# $2^{nd}$ Assingment - Optimization - Dafni Tziakouri

## Exercise 2.1:

(1) Verify that $p_k$ is indeed a descent direction, that is, $f(x_{k+1}) < f(x_k)$, $f(x) = x^2$

$x_0 = 2$, $p_k = -\text{sign}(x_k)$, $a_k = 2 + 3(2^{-k-1})$

I will use the induction method:

- Valid for $k=0$: $x_0 = 2/1 = 2$
- Suppose that is valid for $k=n$: $x_n = (-1)^n \dfrac{2^n+1}{2^n}$
- Let's show that for $k = n+1 \Rightarrow x_{n+1} = (-1)^{n+1} \dfrac{2^{n+1}+1}{2^{n+1}}$

So, $x_{n+1} = x_n + a_n p_n$ where $p_n = (-1)^{n+1}$

$$\Rightarrow x_{n+1} = (-1)^n \frac{2^n+1}{2^n} + (-1)^{n+1}\left[2 + \frac{3}{2^{n+1}}\right] = \frac{(-1)^n}{2^{n+1}}\left[2^{n+1} + 2 - 2^{n+2} - 3\right] = \frac{(-1)^n}{2^{n+1}}\left[-2^{n+1} - 1\right] = (-1)^{n+1} \cdot \frac{2^{n+1}+1}{2^{n+1}}$$

$$\text{So, } f(x_k) - f(x_{k+1}) = \frac{2^{2k}+1+2^{k+1}}{2^{2k}} - \frac{2^{2k+2}+1+2^{k+2}}{2^{2k+2}} = \frac{2^{2k+2}+4+2^{k+3}-2^{2k+2}-1-2^{k+2}}{2^{2k+1}}$$

$$= \frac{3+2^{k+2}}{2^{2k+2}} > 0, \quad k = 0, 1, 2, \dots \blacksquare$$

(2) Perform 5 steps of the descent algorithm

step 1 $(k=0)$: $a_0 = 2 + 3/2 = 7/2$

$x_1 = x_0 + a_0 p_0 = 2 + 7/2(-1) = -3/2$

step 2 $(k=1)$: $a_1 = 2 + 3/4 = 11/4$

$x_2 = x_1 + a_1 p_1 = -3/2 + 11/4(1) = 5/4$

step 3 $(k=2)$: $a_2 = 2 + 3/8 = 19/8$

$x_3 = x_2 + a_2 p_2 = 5/4 - 19/8 = -9/8$

step 4 $(k=3)$: $a_3 = 2 + 3/16 = 35/16$

$x_4 = x_3 + a_3 p_3 = -9/8 + \dfrac{35}{16} = 17/16$

step 5 $(k=4)$: $a_4 = 2 + 3/32 = 67/32$

$x_5 = x_4 + a_4 p_4 = 17/16 + \dfrac{67}{32}(-1) = -33/32$

$\rightarrow$

(3) Does this descent converge? What Wolfe conditions are violated?

The given method may not converge and it likely violates the Wolfe conditions, primarly due to the increasing step size and oscillations in the descent direction. Probably, ussing a different step size or a different optimization method, might be better to ensure convergence for this specific function.

## Exercise 2.2:

$$m_k(p_k) = f_k + G_k \cdot p_k + \frac{1}{2} p_k^T \cdot B_k \cdot p_k \to \min \qquad \|p_k\| < \delta$$

Let $p_k^c = \tau_k p_k^\ell$ be the Cauchy Point, where

$$p_k^\ell = \arg \min_{\substack{p \in \mathbb{R}^n \\ \|p\| < \delta}} (f_k + G_k \cdot p), \quad \tau_k = \arg \min_{\substack{\tau \in \mathbb{R} \\ \|\tau p_k^\ell\| < \delta}} m_k(\tau p_k^\ell)$$

(1) Show that $p_k^\ell = -\frac{\delta}{\|G_k\|} G_k$

Let's first define the variables:

$p_k^c$: cauchy point $\qquad$ $f_k$: value of function at $x_k$

$G_k$: gradient vector at $x_k$

I will differentiate with respect to $p$ the $m_k(p_k)$ and set the derivative equal to zero and then solve for $p_k^\ell$.

→

$\nabla(f_k + G_k \cdot p) = 0 \iff \nabla f_k + G_k = 0$

$\Rightarrow p_k^e = -\nabla f_k - G_k$ , $\|G_k\| = \sqrt{G_k \cdot G_k}$

$\Rightarrow p_k^e = -(\nabla f_k + G_k)$

$\Rightarrow p_k^e = -\frac{\delta}{\|G_k\|}\left(\|G_k\| \cdot (\nabla f_k + G_k)\right)$

where $\|G_k\| \cdot (\nabla f_k + G_k) = G_k$

$\Rightarrow p_k^c = -\frac{\delta}{\|G_k\|} \cdot G_k$ ■

(2) $\tau_k = \begin{cases} 1 & , G_k \cdot B_k \cdot G_k^T \leq 0 \\ \min\{1, \hat{\tau}_k\} & , \text{otherwise} \end{cases}$ where $\hat{\tau}_k = \frac{\|G_k\|^3}{\delta\, G_k \cdot B_k \cdot G_k^T}$

$\tau_k$ takes a speficic value depending on the sign of $G_k \cdot B_k \cdot G_k^T$:

case 1: If $G_k \cdot B_k \cdot G_k^T \leq 0$ the trust region problem is not stricly convex or have no local minimum within the trust region. Therefore, we set $\tau_k = 1$ so we will ensure progress.

case 2: If $G_k \cdot B_k \cdot G_k^T > 0$ the problem is stricly convex and has a unique minimum within the trust region. In this case we want to calculate $\tau_k$ s.t. it maximizes the decrease in ~~ob~~ the objective function. That's why we consider $\tau_k = \min\{1, \hat{\tau}_k\}$ , $\hat{\tau}_k = \frac{\|G_k\|^3}{\delta\, G_k \cdot B_k \cdot G_k^T}$ →

Here, $\hat{\tau}_k$ is the step size that maximizes the decrease in the objective function while staying within the trust region.

If $\hat{\tau}_k \leq 1$ : we take step of size $\hat{\tau}_k$

If $\hat{\tau}_k > 1$ : we take 1 to ensure we stay in the trust region.

## Exercise 2.3:

Implement 2 steps of Cauchy point search for Rosenbrock function $f(x_1, x_2) = (1-x_1)^2 + 5(x_2 - x_1^2)^2$ with starting point at $(-2, -2)$ and trust region being balls with $r = 0.5$.

First, I will find the gradient of $f$ at $(-2,-2)$:

$$\nabla f(x_1, x_2) = (-2(1-x_1) + 10(x_2 - x_1^2)(-2x_1),\ 10(x_2 - x_1^2))$$

$$\nabla f(-2,-2) = (-246, -60)$$

Now, the Hessian matrix at $(-2,-2)$

$$\frac{\partial^2 f}{\partial x_1^2} = 2 - 20x_1(-2x_1) - 20(x_2 - x_1^2)$$

$$= 2 + 40x_1^2 - 20x_2 + 20x_1^2 = 2 + 60x_1^2 - 20x_2$$

$$\frac{\partial^2 f}{\partial x_2^2} = 10 \qquad \frac{\partial^2 f}{\partial x_1 \partial x_2} = -20x_1 \qquad \frac{\partial^2 f}{\partial x_2 \partial x_1} = -20x_1$$

$\rightarrow$

$$\Rightarrow Hf(x_1, x_2) = \begin{pmatrix} 2+60x_1^2-20x_2 & -20x_1 \\ -20x_1 & 10 \end{pmatrix}$$

$$\Rightarrow Hf(-2,-2) = \begin{pmatrix} 282 & 40 \\ 40 & 10 \end{pmatrix}$$

Also, $f(-2,-2) = 9 + 180 = 189$

Let's use the ex. 2.2:

- $f_1 = f(x_0) = 189$ • $G_1 = \nabla f(x_0) = (-246, -60)$
- $B_1 = Hf(x_0) = \begin{pmatrix} 282 & 40 \\ 40 & 10 \end{pmatrix}$ • $\delta = 0.5$

$$\Rightarrow P_1^c = -\frac{0.5}{\|G_1\|} G_1 = -\frac{0.5}{\sqrt{(-246)^2+(-60)^2}}(-246, -60)$$

$$\Rightarrow P_1^c \simeq -1.97\times10^{-3}(-246,-60) \simeq (0.4858, 0.1185)$$

Also, $G_1^T B_1 G_1 = \underset{1\times2}{(-246, -60)} \begin{pmatrix} 282 & 40 \\ 40 & 10 \end{pmatrix} \begin{pmatrix} -246 \\ -60 \end{pmatrix}$

$$= (-71772 \quad -10440) \begin{pmatrix} -246 \\ -60 \end{pmatrix} = 18282312 > 0$$

$$\Rightarrow \hat{\tau}_1 = \frac{\|G_1\|^3}{0.5 G_1^T B_1 G_1} = 1.77 > 1 \Rightarrow \boxed{\tau_1 = 1}$$

$$\Rightarrow P_1^c = (0.4858, 0.1185) \Rightarrow X_1 = (-1.5142, -1.8815)$$

- $f_2 = f(X_1) = 93.45$ • $G_2 = \nabla f(X_1) = (-131.45, -41.74)$
- $B_2 = Hf(X_1) = \begin{pmatrix} 177 & 30.28 \\ 30.28 & 10 \end{pmatrix}$

$\longrightarrow$

$$\Rightarrow p_2^{c} = -\frac{0.5}{\|G_2\|} G_2 = -\frac{0.5}{\sqrt{(-131.45)^2+(-41.74)^2}}(-131.45, -41.74)$$

$$= (0.4765, 0.1513)$$

Also, $G_2^T B_2 G_2 = (-131.45, -41.74)\begin{pmatrix} 177 & 30.28 \\ 30.28 & 10 \end{pmatrix}\begin{pmatrix} -131.45 \\ -41.74 \end{pmatrix}$

$$= 3.411.779 > 0$$

$\Rightarrow \hat{\tau}_2 = 1.5379 > 1 \Rightarrow \tau_2 = 1 \Rightarrow p_2^c = (0.4765, 0.1513)$

$\Rightarrow X_2 = (-1.0377, -1.7302)$ with $f(X_2) = 43.548$