最佳化理論HW2. 人工智慧學程. 李少琪 0860908

8.1 Perform two iterations leading to the minimization of $f(x_1,x_2)=x_1+\frac{1}{2}x_2+\frac{1}{2}x_1^2+x_2^2+3$ using the steepest desent method with the starting point $x^{(0)}=0$. Also determine an optimal solution analytically.

Ans. $x=\begin{bmatrix}x_1\\x_2\end{bmatrix}$ $f(x)=x^T\begin{bmatrix}1\\\frac{1}{2}\end{bmatrix}+x^T\begin{bmatrix}\frac{1}{2}&0\\0&1\end{bmatrix}x+3$

$$=\frac{1}{2}x^T\begin{bmatrix}1&0\\0&2\end{bmatrix}x-x^T\begin{bmatrix}-1\\-\frac{1}{2}\end{bmatrix}+3$$

$$=\frac{1}{2}x^TQx-x^Tb+c \rightarrow Q=\begin{bmatrix}1&0\\0&2\end{bmatrix}\quad b=\begin{bmatrix}-1\\-\frac{1}{2}\end{bmatrix}\quad c=3$$

1st:

$$x^{(1)}=x^{(0)}-\alpha_0\nabla f(x^{(0)}) \qquad \nabla f(x^{(0)})=g^{(0)}=Qx^{(0)}-b=\begin{bmatrix}1\\\frac{1}{2}\end{bmatrix}$$

$$=x^{(0)}-\frac{g^{(0)T}g^{(0)}}{g^{(0)T}Qg^{(0)}}g^{(0)} \qquad \alpha_0=\frac{g^{(0)T}g^{(0)}}{g^{(0)T}Qg^{(0)}}=\frac{[1\ \frac{1}{2}]\begin{bmatrix}1\\\frac{1}{2}\end{bmatrix}}{[1\ \frac{1}{2}]\begin{bmatrix}1&0\\0&2\end{bmatrix}\begin{bmatrix}1\\\frac{1}{2}\end{bmatrix}}=\frac{\frac{5}{4}}{\frac{3}{2}}=\frac{5}{6}$$

$$=\begin{bmatrix}0\\0\end{bmatrix}-\frac{5}{6}\begin{bmatrix}1\\\frac{1}{2}\end{bmatrix}=\begin{bmatrix}-\frac{5}{6}\\-\frac{5}{12}\end{bmatrix}$$

2nd:

$$x^{(2)}=x^{(1)}-\alpha_1\nabla f(x^{(1)}) \qquad \nabla f(x^{(1)})=g^{(1)}=Qx^{(1)}-b=\begin{bmatrix}1&0\\0&2\end{bmatrix}\begin{bmatrix}-\frac{5}{6}\\-\frac{5}{12}\end{bmatrix}-\begin{bmatrix}-1\\-\frac{1}{2}\end{bmatrix}=\begin{bmatrix}\frac{1}{6}\\-\frac{1}{3}\end{bmatrix}$$

$$=\begin{bmatrix}-\frac{5}{6}\\-\frac{5}{12}\end{bmatrix}-\frac{5}{9}\begin{bmatrix}\frac{1}{6}\\-\frac{1}{3}\end{bmatrix} \qquad \alpha_1=\frac{g^{(1)T}g^{(1)}}{g^{(1)T}Qg^{(1)}}=\frac{[-\frac{5}{6}\ -\frac{5}{12}]\begin{bmatrix}-\frac{5}{6}\\-\frac{5}{12}\end{bmatrix}}{[-\frac{5}{6}\ -\frac{5}{12}]\begin{bmatrix}1&0\\0&2\end{bmatrix}\begin{bmatrix}-\frac{5}{6}\\-\frac{5}{12}\end{bmatrix}}=\frac{5}{9}$$

$$=\begin{bmatrix}-\frac{25}{27}\\-\frac{25}{108}\end{bmatrix}$$

optimal: $Qx^*-b=0$ $\qquad x^*=Q^{-1}b$

$Qx^*=b$ $\qquad =\frac{1}{2}\begin{bmatrix}2&0\\0&1\end{bmatrix}\begin{bmatrix}-1\\-\frac{1}{2}\end{bmatrix}$

$$=\frac{1}{2}\begin{bmatrix}-2\\-\frac{1}{2}\end{bmatrix}=\begin{bmatrix}-1\\-\frac{1}{4}\end{bmatrix}\#$$

8.6 support that we use the golden section algorithm to find the minimizer of a function. Let $u_k$ be the uncertainty range at the $k^{th}$ iteration. Find the order of convergence of $\{u_k\}$.

Ans. $u_k\rightarrow 0$ $\qquad \lim_{k\to\infty}\frac{|u_{k+1}|}{|u_k|}=1-\rho>0. \rightarrow 1\#.$

$u_{k+1}=(1-\rho)u_k$

8.18 Let $f:\mathbb{R}^n\rightarrow\mathbb{R}$ be given by $f(x)=\frac{1}{2}x^TQx-x^Tb$, where $b\in\mathbb{R}^n$ and $Q$ is a real symmetric positive definite $n\times n$ matrix. Suppose that we apply the steepest descent method to this function, with $x^{(0)}\neq Q^{-1}b$. Show that the method converges in one step, that is, $x^{(1)}=Q^{-1}b$, if and only if $x^{(0)}$ is chosen such that $g^{(0)}=Qx^{(0)}-b$ is an eigenvector of $Q$.

Ans. $x^{(1)}=x^{(0)}-\alpha_0g^{(0)}=x^{(0)}-\frac{g^{(0)T}g^{(0)}}{g^{(0)T}Qg^{(0)}}g^{(0)}$

$x^{(1)}=Q^{-1}b=x^{(0)}-\alpha_0g^{(0)}$

$\rightarrow Qx^{(0)}-b=\alpha_0Qg^{(0)}$

$\because x^{(0)}\neq Q^{-1}b\ \therefore g^{(0)}=Qx^{(0)}-b\neq 0.$

$\rightarrow g^{(0)}=\alpha_0Qg^{(0)}$

$Qg^{(0)}=\frac{1}{\alpha_0}g^{(0)}$ : $g^{(0)}$是Q的eigenvector. $\frac{1}{\alpha_0}$是eigenvalue.

$$Qx^{(1)}=Q\left(x^{(0)}-\frac{g^{(0)T}g^{(0)}}{g^{(0)T}Qg^{(0)}}g^{(0)}\right)$$

$$=Qx^{(0)}-\alpha_0\frac{g^{(0)T}g^{(0)}}{\cancel{g^{(0)T}g^{(0)}}}Qg^{(0)}$$

$$=Qx^{(0)}-g^{(0)}$$

$$=b \qquad \rightarrow x^{(1)}=Q^{-1}b.$$

8.24. Given $f: \mathbb{R}^n \to \mathbb{R}$, consider the general iterative algorithm $x^{(k+1)} = x^{(k)} + \alpha_k d^{(k)}$, where $d^{(1)}, d^{(2)}$, are given vectors in $\mathbb{R}^n$ and $\alpha_k$ is chosen to minimize $f(x^{(k)} + \alpha d^{(k)})$; that is, $\alpha_k = \arg\min f(x^{(k)} + \alpha d^{(k)})$. Show that for each k, the vector $x^{(k+1)} - x^{(k)}$ is orthogonal to $\nabla f(x^{(k+1)})$ (assuming that the gradient exists)

Ans. $x^{(k+1)} = x^{(k)} + \alpha_k d^{(k)} \to x^{(k+1)} - x^{(k)} = \alpha_k d^{(k)}$

$\langle x^{(k+1)} - x^{(k)}, \nabla f(x^{(k+1)}) \rangle = \alpha_k \langle d^{(k)}, \nabla f(x^{(k+1)}) \rangle$

令 $\phi_k(\alpha) = f(x^{(k)} + \alpha d^{(k)})$. $\alpha_k = \arg\min \phi_k$ $\because$ FONC $\therefore \phi_k'(\alpha_k) = 0$

$\phi_k'(\alpha_k) = d^{(k)T} \nabla f(x^{(k)} + \alpha_k d^{(k)}) = \langle d^{(k)}, \nabla f(x^{(k+1)}) \rangle = 0$

$\to \langle x^{(k+1)} - x^{(k)}, \nabla f(x^{(k+1)}) \rangle = \alpha_k \langle d^{(k)}, \nabla f(x^{(k+1)}) \rangle = 0 \to$ orthogonal

9.1 Let $f: \mathbb{R} \to \mathbb{R}$ ~~by~~ given by $f(x) = (x - x_0)^4$, where $x_0 \in \mathbb{R}$ is a constant. Suppose that we apply Newton's method to the problem of minimizing $f$.

(a.) Write down the update equation for Newton's method applied to the problem.

Ans. $f'(x) = 4(x-x_0)^3 \xrightarrow{\text{Newton}} x^{(k+1)} = x^{(k)} - \frac{x^{(k)} - x_0}{3}$ #.

$f''(x) = 12(x-x_0)^2$

(b) Let $y^{(k)} = |x^{(k)} - x_0|$, where $x^{(k)}$ is the $k^{th}$ iterate in Newton's method. Show that the sequence $\{y^{(k)}\}$ satisfies $y^{(k+1)} = \frac{2}{3} y^{(k)}$

Ans. $x^{(k+1)} - x_0 = \frac{2}{3}(x^{(k)} - x_0)$

$y^{(k)} = |x^{(k)} - x_0| = \frac{2}{3}|x^{(k-1)} - x_0| = \frac{2}{3} y^{(k-1)} \to y^{(k+1)} = \frac{2}{3} y^{(k)}$

(c.) Show that $x^{(k)} \to x_0$ for any initial guess $x^{(0)}$.

Ans. (b) $\to y^{(k)} = (\frac{2}{3})^k y^{(0)} \to 0$.

$|x^{(k)} - x_0| \to 0 \quad \therefore x^{(k)} \to x_0$ for any $x^{(0)}$

(d). Show that the order of convergence of the sequence $\{x^{(k)}\}$ in part b. is 1.

Ans. $\lim_{k \to \infty} \frac{|x^{(k+1)} - x_0|}{|x^{(k)} - x_0|} = \frac{2}{3} > 0$

(e.) Theorem 9.1 states that under certain conditions, the order of convergence of Newton's method is at least 2. Why does that theorem not hold in this particular problem?

Ans. theorem 是假設 $f''(x^*) \neq 0$

9.4 Consider Rosenbrock's Function: $f(x) = 100(x_2 - x_1^2)^2 + (1-x_1)^2$, where $x = [x_1, x_2]^T$ (known to be a "nasty" function – often used as a benchmark for testing algorithms.) This function is also known as the banana function because of ~~the~~ its level sets.

(a) Prove the $[1,1]^T$ is the unique global minimizer of f over $\mathbb{R}^2$.

Ans. $f(x) = 100(x_2 - x_1^2)^2 + (1-x_1)^2 \geq 0$.

$f(x^*) = 0. \to \begin{cases} x_2 - x_1^2 = 0 \\ 1 - x_1 = 0 \end{cases} \to x = [1,1]^T$

$f(x) > f([1,1]^T)$ for all $x \neq [1,1]^T$

$\to [1,1]^T$ 是 unique global minimizer.

(b) With a starting point of $[0,0]^T$, apply two iterations of Newton's method.

Hint: $\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$

Ans. $\nabla f(x) = \begin{bmatrix} \frac{\partial f(x)}{x_1} \\ \frac{\partial f(x)}{x_2} \end{bmatrix} = \begin{bmatrix} 400x_1^3 - 400x_1x_2 + 2x_1 - 2 \\ 200(x_2 - x_1^2) \end{bmatrix}$

$F(x) = \begin{bmatrix} 1200x_1^2 - 400x_2 + 2 & -400x_1 \\ -400x_1 & 200 \end{bmatrix}$

$F^{-1}(x) = \frac{1}{80000(x_1^2 - x_2) + 400} \begin{bmatrix} 200 & 400x_1 \\ 400x_1 & 1200x_1^2 - 400x_2 + 2 \end{bmatrix}$

$x^{(1)} = x^{(0)} - F(x^{(0)})^{-1} \nabla f(x^{(0)})$

$= [1,0]^T$ #

$x^{2} = x^{(1)} - F(x^{(1)})^{-1} \nabla f(x^{(1)})$

$= [1,1]^T$ #.

1. (Adopted from [88, Exercise 9.8(1)]) Let $Q$ be a real symmetric positive definite $n\times n$ matrix. Given an arbitrary set of linearly independent vector $\{p^{(0)},\cdots,p^{(n-1)}\}$ in $\mathbb{R}^n$, the Gram-Schmidt procedure generates a sets of vectors $\{d^{(0)},\cdots,d^{(n-1)}\}$ as follow:

$$d^{(0)}=p^{(0)}$$

$$d^{(k+1)}=p^{(k+1)}-\sum_{i=0}^{k}\frac{p^{(k+1)T}Qd^{(i)}}{d^{(i)T}Qd^{(i)}}d^{(i)}$$

Show that the vectors $d^{(0)},\cdots,d^{(n-1)}$ are $Q$-conjugate.

Ans: by induction.

假設 $d^{(i)}\neq 0$, $i=1,\cdots,n-1$. $\rightarrow d^{(i)T}Qd^{(i)}\neq 0$.

$i=0$, ok.

假設 $i<n-1$ 都成立. $\rightarrow$ $\{d^{(0)},\cdots,d^{(i)}\}$ : $Q$-conjugate.

$i=n-1$

$$d^{(i)T}Qd^{(j)}=\left(p^{(i)T}-\sum_{a=1}^{i-1}\frac{p^{(i)T}Qd^{(a)}}{d^{(a)T}Qd^{(a)}}d^{(a)T}\right)Qd^{(j)}$$

$$=p^{(i)T}Qd^{(j)}-\sum_{i=0}^{i-1}\frac{p^{(i)T}Qd^{(a)}}{d^{(a)}Qd^{(a)}}d^{(a)T}Qd^{(j)}=0.$$

$\rightarrow d^{(i)T}Qd^{(j)}=0$, for $i\neq j$.

by induction: $\underset{\text{(nonzero)}}{d^{(k)}}$ is a linear combination of $p^{(0)},\cdots,p^{(k)}$

$k=0$. $d^{(0)}=p^{(0)}$ 成立.

假設 $k<n-1$ 成立. $d^{(k)}=\sum_{j=0}^{k}\alpha_j^{(k)}p^{(j)}$, $\alpha_j\neq 0$.

$$\rightarrow d^{(k+1)}=p^{(k+1)}-\sum_{i=0}^{k}\beta_i d^{(i)}$$

$$=p^{(k+1)}-\sum_{i=0}^{k}\beta_i\sum_{j=0}^{i}\alpha_j^{(k)}p^{(j)}$$

$$=p^{(k+1)}-\sum_{i=0}^{k}\sum_{j=0}^{i}\beta_i\alpha_j^{(k)}p^{(j)}$$

$\rightarrow d^{(k+1)}$: nonzero linear combination of $p^{(0)},\cdots,\underset{\uparrow\atop n-1}{p^{(k+1)}}$

```python
import numpy as np

def ConjugateGradientAlgo(lite, Q, b, c, x):
    """
    f(x) = 0.5 * X.T * Q * X - X.T * b + c
    """
    g = np.matmul(Q, x) - b
    d = g
    for _ in range(lite):
        print(x, d, g)
        alpha = - (g.T * d) / (np.matmul(np.matmul(d.T, Q), d))
        x = x + alpha * d
        g = np.matmul(Q, x) - b
        beta = (np.matmul(g.T, Q) * d) / (np.matmul(np.matmul(d.T, Q), d))
        d = - g + np.matmul(beta, d)

Q = np.array([[5, -3], [-3, 2]])
b = np.array([0,1])
c = -7
x = np.array([0, 0])
lite = 3
ConjugateGradientAlgo(lite, Q, b, c, x)
```

```
python3 10_9.py
[0 0] [ 0 -1] [ 0 -1]
[0.    0.5] [3.75 2.25] [-1.5    0. ]
[0.70754717 0.5         ] [4.01299395 8.17337131] [ 2.03773585 -2.12264151]
```