Midterm for Introduction to Optimization

By Shun-Feng Su

April 15, 2022

General Guideline:

1. **There are 13 questions in this exam. Please do all questions. The total score is 400. Do your best and good luck.**
2. **Your answers can be written either in English or in Chinese.**
3. **Please give proper definition for the symbols you use.**
4. **Write all answers on the blank place immediately following the questions. If there is no enough space, continue your answers on the back of the sheet with proper indications.**
5. **Partial grade may be given. Write down any derivation to demonstrate your knowledge in case your answer is incorrect.**
6. **If the problem is infeasible, please state why it is infeasible.**

Your Name: 1101 助教

Student ID: ______________________

Problem Number	Score
I. (25pts)	
II. (25pts)	
III. (10pts)	
IV. (30pts)	
V. (30pts)	
VI. (15pts)	
VII. (30pts)	
VIII. (20pts)	
IX. (25 pts)	
X. (60 pts)	
XI. (20 pts)	
XII. (60 pts)	
XIII. (50pts)	
Total: (400pts)	

I. Define $f(x_1,x_2,x_3,x_4)=x_1^2+2x_1x_2^2+x_2^3+2x_2^2x_3+x_3^4+2x_2x_4+x_4^2$ and let $\boldsymbol{d}=[2,\ 1,\ 2,\ 0]^T$.

(a) (10 points) Please find the directional derivative of *f* on the direction ***d***.

(b) (5 points) Also find the rate of increase of *f* at the point $(1,\ 0,\ 1,\ 1)^T$ on the direction ***d***.

(c) (10 points) Find the gradient and the Hessian matrix of *f* at the point $(1,\ 0,\ 0,\ 1)^T$.

(a) $\nabla f(x)^T \dfrac{d}{\|d\|}$

$$= \begin{bmatrix} 2x_1+2x_2^2 \\ 4x_1x_2+3x_2^2+4x_2x_3+2x_4 \\ 2x_2^2+4x_3^3 \\ 2x_2+2x_4 \end{bmatrix}^T \begin{bmatrix} 2 \\ 1 \\ 2 \\ 0 \end{bmatrix} \times \frac{1}{\sqrt{9}}$$

$$= \frac{1}{3}(4x_1+11x_2^2+4x_1x_2+4x_2x_3+2x_4+8x_3^3) \ \#$$

(b)

$$\begin{bmatrix} 2 \\ 2 \\ 4 \\ 2 \end{bmatrix}^T \begin{bmatrix} 2 \\ 1 \\ 2 \\ 0 \end{bmatrix} \times \frac{1}{\sqrt{9}} = \frac{14}{3} \ \#$$

(c)

$$\text{gradient} = \begin{bmatrix} 2 \\ 2 \\ 0 \\ 2 \end{bmatrix} \ \# \qquad H(x) = \begin{bmatrix} 2 & 4x_2 & 0 & 0 \\ 4x_2 & 4x_1+6x_2+4x_3 & 4x_2 & 2 \\ 0 & 4x_2 & 12x_3^2 & 0 \\ 0 & 2 & 0 & 2 \end{bmatrix}$$

$$H(1,0,0,1) = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 4 & 0 & 2 \\ 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 2 \end{bmatrix} \ \#$$

II. Consider the problem

Minimize $2x_1^2 + x_2^2 - 2x_1 - 4x_2 + 3$

Subject to $x_1 \le 1$ and $x_2 \ge 1$.

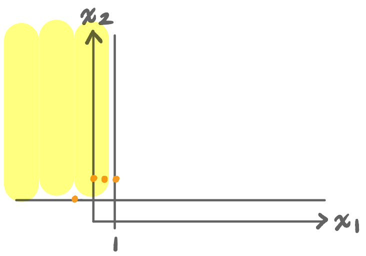
(a) (10 points) Please check whether the following points satisfy the FONC: $[1, 2]^T$, $[0.5, 2]^T$, $[0, 2]^T$, and $[-1, 1]^T$.

(b) (10 points) Whether the directional derivative of f at $[0, 1]^T$ on the direction $[2, 1]^T$ exists? If yes, find it; if not, prove it.

(c) (5 points) Which directions $[-1, 1]^T$, $[1, 1]^T$, $[1, -1]^T$ or $[-1, -1]^T$ are feasible directions for the point $[1, 1]^T$?

(a) FONC: $d^T \nabla f(x) \ge 0$

$$\nabla f(x) = \begin{bmatrix} 4x_1 - 2 \\ 2x_2 - 4 \end{bmatrix}$$



$[1,2]^T$ boundary

$$\begin{bmatrix} 2 \\ 0 \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = 2d_1$$ No ※

$[0.5, 2]^T$ in

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = 0$$ Yes ※

$[0, 2]^T$ in

$$\begin{bmatrix} -2 \\ 0 \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = -2d_1$$ No ※

$[-1, 1]^T$ boundary

$$\begin{bmatrix} -6 \\ -2 \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = -6d_1 - 2d_2$$ No ※

(b)
$$\nabla f(x)^T \frac{d}{\|d\|}$$
$$= [4x_1 - 2 \quad 2x_2 - 4] \frac{\begin{bmatrix} 2 \\ 1 \end{bmatrix}}{\sqrt{5}}$$
$$= \frac{-6}{\sqrt{5}} = \frac{-6\sqrt{5}}{5}$$ ※

(c) $[1, 1]^T$ boundary

$d_1 \le 1$, $d_2 \ge 1$ $\Rightarrow$ $[-1, 1]$ 符合。

III. (10 points) As shown in the following figure, in Golden section search, by choosing the intermediate points so that a_1-a_0=b_0-b_1=$\rho(b_0$-$a_0)$ where ρ<1/2. The search can repeat the above process by using the same ρ and make one intermediate point coincide with the one already used in the previous search. Please show that in this case $\rho=\frac{3-\sqrt{5}}{2}$.

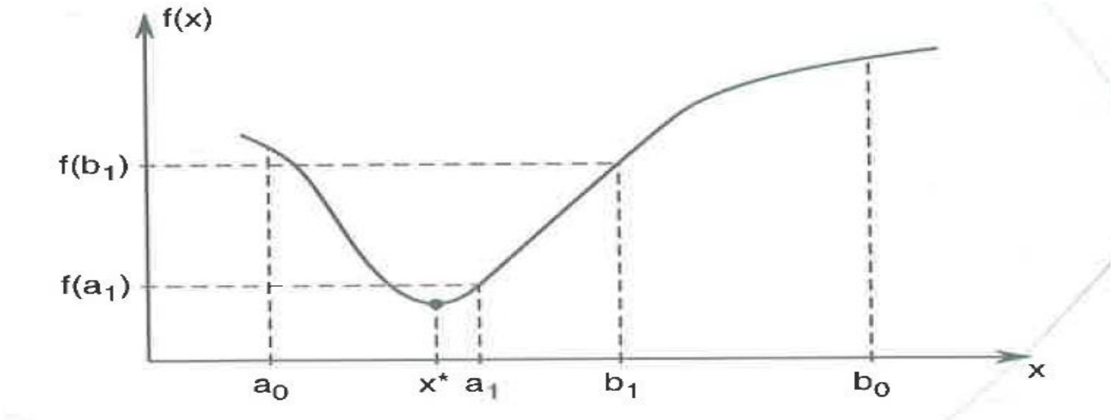


Figure 7.3 The case where $f(a_1) < f(b_1)$; the minimizer $x^* \in [a_0, b_1]$

$\rho(b_1 - a_0) = b_1 - b_2$

$b_0 - a_1 = 1 - \rho$

$b_1 - b_2 = 1 - 2\rho$

$\rho(1-\rho) = 1 - 2\rho \Rightarrow \rho^2 - 3\rho + 1 = 0$

$\rho = \frac{3 \pm \sqrt{5}}{2}$ ($\frac{3+\sqrt{5}}{2}$ 不符)

$\rho = \frac{3-\sqrt{5}}{2}$ ※

IV. Let $f(x) = x^2 + 4\cos x$, $x \in [0, 2]$.

(a) (10 points) Perform 3 iterations of search by using the Golden section method.

(b) (5 points) If the required accuracy must be within 0.00001, then how many iterations are needed when using the Golden section method?

(c) (10 points) Perform 3 iterations of search by using the Fibonacci method with ε=0.01.

(d) (5 points) If the required accuracy must be within 0.00001, then how many iterations are needed when using the Fibonacci method with ε=0.01?

(a) $\rho = 0.382$, $1-\rho = 0.618$

1. $a_1 = a_0 + \rho(b_0 - a_0) = 2 \times 0.382 = 0.764 \Rightarrow f(a_1) = 3.472$
   $b_1 = a_0 + (1-\rho)(b_0 - a_0) = 2 \times 0.618 = 1.236 \Rightarrow f(b_1) = 2.842$
   $> f(a_1) > f(b_1)$ ※

2. $[a_0, a_1]$
   $a_2 = b_1 = 1.236 \Rightarrow f(a_2) = 2.842$
   $b_2 = 1.528 \Rightarrow f(b_2) = 2.506$
   $> f(a_2) > f(b_2)$ ※

3. $[a_0, b_2]$
   $a_3 = b_2 = 1.528 \Rightarrow f(a_3) = 2.506$
   $b_3 = 1.708 \Rightarrow f(b_3) = 2.39$
   $> f(a_3) > f(b_3)$ ※

Next $[a_3, b_0]$

(b)

$0.618^N \le \frac{0.00001}{2} \Rightarrow N \ge 25.36 \Rightarrow N = 26$ ※

(c)

$\rho_1 = 1 - \frac{3}{5} = \frac{2}{5}$, $\rho_2 = 1 - \frac{2}{3} = \frac{1}{3}$, $\rho_3 = \frac{1}{2} - 0.01 = 0.49$

1. $a_1 = a_0 + \rho_1(b_0 - a_0) = 0.8 \Rightarrow f(a_1) = 3.427$
   $b_1 = 1.2 \Rightarrow f(b_1) = 2.89$
   $> f(a_1) > f(b_1)$

2. $[a_0\ b_1]$
   $a_2 = b_1 = 1.2 \Rightarrow f(a_2) = 2.89$
   $b_2 = 1.6 \Rightarrow f(b_2) = 2.44$
   $> f(a_2) > f(b_2)$

3. $[a_0\ b_2]$
   $a_3 = 1.59 \Rightarrow f(a_3) = 2.45$
   $b_3 = 1.61 \Rightarrow f(b_3) = 2.44$
   $> f(a_3) > f(b_3)$

(d)

$\frac{1+2\varepsilon}{F_{N+1}} \le \frac{0.00001}{2} \Rightarrow F_{N+1} \ge 204000 \Rightarrow N = 27$ ※

$F_{26} = 121393$
$F_{27} = 196418$
$F_{28} = 317811$

V. Consider the minimization problem for $f(x)=(x-1)^2(x-2)(x-3)$

(a) (10 points) Perform Newton's method for 3 iterations with $x^{(0)}=1.3$;

(b) (10 points) Repeat (a) with $x^{(0)}=2.2$;

(c) (10 points) Given $x^{(-1)}=2.1$ and $x^{(0)}=2.2$, using the Secant method for 3 iterations.

(a) $f(x)=(x^2-2x+1)(x^2-5x+6)$

$= x^4-7x^3+17x^2-17x+6$

$x^{k+1}=x^k-\dfrac{f'(x^k)}{f''(x^k)}$

$f'(x)=4x^3-21x^2+34x-17$

$f''(x)=12x^2-42x+34$

$x^1=1.3-\dfrac{0.498}{5.764}=2.856$ ※

$x^2=2.856-\dfrac{1.998}{11.929}=2.688$ ※

$x^3=2.644$ ※

(b)

$x^0=2.2$

$x^1=-1.7$ ※

$x^2=-0.592$ ※

$x^3=0.127$ ※

(c)

$x^1=2.2-\dfrac{2.2-2.1}{-1.248-(-1.166)}\times(-1.248)=0.678$ ※

$x^2=x^1-\dfrac{x^1-x^0}{f'(x^1)-f'(x^0)}f'(x^1)=3.916$ ※

$x^3=x^2-\dfrac{x^2-x^1}{f'(x^2)-f'(x^1)}f'(x^2)=0.886$ ※

VI. (15 points) If $\{x^{(k)}\}_{k=0}^{\infty}$ is a steepest descent sequence for a given function, please prove that for each k, $(x^{(k+1)}-x^{(k)})$ is orthogonal to $(x^{(k+2)}-x^{(k+1)})$, where orthogonal means $<(x^{(k+1)}-x^{(k)}), (x^{(k+2)}-x^{(k+1)})>=0$

VII. Consider $f(x_1, x_2) = x_1^4 + 2x_1^2 - 2x_2^2 + 2x_1x_2 + 2$

(a) (15 points) Use Newton's method for 2 iterations with $x^{(0)} = [0,\ 0]$;

(b) (15 points) Repeat (a) if possible, apply the Levenberg-Marquardt modification.

$$X^{k+1} = X^k - H(X^k)^{-1} \cdot g(X^k)$$

$$\nabla f(x_1, x_2) = \begin{bmatrix} 4x_1^3 + 4x_1 + 2x_2 \\ -4x_2 + 2x_1 \end{bmatrix} = g(x)$$

$$H(x_1, x_2) = \begin{bmatrix} 12x^2 + 4 & 2 \\ 2 & -4 \end{bmatrix}$$

(a)

$$X^1 = X^0 - H(X_0)^{-1} g(X_0) = \begin{bmatrix} 0 \\ 0 \end{bmatrix} - \begin{bmatrix} 4 & 2 \\ 2 & -4 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \text{※}$$

$$X^1 = X^1 - H(X_1)^{-1} g(X_1) = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \text{※}$$

(b)

$H(x^0)^{-1}$, $H(x^1)^{-1}$ 不正定

$$X^1 = X^0 - (H(X^0) + \mu_k I)^{-1} g(X^0) = \begin{bmatrix} 0 \\ 0 \end{bmatrix} - \left(\begin{bmatrix} 4 & 2 \\ 2 & -4 \end{bmatrix} + \begin{bmatrix} \mu_k & 0 \\ 0 & \mu_k \end{bmatrix} \right)^{-1} \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} 0 \\ 0 \end{bmatrix} \text{※}$$

VIII. Let $f(x_1,x_2) = x_1^2 + 2x_2^2 + x_1x_2 - 2x_1 - x_2$.

(a) (5 points) Express $f(x_1,x_2)$ in the form of $f(x) = \frac{1}{2}x^TQx - x^Tb$

(b) (15 points) Find the minimizer of f using the conjugate gradient algorithm with the initial point $x^{(0)} = [0,\ 0]^T$.

(a)

$$f(x) = \frac{1}{2}x^T \underbrace{\begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix}}_{Q} x - x^T \underbrace{\begin{bmatrix} 2 \\ 1 \end{bmatrix}}_{b} \quad \#$$

(b) 求 x^1, α^0

$$\nabla f(x) = \begin{bmatrix} 2x_1 + x_2 - 2 \\ 4x_2 + x_1 - 1 \end{bmatrix} \qquad \nabla f(x^0) = \begin{bmatrix} -2 \\ -1 \end{bmatrix} = g(x^0)$$

$$d^0 = -g^0 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

$$\alpha^0 = -\frac{[-2\ -1]\begin{bmatrix} 2 \\ 1 \end{bmatrix}}{\begin{bmatrix} 2 \\ 1 \end{bmatrix}^T \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix}\begin{bmatrix} 2 \\ 1 \end{bmatrix}} = \frac{5}{16} = 0.3125$$

$$x^1 = x^0 + \alpha^0 d^0 = \begin{bmatrix} 0.625 \\ 0.3125 \end{bmatrix}$$

求 $x_2, \alpha_1, d_1, \beta_0$

$$g^1 = \begin{bmatrix} -0.4375 \\ 0.875 \end{bmatrix} \qquad \beta_0 = \frac{[-0.4375\ \ 0.875]\begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix}\begin{bmatrix} 2 \\ 1 \end{bmatrix}}{16} = 0.1914$$

$$d^1 = -g^1 + \beta_0 d^0 = -\begin{bmatrix} -0.4375 \\ 0.875 \end{bmatrix} + 0.1914\begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 0.82 \\ -0.68 \end{bmatrix}$$

$$\alpha^1 = -\frac{[-0.4375\ \ 0.875]\begin{bmatrix} 0.82 \\ -0.68 \end{bmatrix}}{[0.82\ \ -0.68]\begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix}\begin{bmatrix} 0.82 \\ -0.68 \end{bmatrix}} = -\frac{-0.954}{2.08} = 0.459$$

$$x^2 = x_1 + \alpha^1 d^1 = \begin{bmatrix} 0.625 \\ 0.3125 \end{bmatrix} + 0.459\begin{bmatrix} 0.82 \\ -0.68 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$Qx_2 = b$? $\Rightarrow \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ Yes !! \#

IX. (25 points) Consider a quadratic function as $f(\boldsymbol{x})=1/2\ \boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x}-\boldsymbol{b}^T\boldsymbol{x}$. In the conjugation direction algorithm, given a starting point $\boldsymbol{x}^{(0)}$ and $\boldsymbol{Q}$-conjugate vector $\boldsymbol{d}^{(0)}, \boldsymbol{d}^{(1)}, \ldots, \boldsymbol{d}^{(n-1)}$, please prove $\boldsymbol{g}^{(k+1)\mathrm{T}}\boldsymbol{d}^{(i)}=0$ for all $0 \le k \le n-1$ and $0 \le i \le k$, where $\boldsymbol{g}^{(k)} = \nabla f(\boldsymbol{x}^{(k)})$.

X. Suppose we want to minimize

$$f(x_1,x_2)=x_1^{\,4}+2x_2^{\,4}+x_1^2x_2^2-2x_1-x_2$$

(a) (20 points) Use the rank one algorithm to find the solution with $H_0=I_2$ and $x^{(0)}=[0,\ 0]^T$.

(b) (20 points) Use the DFP algorithm to find the solution with $H_0=I_2$ and $x^{(0)}=[1,\ 0]^T$.

(c) (20 points) Use the BFGS algorithm to find the solution with $H_0=I_2$ and $x^{(0)}=[0,\ 1]^T$.

(a)

$\nabla f=\begin{bmatrix}4x_1^3+2x_1x_2^2-2\\ 8x_2^3+2x_1^2x_2-1\end{bmatrix}$ $x^{(0)}=\begin{bmatrix}0\\0\end{bmatrix}$ $\Rightarrow g^{(0)}=\begin{bmatrix}-2\\-1\end{bmatrix}$ $\Rightarrow d^{(0)}=-H_0g^{(0)}=\begin{bmatrix}2\\1\end{bmatrix}$

$\alpha^0=\arg\min_{\alpha\geq 0}(f(x^{(0)}+\alpha d^{(0)}))$, $d^{(0)T}\nabla f(x^{(0)}+\alpha^0 d^{(0)})=0 \Rightarrow [2\ 1]\begin{bmatrix}32\alpha_0^3-2\\16\alpha_0^3-1\end{bmatrix}=0 \Rightarrow \alpha_0=0.384$

$x^{(1)}=x^{(0)}+\alpha^0d^{(0)}=\begin{bmatrix}0.769\\0.384\end{bmatrix}$ $\Rightarrow g^{(1)}=\begin{bmatrix}0.0458\\-0.0928\end{bmatrix}$ $\Rightarrow \Delta g^{(0)}=g^{(1)}-g^{(0)}=\begin{bmatrix}2.0458\\0.9072\end{bmatrix}$, $\Delta x^{(0)}=\alpha_0d^{(0)}=\begin{bmatrix}0.769\\0.384\end{bmatrix}$

$H_1=H_0+\dfrac{(\Delta x^0-H_0\Delta g^{(0)})(\Delta x^{(0)}-H_0\Delta g^{(0)})^T}{\Delta g^{(0)T}(\Delta x^{(0)}-H_0\Delta g^{(0)})}=\begin{bmatrix}0.473&-0.216\\-0.216&0.911\end{bmatrix}$ $\Rightarrow d^{(1)}=-H_1g^{(1)}=\begin{bmatrix}-0.042\\0.094\end{bmatrix}$

$\alpha^1=\arg\min_{\alpha\geq 0}(f(x^{(1)}+\alpha d^{(1)}))$, $d^{(1)T}\nabla f(x^{(1)}+\alpha^1 d^{(1)})=0 \Rightarrow \alpha_1=0.22$

$x^{(2)}=x^{(1)}+\alpha_1d^{(1)}=\begin{bmatrix}0.76\\0.40\end{bmatrix}$, $g^{(1)}\approx 0$ *

(b)

$x^{(0)}=\begin{bmatrix}1\\0\end{bmatrix}$ $\Rightarrow g^{(0)}=\begin{bmatrix}2\\-1\end{bmatrix}$ $\Rightarrow d^{(0)}=-H_0g^{(0)}=\begin{bmatrix}-2\\1\end{bmatrix}$ $\Rightarrow x^{(1)}=x^{(0)}+\alpha^0d^{(0)}=\begin{bmatrix}1-2\alpha_0\\ \alpha_0\end{bmatrix}$

$\alpha^0=\arg\min_{\alpha\geq 0}(f(x^{(0)}+\alpha d^{(0)}))$, $d^{(0)T}\nabla f(x^{(0)}+\alpha^0 d^{(0)})=0 \Rightarrow [-2\ 1]\begin{bmatrix}-36\alpha_0^3+50\alpha_0^2-24\alpha+2\\16\alpha_0^3-8\alpha_0^2+2\alpha_0-1\end{bmatrix}=0$

$88\alpha_0^3-108\alpha_0^2+50\alpha_0-5=0 \Rightarrow 0.135$

$x^{(1)}=x^{(0)}+\alpha^0d^{(0)}=\begin{bmatrix}0.73\\0.135\end{bmatrix}$ $\Rightarrow g^{(1)}=\begin{bmatrix}-0.417\\-0.836\end{bmatrix}$ $\Rightarrow \Delta g^{(0)}=g^{(1)}-g^{(0)}=\begin{bmatrix}-2.42\\0.16\end{bmatrix}$, $\Delta x^{(0)}=\alpha_0d^{(0)}=\begin{bmatrix}-0.27\\0.135\end{bmatrix}$

(BACK)

(b)

$$H_1 = H_0 + \frac{\Delta X^{(0)} \Delta X^{(0)T}}{\Delta X^{(0)T} \Delta g^{(0)}} - \frac{[H_0 \Delta g^{(0)}][H_0 \Delta g^{(0)}]^T}{\Delta g^{(0)T} H_0 \Delta g^{(0)}} = \begin{bmatrix} 0.11 & 0.01 \\ 0.01 & 1.02 \end{bmatrix} \Rightarrow d^{(1)} = -H_1 g^{(1)} = \begin{bmatrix} 0.06 \\ 0.86 \end{bmatrix}$$

$$\alpha^1 = \underset{\alpha > 0}{\arg\min}\,(f(x^{(1)} + \alpha d^{(1)})),\ d^{(1)T} \nabla f(x^{(1)} + \alpha^1 d^{(1)}) = 0 \Rightarrow \alpha_1 = 0.316$$

$$x^{(2)} = x^{(1)} + \alpha_1 d^{(1)} = \begin{bmatrix} 0.75 \\ 0.41 \end{bmatrix},\ g^{(2)} = \begin{bmatrix} -0.08 \\ -0.006 \end{bmatrix},\ \Delta g^{(1)} = \begin{bmatrix} 0.34 \\ 0.83 \end{bmatrix},\ \Delta X^{(1)} = \begin{bmatrix} 0.02 \\ 0.27 \end{bmatrix}$$

$$H_2 = \begin{bmatrix} 0.11 & -0.02 \\ -0.02 & 0.37 \end{bmatrix},\ d^{(2)} = \begin{bmatrix} 0.0082 \\ 0.0002 \end{bmatrix},\ \alpha^2 = 1.2807,\ x^3 = \begin{bmatrix} 0.758 \\ 0.409 \end{bmatrix},\ g^{(3)} \approx 0\ \#$$

(c)

$$x^{(0)} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \Rightarrow g^{(0)} = \begin{bmatrix} -2 \\ 7 \end{bmatrix} \Rightarrow d^{(0)} = -H_0 g^{(0)} = \begin{bmatrix} 2 \\ -7 \end{bmatrix} \Rightarrow x^{(1)} = x^{(0)} + \alpha^0 d^{(0)} = \begin{bmatrix} 2\alpha^0 \\ 1 - 7\alpha^0 \end{bmatrix}$$

$$\alpha^0 = \underset{\alpha > 0}{\arg\min}\,(f(x^{(0)} + \alpha d^{(0)})),\ d^{(0)T} \nabla f(x^{(0)} + \alpha^0 d^{(0)}) = 0 \Rightarrow \alpha_0 = 0.0891$$

$$x^{(1)} = x^{(0)} + \alpha^0 d^{(0)} = \begin{bmatrix} 0.178 \\ 0.376 \end{bmatrix},\ g^{(1)} = \begin{bmatrix} -1.93 \\ -0.55 \end{bmatrix}$$

$$\Delta g^{(0)} = g^{(1)} - g^{(0)} = \begin{bmatrix} 0.073 \\ -7.55 \end{bmatrix},\ \Delta X^{(0)} = \alpha_0 d^{(0)} = \begin{bmatrix} 0.178 \\ -0.624 \end{bmatrix}$$

$$H_1 = H_0 + \left(1 + \frac{\Delta g^{(0)T} H_0 \Delta g^{(0)}}{\Delta g^{(0)T} \Delta X^{(0)}}\right) \frac{\Delta X^{(0)} \Delta X^{(0)T}}{\Delta X^{(0)T} \Delta g^{(0)}} - \frac{H_0 \Delta g^{(0)} \Delta X^{(0)T} + (H_0 \Delta g^{(0)} \Delta X^{(0)T})^T}{\Delta g^{(0)T} \Delta X^{(0)}}$$

$$= \begin{bmatrix} 1.09 & -0.26 \\ -0.26 & -3 \end{bmatrix}$$

(c)

$$d^{(1)} = -H_1 g^{(1)} = \begin{bmatrix} 1.95 \\ -2.15 \end{bmatrix}$$

$$\alpha^1 = \underset{\alpha \geq 0}{\arg\min}\,(f(x^{(1)} + \alpha d^{(1)})),\quad d^{(1)T} \nabla f(x^{(1)} + \alpha^1 d^{(1)}) = 0 \Rightarrow \alpha_1 = 0.21$$

$$x^{(2)} = x^{(1)} + \alpha_1 d^{(1)} = \begin{bmatrix} 0.59 \\ -0.08 \end{bmatrix},\quad g^{(2)} = \begin{bmatrix} -1.17 \\ -1.06 \end{bmatrix},\quad \Delta g^{(1)} = \begin{bmatrix} 0.76 \\ -0.51 \end{bmatrix}$$

$$\Delta x^{(1)} = \begin{bmatrix} 0.41 \\ -0.45 \end{bmatrix},\quad H_2 = \begin{bmatrix} 0.79 & -1.52 \\ -1.52 & -3.81 \end{bmatrix},\quad d_2 = \begin{bmatrix} -0.68 \\ -5.8 \end{bmatrix}$$

$$\alpha^2 = -0.09$$

$$x^3 = \begin{bmatrix} 0.65 \\ 0.44 \end{bmatrix}$$

※

XI. (20 points) For the rank one algorithm applied to the quadratic with Hessian $Q = Q^T$, please show that $H_{k+1}\Delta g^{(i)} = \Delta x^{(i)}$, for $0 \le i \le k$, where $\Delta g^{(i)} = g^{(i+1)} - g^{(i)}$, $\Delta x^{(i)} = x^{(i+1)} - x^{(i)}$, and g is the gradient of the objective function.

XII. Short Answers. (*5 pts each*)

1. Let $\Omega \subset \mathbb{R}^n$ and $g \in \mathcal{C}^1$. If is a local minimizer of g over Ω and is an interior point of Ω, then from FONC, what can we conclude?

ANS:

2. Let $\Omega \subset \mathbb{R}^n$ and $g \in \mathcal{C}^1$. If x^* is a local minimizer of g over Ω, then for any feasible direction $\boldsymbol{w}$ at x^*, from Second Order Necessary Condition (SONC), what can we conclude?

ANS:

3. When the steepest descent method is applied to a quadratic function $f(x) = \frac{1}{2}x^T Q x - x^T b$, where Q is a symmetric positive definite matrix, how many steps it needs to converge to the minimizer in general?

ANS:

4. Consider a quadratic function $f: \mathbb{R}^n \to \mathbb{R}$. If the Quasi-Newton method is applied, how many iterations does it need to reach the minimizer?

ANS:

5. When the Fibonacci search method is used, please specify how we choose the partition ratio ρ.

ANS:

6. Since the commonly used search algorithm in learning is the gradient descent algorithm, please specify the algorithm.

ANS:

7. When a gradient descent algorithm is used, please specify the effects when a small step size is used.

ANS:

8. Which lemma is used to convert the Hestenes-Stiefel formula to the Polak-Ribiere formula?

ANS:

9. What is the main difference between the Newton method and the quasi-Newton method?

ANS:

10. In all quasi-Newton methods, which property the methods need to satisfy?

ANS:

11. What is the difference between the DFP algorithm and the rank one algorithm?

ANS:

12. What is the difference between the DFP algorithm and the BFGS algorithm?

ANS:

XIII. True/False. Be sure to justify your answer. (*5 pts each*)

1. If a point is a global minimizer for an objective function defined over Ω, then the gradient of this function at this point is always equal to zero.

ANS:___ Why:

2. The Fibonacci search method is always more effective than the golden section search method.

ANS:____ Why:

3. If $\left\{x^{(k)}\right\}_{k=0}^{\infty}$ is a steepest descent search sequence for a given function $f: \mathbb{R}^n \to \mathbb{R}$ and if $\nabla f(x^{(k)}) \neq 0$, then always $f(x^{(k+1)}) < f(x^{(k)})$.

ANS:_____ Why:

4. If $\left\{x^{(k)}\right\}_{k=0}^{\infty}$ is a gradient descent sequence for a given function $f: \mathbb{R}^n \to \mathbb{R}$, then for each k, the vector $x^{(k+1)} - x^{(k)}$ is orthogonal to the vector $x^{(i+1)} - x^{(i)}$ for all $i<k$.

ANS:_____ Why:

5. Consider a function $f: \Omega \to \mathbb{R}$, where $\Omega \subset \mathbb{R}^n$ is a convex set and $f \in \mathcal{C}^1$. Given $x^* \in \Omega$, suppose there exists a constant $c>0$ such that $\boldsymbol{d}^T \nabla f(x^*) \geq c\|\boldsymbol{d}\|$ for all feasible direction $\boldsymbol{d}$ at x^*. Then x^* is always a strict local minimizer of f over Ω.

ANS:_____ Why:

6. The Hessian of any objective function $f(x)$ is always a symmetric matrix.

ANS:_____ Why:

7. Consider the general iterative algorithm, $x^{(k+1)} = x^{(k)} + \alpha_k \boldsymbol{d}^{(k)}$, for a non-quadratic function $f(x)$, where $\boldsymbol{d}^{(1)}, \boldsymbol{d}^{(2)}, \cdots$ are given vectors in $\mathbb{R}^n$. If α_k is chosen to minimize $f(x^{(k)} + \alpha_k \boldsymbol{d}^{(k)})$ for each k, then $x^{(k+1)} - x^{(k)}$ is always orthogonal to $\nabla f(x^{(k+1)})$.

ANS:_____ Why:

8. In a Quasi-Newton algorithm for a quadratic function, the resultant search direction are always Q-conjugate.

ANS:_____ Why:

9. DFP can always result in the positive definiteness of H_k in the calculation process.

ANS:_____ Why:

10. In general cases, when the line search is known to be inaccurate, the Polak-Ribiere formula is recommended.

ANS:_____ Why:

— END —

Supplement formulas for Nonlinear programming Midterm

1. In Golden section, $\rho = \frac{3-\sqrt{5}}{2}$ =0.382.

2. the secant method: $x^{(k+1)} = x^{(k)} - \frac{x^{(k)} - x^{(k-1)}}{f'(x^{(k)}) - f'(x^{(k-1)})} f'(x^{(k)})$

3. Basic conjugate direction algorithm: $x(k+1)=x(k)+\alpha_k d(k)$ with $\alpha_k = -\frac{\nabla f(x^{(k)})^T d^{(i)}}{d^{(k)T} Q d^{(k)}}$

4. conjugate gradient algorithm, $\boldsymbol{d}^{(k+1)} = -\boldsymbol{g}^{(k)} + \beta_k \boldsymbol{d}^{(k)}$ with $\beta_k = \frac{g^{(k+1)T} Q d^{(k)}}{d^{(k)T} Q d^{(k)}}$

5. The Hestenes-Stiefel formula: $\beta_k = \frac{g^{(k+1)T}(g^{(k+1)} - g^{(k)})}{d^{(k)T}(g^{(k+1)} - g^{(k)})}$

6. The Polak-Ribiere formula: $\beta_k = \frac{g^{(k+1)T}(g^{(k+1)} - g^{(k)})}{g^{(k)T} g^{(k)}}$

7. The Fletcher-Reeves formula: $\beta_k = \frac{g^{(k+1)T} g^{(k+1)}}{g^{(k)T} g^{(k)}}$

8. Rank one:

$$\boldsymbol{H}_{k+1} = \boldsymbol{H}_k + \frac{(\Delta \boldsymbol{x}^{(k)} - \boldsymbol{H}_k \Delta \boldsymbol{g}^{(k)})(\Delta \boldsymbol{x}^{(k)} - \boldsymbol{H}_k \Delta \boldsymbol{g}^{(k)})^T}{\Delta \boldsymbol{g}^{(k)T}(\Delta \boldsymbol{x}^{(k)} - \boldsymbol{H}_k \Delta \boldsymbol{g}^{(k)})}.$$

9. The DFP update:

$$\boldsymbol{H}_{k+1} = \boldsymbol{H}_k + \frac{\Delta \boldsymbol{x}^{(k)} \Delta \boldsymbol{x}^{(k)T}}{\Delta \boldsymbol{x}^{(k)T} \Delta \boldsymbol{g}^{(k)}} - \frac{[\boldsymbol{H}_k \Delta \boldsymbol{g}^{(k)}][\boldsymbol{H}_k \Delta \boldsymbol{g}^{(k)}]^T}{\Delta \boldsymbol{g}^{(k)T} \boldsymbol{H}_k \Delta \boldsymbol{g}^{(k)}}$$

10. The BFGS algorithm

$$\boldsymbol{H}_{k+1}^{BFGS} = \boldsymbol{H}_k + \left(1 + \frac{\Delta \boldsymbol{g}^{(k)T} \boldsymbol{H}_k \Delta \boldsymbol{g}^{(k)}}{\Delta \boldsymbol{g}^{(k)T} \Delta \boldsymbol{x}^{(k)}}\right) \frac{\Delta \boldsymbol{x}^{(k)} \Delta \boldsymbol{x}^{(k)T}}{\Delta \boldsymbol{x}^{(k)T} \Delta \boldsymbol{g}^{(k)}} - \frac{\boldsymbol{H}_k \Delta \boldsymbol{g}^{(k)} \Delta \boldsymbol{x}^{(k)T} + (\boldsymbol{H}_k \Delta \boldsymbol{g}^{(k)} \Delta \boldsymbol{x}^{(k)T})^T}{\Delta \boldsymbol{g}^{(k)T} \Delta \boldsymbol{x}^{(k)}}.$$