CS532100 Numerical Optimization Homework 1

Student ID Name

Due Nov 11

1. (45%) Consider a function $f(x_1, x_2) = x_1^3 x_2 - 2x_1 x_2^2 + x_1 x_2^3$.

   (a) Compute the gradient and Hessian of f.

   (b) Is $(x_1, x_2) = (1, 1)$ a local minimizer? Justify your answer.

   (c) What is the steepest descent direction of f at $(x_1, x_2) = (1, 2)$?

   (d) What is the Newton's direction of f at $(x_1, x_2) = (1, 2)$?

   (e) Compute the LDL decomposition of the Hessian of f at $(x_1, x_2) = (1, 2)$. (No pivoting)

   (f) Is the Newton's direction of f at $(x_1, x_2) = (1, 2)$ a descent direction? Justify your answer.

   (g) Modify the LDL decomposition computed in (d) such that all diagonal elements of D is larger than or equal to 1, and use the modified LDL decomposition to compute a modified Newton's direction at $(x_1, x_2) = (1, 2)$.

   (h) Suppose $\vec{x_0} = (1, 1)$ and $\vec{x_1} = (1, 2)$ and $B_0 = I$, compute the quasi Newton direction p_1 using SR1.

   (i) Suppose $\vec{x_0} = (1, 1)$ and $\vec{x_1} = (1, 2)$ and $B_0 = I$, compute the quasi Newton direction p_1 using BFGS.

(a)

$\frac{df}{dx_1} = 3x_1^2 x_2 - 2x_2^2 + x_2^3 \qquad \frac{df}{dx_2} = x_1^3 - 4x_1 x_2 + 3x_1 x_2^2$

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

$\frac{d^2 f}{dx_1^2} = 6x_1 x_2 \qquad \frac{d^2 f}{dx_1 x_2} = \frac{d^2 f}{dx_2 x_1} = 3x_1^2 - 4x_2 + 3x_2^2 \qquad \frac{d^2 f}{dx_2^2} = -4x_1 + 6x_1 x_2$

$$\bigtriangledown^2 f(x) = Hessian = \begin{bmatrix} 6x_1 x_2 & 3x_1^2 - 4x_2 + 3x_2^2 \\ 3x_1^2 - 4x_2 + 3x_2^2 & -4x_1 + 6x_1 x_2 \end{bmatrix}$$

(b)

(1,1) is not local minimizer. To be a local minimizer have condition
(1) Gradient(x) = 0 (2) Hessian(x) is Positive definition

Then,

Assume $x^* = (1,1)$

$H(x^*) = \begin{bmatrix} 6 & 2 \\ 2 & 2 \end{bmatrix}$ is PD , but $\bigtriangledown f(x^*) = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$ isn't zero. It is not satisfy condition (1)

(c)

$\vec{p} = - \bigtriangledown f(1,2) = \begin{bmatrix} -6 \\ -5 \end{bmatrix}$

(d)

Because, $\vec{p} = -H(X)^{-1} \bigtriangledown f(x)$

$H(1,2) = \begin{bmatrix} 12 & 7 \\ 7 & 8 \end{bmatrix}, H^{-1} = \begin{bmatrix} 8/47 & -7/47 \\ -7/47 & 12/47 \end{bmatrix}$

$\vec{p} = \begin{bmatrix} 8/47 & -7/47 \\ -7/47 & 12/47 \end{bmatrix} \begin{bmatrix} 6 \\ 5 \end{bmatrix} = \begin{bmatrix} 13/47 \\ 18/48 \end{bmatrix}$

(e)

$H(1,2) = \begin{bmatrix} 12 & 7 \\ 7 & 8 \end{bmatrix}$ Assume $H = LL^T$ by cholesky we can get

$=¿ \begin{bmatrix} 12 & 7 \\ 7 & 8 \end{bmatrix} = \begin{bmatrix} I_{11} & 0 \\ I_{21} & I_{22} \end{bmatrix} \begin{bmatrix} I_{11} & I_{21} \\ 0 & I_{22} \end{bmatrix} = \begin{bmatrix} I_{11}^2 & I_{11}I_{21} \\ I_{11}I_{21} & I_{21}^2 + I_{22}^2 \end{bmatrix}$

Solve them get,$I_{11}^2 = 12, I_{11} = 2\sqrt{3}, I_{11}I_{21} = 7, I_{21} = \frac{7\sqrt{3}}{6}$, and $I_{22} = \frac{\sqrt{141}}{6}$

$H(1,2) = LL^T = \begin{bmatrix} 2\sqrt{3} & 0 & \\ 7\sqrt{3}/6 & \sqrt{141}/6 & 0 \end{bmatrix} \begin{bmatrix} 2\sqrt{3} & 7\sqrt{3}/6 \\ 0 & \sqrt{141}/6 \end{bmatrix}$

Transfor to LDL^T,by $H = (LD^{1/2})(D^{1/2}L^T)$

$H(1,2) = \begin{bmatrix} 1 & 0 \\ 7/12 & 1 \end{bmatrix} \begin{bmatrix} 2\sqrt{3} & 0 \\ 0 & \sqrt{141}/6 \end{bmatrix} \begin{bmatrix} 2\sqrt{3} & 0 \\ 0 & \sqrt{141}/6 \end{bmatrix} \begin{bmatrix} 1 & 7/12 \\ 0 & 1 \end{bmatrix}$

$= \begin{bmatrix} 1 & 0 \\ 7/12 & 1 \end{bmatrix} \begin{bmatrix} 12 & 0 \\ 0 & 47/12 \end{bmatrix} \begin{bmatrix} 1 & 7/12 \\ 0 & 1 \end{bmatrix} = LDLdecomposition$

(f)

P_k is a descent direction if $-90 < \theta < 90$, it means $cos\theta > 0$

$cos\theta = \frac{\vec{p_k} \bigtriangledown f(x)}{\|\vec{p_k}\| \|\bigtriangledown f(x)\|} > 0$, it mean $\vec{p_k} \bigtriangledown f(x) = \begin{bmatrix} 13/47 \\ 18/47 \end{bmatrix} \begin{bmatrix} 6 \\ 5 \end{bmatrix} > 0$

so, it is decent direction.

(g)

(h)

$\vec{s_0} = \vec{x_1} - \vec{x_0}, \quad \vec{g_0} = \triangledown f(1,1) = (2,0)$
$\vec{g_1} = \triangledown f(1,2) = (6,5), \quad \vec{y_0} = (4,5)$

$B_1 = B_0 + \frac{[(4,5)-(0,1)][(4,5)-(0,1)]^T}{(4,4)^T(0,1)} = \begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix}$

$\vec{p_1} = -\begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix}^{-1} \begin{bmatrix} 6 \\ 5 \end{bmatrix} = -\begin{bmatrix} 10/9 \\ 1/9 \end{bmatrix}$

(i)

y_0, s_o from (h)
$B_1 = B_0 - \frac{[(0,1)(0,1)^T]}{(0,1)^T(0,1)} + \frac{[(4,5)(4,5)^T]}{(4,5)^T(0,1)}$
$= I - \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 16/5 & 4 \\ 4 & 5 \end{bmatrix} = \begin{bmatrix} 21/5 & 4 \\ 4 & 5 \end{bmatrix}$
$\vec{p_1} = -B_1^{-}1 g_1 = -\begin{bmatrix} 21/5 & 4 \\ 4 & 5 \end{bmatrix}^{-1} \begin{bmatrix} 6 \\ 5 \end{bmatrix} = \begin{bmatrix} -2 \\ 3/5 \end{bmatrix}$

2. (20%)

   (a) A set $S \subseteq R^n$ is a *convex* set if the straight line connecting any two points in S is also entirely in S. A function $f : S \to R$ is a *convex* function if S is a convex set. The following properties are equivalent:

      i. $S \subseteq R^n$ is a convex set, $f : S \to R$ is a convex function.
      ii. $f(\alpha\vec{x}+(1-\alpha)\vec{y}) \le \alpha f(\vec{x})+(1-\alpha)f(\vec{y})$ for all $\alpha \in [0,1]$, $\vec{x}, \vec{y} \in S$.

      Prove that when f is a convex function, any local minimizer $\vec{x}^*$ is a global minimizer of f.
      (Hint: Suppose there is another point $\vec{z} \in S$ such that $f(\vec{z}) \le f(\vec{x}^*)$, then $\vec{x}^*$ is not a local minimizer.)

   (b) Suppose $f(\vec{x}) = \vec{x}^T Q \vec{x}$, where Q is a symmetric positive semidefinite matrix, show that $f(\vec{x})$ is a convex function.
      (Hint: It might be easier to show $f(\vec{y} + \alpha(\vec{x} - \vec{y})) - \alpha f(\vec{x}) - (1 - \alpha)f(\vec{y}) \le 0$.)

(a)

Assume we have a local minimum at x. And there is some $\epsilon > 0$ such that if $\parallel y$ - $x \parallel < \epsilon$, then$f(y) \ge f(x)$ ∘ Then consider some z in R^n.According to local minimum property,we can get

$$y = \alpha z + (1-\alpha)x,$$

where $\alpha \in [0,1]$ is satisfied $\parallel y$ - $x \parallel < \epsilon$

Solving

$$\epsilon > \parallel y - x \parallel = \parallel \alpha z + (1-\alpha)x - x \parallel = \alpha \parallel z - x \parallel$$

Choose $\alpha < min(1, \epsilon/ \parallel x - z \parallel)$, any α as long as it meets the requirements.Then,

$$f(x) \le f(y) \le \alpha f(x) + (1-\alpha)f(z) \Rightarrow -(1-\alpha)f(x) + (1-\alpha)f(z) \ge 0$$
$$\Rightarrow f(x) \le f(z)$$

according to this result, x is a global minimum

(b)

Because

$$f(\vec{y} + \alpha(\vec{x} - \vec{y})) = \alpha^2 \vec{x}^T Q \vec{x} + (1-\alpha)^2 \vec{y}^T Q \vec{y} + 2\alpha(1-\alpha)\vec{x}^T Q \vec{y}$$

and

$$\alpha f(\vec{x}) - (1-\alpha)f(\vec{y}) = \alpha^2 \vec{x}^T Q \vec{x} + (1-\alpha)^2 \vec{y}^T Q \vec{y} + \alpha(1-\alpha)(\vec{x}^T Q \vec{x} + \vec{y}^T Q \vec{y})$$

We can get it

$$f(\vec{y} + \alpha(\vec{x} - \vec{y})) - \alpha f(\vec{x}) - (1-\alpha)f(\vec{y}) = -\alpha(1-\alpha)(x-y)^T Q(x-y) \ (1)$$

Since Q is positive semidefinite, we have the $(x-y)^T Q(x-y) \ge 0$.And $\alpha \le 1 \Rightarrow (1-\alpha) \in [0,1]$ and hence $-\alpha(1-\alpha) \le 0$ .combine these , we can get the properties always true $((1) \le 0)$ .Hence f(x) is convex function

3. (20%) (Line search method) Suppose $\phi(\alpha) = f(\vec{x_k} + \alpha \vec{p_k}) = (\alpha - 1)^2$.

   (a) The sufficient decrease condition asks $\phi(\alpha) \le \phi(0) + c_1 \alpha \phi'(0)$, $\alpha \in [0, \infty)$. Suppose $c_1 = 0.1$, what is the feasible interval of α satisfying this condition ?

   (b) The curvature condition asks $\phi'(\alpha) \ge c_2 \phi'(0)$. Suppose $c_2 = 0.9$, what is the feasible interval of α satisfying this condition ?

   Answers are put here.

   可以用中文作答。

4. (15%) The conjugate gradient method for solving $Ax = b$ is given in Figure 1, where z_k is the approximate solution. In class we only showed that $\alpha_k = (\vec{p_k}^T \vec{r_k})/(\vec{p_k}^T A \vec{p_k})$ and $\beta_k = -(\vec{p_k}^T A \vec{r_{k+1}})/(\vec{p_k}^T A \vec{p_k})$. Prove that the above formula of α_k and β_k are equivalent to the ones in step (3) and step (6). You may need the relations in step (4) and step (5), and the following properties:

   (a) $\vec{r_i}^T \vec{r_j} = 0$ for all $i \ne j$.

   (b) $\vec{p_i}^T A \vec{p_j} = 0$ for all $i \ne j$.

   (c) $\vec{p_k}$ is a linear combination of $\vec{r_0}, ...\vec{r_k}$, $\vec{p_k} = \sum_{i=1}^{k} \gamma_i \vec{r_i}$.
   (which can be shown from step (7) by mathematical induction.)

(1)	Given $\vec{z}_0$. Let $\vec{p}_0 = \vec{b} - A\vec{z}_0$, and $\vec{r}_0 = \vec{p}_0$.
(2)	For $k = 0, 1, 2, \ldots$ until $\|\vec{r}_k\| \leq \epsilon$
(3)	$\alpha_k = (\vec{r}_k^T \vec{r}_k)/(\vec{p}_k^T A\vec{p}_k)$
(4)	$\vec{z}_{k+1} = \vec{z}_k + \alpha_k \vec{p}_k$
(5)	$\vec{r}_{k+1} = \vec{r}_k - \alpha_k A\vec{p}_k$
(6)	$\beta_k = (\vec{r}_{k+1}^T \vec{r}_{k+1})/(\vec{r}_k^T \vec{r}_k)$
(7)	$\vec{p}_{k+1} = \vec{r}_{k+1} + \beta_k \vec{p}_k$

Figure 1: The CG algorithm.

Ans：

Figure(3) and description show that $\vec{p_k}^T\vec{r_k} = \vec{r_k}^T\vec{r_k}$. Proof it from properties (c),we know

$$\vec{p_k}^T\vec{r_k} = \sum_{i=1}^{k}\gamma_i\vec{r_i}^T\vec{r_k}$$

Because $\vec{r_i}\vec{r_k} = 0$(Properties (a)), we get $\vec{p_k}^T\vec{r_k} = \vec{r_k}^T\vec{r_k}$. (1)

Then

$$\begin{aligned}\beta_k = -(\vec{p_k}^T A\vec{r_{k+1}})/(\vec{p_k}^T A\vec{p_k}) &= -\frac{\vec{r_i}^T\vec{r_k}}{(\vec{p_k}^T A\vec{p_k})}\frac{\vec{p_k}^T A\vec{r_{k+1}}}{\vec{p_k}^T A\vec{p_k}} \\ &= -\alpha_k\vec{p_k}^T A\vec{r_{k+1}}/\vec{r_k}^T\vec{r_k}\end{aligned}$$

For show $-\alpha_k\vec{p_k}^T A\vec{r_{k+1}} = \vec{r_{k+1}}^T\vec{r_{k+1}}$.From Figure1.(5) multiply by $\vec{r_{k+1}}^T$ get

$$\vec{r_{k+1}}^T\vec{r_{k+1}} = (\vec{r_k} - \alpha_k A\vec{p_k})^T\vec{r_{k+1}} = \vec{r_{k+1}}^T\vec{r_{k+1}} - \alpha_k\vec{p_k}^T A\vec{r_{k+1}}$$

with property (a) get $\vec{r_k}^T\vec{r_{k+1}} = 0$,
So,

$$-\alpha_k\vec{p_k}^T A\vec{r_{k+1}} = \vec{r_k}^T\vec{r_{k+1}} \ (2)$$

combine (1) and (2), we proof it