}\theta(-10) &\geq \theta^4 - 4\theta^3 + \theta^2 - 10\theta + 12 \\ \therefore \theta\left(\theta^3 - 4\theta^2 + \theta - \frac{15}{2}\right) &\leq 0 \\ \therefore 0 \leq \theta &\leq 4.189 \ldots (1). \end{aligned}$$

Applying Armijo's second condition, we have

$$\begin{aligned} a'(\theta) &\geq \sigma a'(0) \\ \therefore 4\theta^3 - 12\theta^2 + 2\theta - 10 &\geq \frac{3}{4}(-10) \\ \therefore 4\theta^3 - 12\theta^2 + 2\theta - \frac{5}{2} &\geq 0 \\ \therefore \theta &\geq 2.9019 \ldots (2). \end{aligned}$$

Combining (1) and (2), we have the range $2.9019 \leq \theta \leq 4.189$ for Armijo's stopping conditions. Applying the bisection method to the interval found above, we set $\alpha_0 = 0$, which violates condition 2, and $\beta_0 = 5$, which violates condition 1. Take $\gamma_0 = \dfrac{\alpha_0 + \beta_0}{2} = 2.5$. $\gamma_0 = 2.5$ violates condition 2, so set $\alpha_1 = \gamma_0 = 2.5$, and $\beta_1 = \beta_0 = 5$. Now take $\gamma_1 = \dfrac{\alpha_1 + \beta_1}{2} = 3.75$. Clearly, $\gamma_1 = 3.75$ satisfies conditions 1 and 2, so $\theta = \gamma_1 = 3.75$ is an acceptable value.