MATH562: Continuous Optimisation
Homework 4

Name: Keegan Gyoery UM-ID: 31799451

1. Consider the function $f(\mathbf{x}) = 2x_1^2 - 2x_1x_2 + x_2^2 + 2x_1 - 2x_2$, with the Steepest Descent method applied, and a sequence $\mathbf{x}^k$.

   a) If $\mathbf{x}^{2k+1} = \left(0, 1 - \dfrac{1}{5^k}\right)^T$, applying two steps of Cauchy's Steepest Descent method, we are required to show that $\mathbf{x}^{2k+3} = \left(0, 1 - \dfrac{1}{5^{k+1}}\right)^T$. Firstly, the gradient of $f(\mathbf{x}$ is

$$\nabla f(\mathbf{x}) = (4x_1 - 2x_2 + 2,\ -2x_1 + 2x_2 - 2)^T.$$

So, substituting in $\mathbf{x}^{2k+1} = \left(0, 1 - \dfrac{1}{5^k}\right)^T$, we have,

$$\begin{aligned}
\therefore \nabla f(\mathbf{x}^{2k+1}) &= \left(2 - 2\left(1 - \frac{1}{5^k}\right),\ 2\left(1 - \frac{1}{5^k}\right) - 2\right)^T \\
&= \left(\frac{2}{5^k},\ -\frac{2}{5^k}\right)^T.
\end{aligned}$$

This gives us $\mathbf{d}^{2k+1} = \left(\dfrac{2}{5^k},\ -\dfrac{2}{5^k}\right)^T$. Clearly, we have,

$$\mathbf{x}^{2k+1} + \theta\mathbf{d}^{2k+1} = \left(-\frac{2\theta}{5^k},\ \frac{5^k + 2\theta - 1}{5^k}\right)^T.$$

Consider now $f\left(\mathbf{x}^{2k+1} + \theta\mathbf{d}^{2k+1}\right)$,

$$\begin{aligned}
f\left(\mathbf{x}^{2k+1} + \theta\mathbf{d}^{2k+1}\right) &= 2\left(-\frac{2\theta}{5^k}\right)^2 - 2\left(-\frac{2\theta}{5^k}\right)\left(\frac{5^k + 2\theta - 1}{5^k}\right) + \left(\frac{5^k + 2\theta - 1}{5^k}\right)^2 \\
&\quad + 2\left(-\frac{2\theta}{5^k}\right) - 2\left(\frac{5^k + 2\theta - 1}{5^k}\right) \\
&= \frac{20}{5^{2k}}\theta^2 - \frac{8}{5^{2k}}\theta + \frac{1}{5^{2k}} \\
\therefore \frac{df\left(\mathbf{x}^{2k+1} + \theta\mathbf{d}^{2k+1}\right)}{d\theta} &= \frac{40}{5^{2k}}\theta - \frac{8}{5^{2k}}.
\end{aligned}$$

Setting the derivative to 0 to find the minimum, we have,

$$\begin{aligned}
\frac{df\left(\mathbf{x}^{2k+1} + \theta\mathbf{d}^{2k+1}\right)}{d\theta} &= 0 \\
\therefore \frac{40}{5^{2k}}\theta - \frac{8}{5^{2k}} &= 0 \\
\therefore \theta_{2k+1} &= \frac{1}{5}. \\
\therefore \mathbf{x}^{2k+2} &= \mathbf{x}^{2k+1} + \theta_{2k+1}\mathbf{d}^{2k+1} \\
&= \left(-\frac{2}{5^{k+1}},\ 1 - \frac{3}{5^{k+1}}\right)^T.
\end{aligned}$$

Applying the next iteration of the Steepest Descent Method, we have

$$
\begin{aligned}
\therefore \nabla f(\mathbf{x}^{2k+2}) &= \left(-\frac{8}{5^{k+1}} - 2 + \frac{6}{5^{k+1}} + 2,\ \frac{4}{5^{k+1}} + 2 - \frac{6}{5^{k+1}} - 2\right)^T \\
&= \left(-\frac{2}{5^{k+1}},\ -\frac{2}{5^{k+1}}\right)^T.
\end{aligned}
$$

This gives us $\mathbf{d}^{2k+2} = \left(\frac{2}{5^{k+1}},\ \frac{2}{5^{k+1}}\right)^T$. Clearly, we have,

$$
\mathbf{x}^{2k+2} + \theta\mathbf{d}^{2k+2} = \left(\frac{2\theta - 2}{5^{k+1}},\ \frac{5^{k+1} + 2\theta - 3}{5^{k+1}}\right)^T.
$$

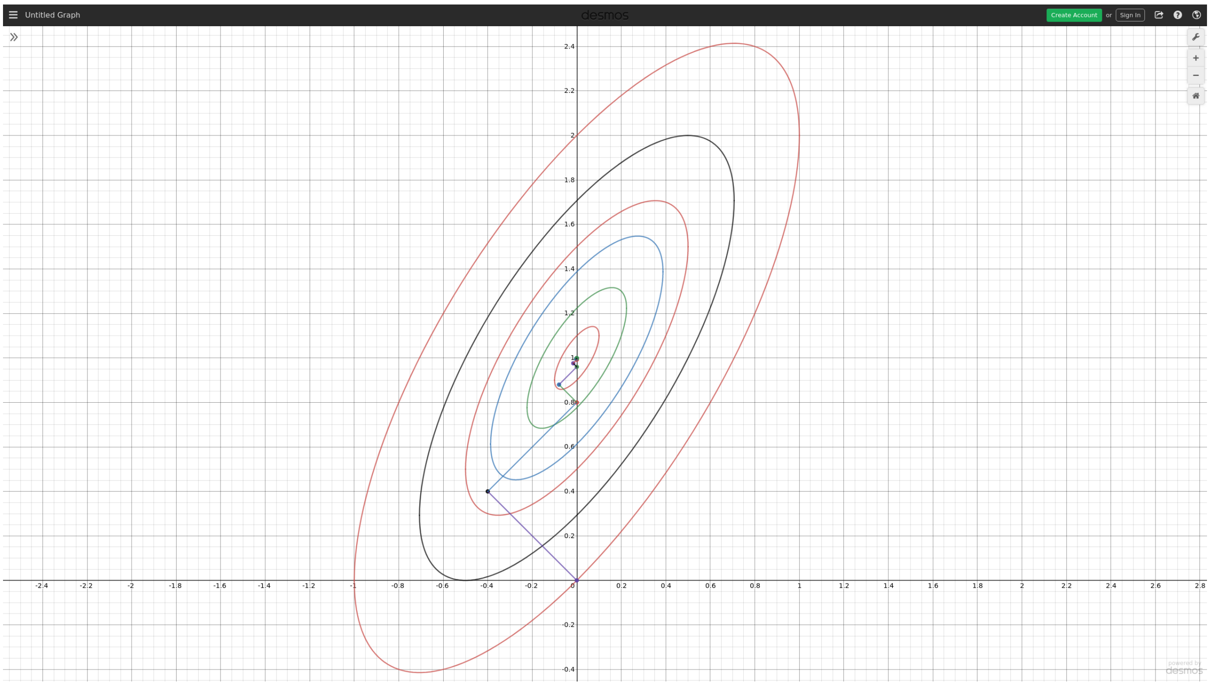
Consider now $f\left(\mathbf{x}^{2k+2} + \theta\mathbf{d}^{2k+2}\right)$,

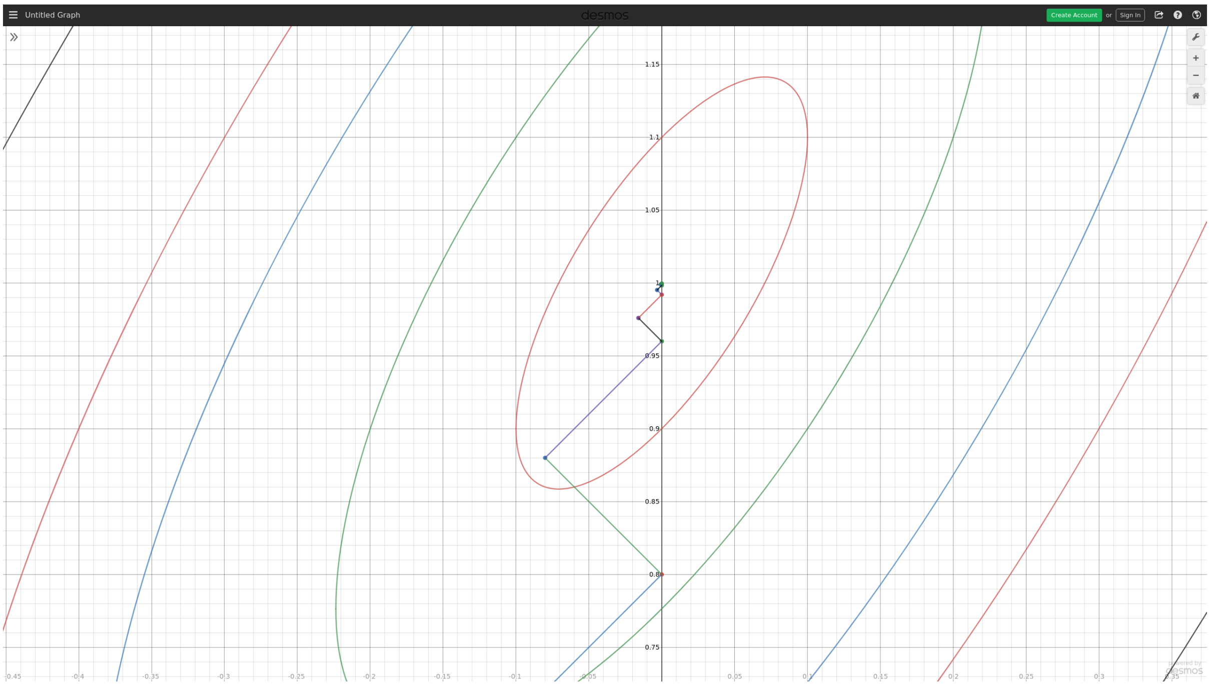
$$
\begin{aligned}
f\left(\mathbf{x}^{2k+2} + \theta\mathbf{d}^{2k+2}\right) &= 2\left(\frac{2\theta - 2}{5^{k+1}}\right)^2 - 2\left(\frac{2\theta - 2}{5^{k+1}}\right)\left(\frac{5^{k+1} + 2\theta - 3}{5^{k+1}}\right) \\
&\quad + \left(\frac{5^{k+1} + 2\theta - 3}{5^{k+1}}\right)^2 + 2\left(\frac{2\theta - 2}{5^{k+1}}\right) - 2\left(\frac{5^{k+1} + 2\theta - 3}{5^{k+1}}\right) \\
&= \frac{4}{5^{2k+2}}\theta^2 - \frac{8}{5^{2k+2}}\theta + \frac{1}{5^{2k+2}} - 1 \\
\therefore \frac{df\left(\mathbf{x}^{2k+1} + \theta\mathbf{d}^{2k+1}\right)}{d\theta} &= \frac{8}{5^{2+2k}}\theta - \frac{8}{5^{2k+2}}.
\end{aligned}
$$

Setting the derivative to 0 to find the minimum, we have,

$$
\begin{aligned}
\frac{df\left(\mathbf{x}^{2k+1} + \theta\mathbf{d}^{2k+1}\right)}{d\theta} &= 0 \\
\therefore \frac{8}{5^{2+2k}}\theta - \frac{8}{5^{2k+2}} &= 0 \\
\therefore \theta_{2k+2} &= 1. \\
\therefore \mathbf{x}^{2k+3} &= \mathbf{x}^{2k+2} + \theta_{2k+2}\mathbf{d}^{2k+2} \\
&= \left(0,\ 1 - \frac{1}{5^{k+1}}\right)^T.
\end{aligned}
$$

b) As seen in the included images, the Steepest Descent Method starting at $\mathbf{x}^1 = \mathbf{0}$ zig-zags towards the minimum, located at $(0, 1)^T$. Note that consecutive descent directions are perpendicular, a fact we will prove later in the assignment.





2. Assuming that $\mathbf{x}^k$ and $\mathbf{x}^{k+1}$ are consecutive points generated by the Steepest Descent Method, we have by defintion,

$$\begin{aligned}
\left.\frac{df\left(\mathbf{x}^k+\theta\mathbf{d}^k\right)}{d\theta}\right|_{\theta=\theta_k} &= 0\\
\therefore \nabla f\left(\mathbf{x}^k+\theta\mathbf{d}^k\right)^T \mathbf{d}^k\Big|_{\theta=\theta_k} &= 0\\
\therefore \nabla f\left(\mathbf{x}^k+\theta\mathbf{d}^k\right)^T\Big|_{\theta=\theta_k} \mathbf{d}^k &= 0\\
\therefore \nabla f\left(\mathbf{x}^{k+1}\right)^T \mathbf{d}^k &= 0\\
\therefore \left(-\mathbf{d}^{k+1}\right)^T \mathbf{d}^k &= 0\\
\therefore \left(\mathbf{d}^{k+1}\right)^T \mathbf{d}^k &= 0.
\end{aligned}$$

Thus, the dot product of $\mathbf{d}^k$ and $\mathbf{d}^{k+1}$ is 0, and so the descent directions of consecutive points are orthogonal.

3. Consider the function $f(\mathbf{x}) = \left(x_1^2 - x_2\right)^2 + 2\left(x_2^2 - x_1 - 4\right)^4$ with $\mathbf{x}^0 = (2,1)^T$.

   a) The gradient of $f(\mathbf{x})$ is given by,

$$\begin{aligned}
\nabla f(\mathbf{x}) &= \begin{bmatrix} 4x_1\left(x_1^2 - x_2\right) - 8\left(x_2^2 - x_1 - 4\right)^3 \\ -2\left(x_1^2 - x_2\right) + 16x_2\left(x_2^2 - x_1 - 4\right)^3 \end{bmatrix},\\
\therefore \nabla f(\mathbf{x}^0) &= (1024, -2006)^T.
\end{aligned}$$

   b) This gives us $\mathbf{d}^0 = (-1024, 2006)^T$. Clearly, we have, $\mathbf{x}^0 + \theta\mathbf{d}^0 = (2 - 1024\theta, 1 + 2006\theta)^T$. Now considering $a(\theta) = f(\mathbf{x}^0 + \theta\mathbf{d}^0)$,

$$\begin{aligned}
a(\theta) &= f(\mathbf{x}^0 + \theta\mathbf{d}^0)\\
&= f(2 - 1024\theta, 1 + 2006\theta)\\
\therefore a(\theta) &= \left[(2 - 1024\theta)^2 - (1 + 2006\theta)\right]^2 + 2\left[(1 + 2006\theta)^2 - (2 - 1024\theta) - 4\right]^4
\end{aligned}$$

   c) From $a(\theta)$ above, we calculate the derivative as

$$\begin{aligned}
a(\theta) &= \left[(2 - 1024\theta)^2 - (1 + 2006\theta)\right]^2 + 2\left[(1 + 2006\theta)^2 - (2 - 1024\theta) - 4\right]^4\\
\therefore a'(\theta) &= 2\left[2(-1024)(2 - 1024\theta) - 2006\right]\left[(2 - 1024\theta)^2 - (1 + 2006\theta)\right]\\
&\qquad + 2 \times 4\left[2 \times 2006(1 + 2006\theta) + 1024\right]\left[(1 + 2006\theta)^2 - (2 - 1024\theta) - 4\right]^3\\
\therefore a'(0) &= 2\left[2(-1024) \times 2 - 2006\right]\left[(2)^2 - (1)\right] + 2 \times 4\left[2 \times 2006 + 1024\right]\left[(1)^2 - (2) - 4\right]^3\\
&= (-12204) \times 3 + 40288 \times (-125)\\
&= -5072612
\end{aligned}$$

$\therefore a'(\theta) < 0.$

d) Let $\beta = 1$. Consider $a'(\beta)$.

$$\begin{aligned}
a'(\beta) &= a'(1) \\
&= 2\left[2(-1024)(2-1024) - 2006\right]\left[(2-1024)^2 - (1+2006)\right] \\
&\quad + 2 \times 4\left[2 \times 2006(1+2006) + 1024\right]\left[(1+2006)^2 - (2-1024) - 4\right]^3 \\
&\approx 4.21 \times 10^{27} \\
\therefore a'(\beta) &> 0.
\end{aligned}$$

e) The code prints out the necessary variables to track, and may be viewed when run.

f) The code prints out the necessary variables to track, and may be viewed when run.

g) Using the MATLAB code for the golden section method, and with a smaller epsilon, we obtain $\theta_0 = 0.0006$. Thus, using $\mathbf{x}^0 = (2,1)^T$ and $\mathbf{d}^0 = (-1024, 2006)^T$, applying the one step of Steepest Descent Method gives us

$$\begin{aligned}
\mathbf{x}^1 &= \mathbf{x}^0 + \theta_0 \mathbf{d}^0 \\
&= (1.3856, 2.2036)^T .
\end{aligned}$$