MATH3230A Numerical Analysis

Tutorial 6 with solution

1 Recall:

1. **Broyden's Method:**

   (a)
   - Scant Condition: $A_k(x_k - x_{k-1}) = F(x_k) - F(x_{k-1})$.
   - Rank one update of A : $B = A + \mathbf{u}\mathbf{v}^T$.
   - Sherman-Morrison formula: $B^{-1} = A^{-1} - \frac{[A^{-1}(\mathbf{u} \otimes \mathbf{v})A^{-1}]}{\mathbf{v} \cdot A^{-1}\mathbf{u}}$

   (b) With extra condition on mimicking behavior of the true Jacobian along the the line joining x_{k-1} and x_k, the following 'bad' Broyden's method is derived:
   Select $\mathbf{x}_0$ and A_0. For $k = 0, 1, 2, \cdots$, do the following
   - Compute $\mathbf{d}_k$ using $\mathbf{d}_k = -A_k^{-1}F(\mathbf{x}_k)$
   - Update $\mathbf{x}_{k+1}$ by $\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{d}_k$
   - Update $\mathbf{u}_k = A_k^{-1}F(\mathbf{x}_{k+1})$, $c_k = \mathbf{d}_k^T\mathbf{d}_k + \mathbf{d}_k \cdot \mathbf{u}_k$.
   - Update $A_{k+1}^{-1} = A_k^{-1} - \frac{1}{c_k}(\mathbf{u}_k \otimes \mathbf{u}_k)$

   One order of computaional expense is saved compared with Newton's Method.

   (c) Convergence of Broyden's method:
   For the 'good' Broyden's method, if
   i. $F(\mathbf{x})$ is differentiable, Jacobian $DF(\mathbf{x})$ is Lipschitz continuous with constant γ on a convex open set $D \subset \mathbb{R}^n$.
   ii. $\mathbf{x}^*$ satisfies $F(\mathbf{x}^*) = 0$ and $DF(\mathbf{x}^*)$ is invertible.
   iii. $||\mathbf{x}_0 - \mathbf{x}^*|| < \epsilon$ and $||A_0 - DF(\mathbf{x}^*|| < \delta$ for some constants ϵ, δ.

   Then $||\mathbf{x}_{k+1} - \mathbf{x}^*|| \leq \frac{1}{2}||\mathbf{x}_k - \mathbf{x}^*||$

2. **Steepest Descent Method:**

   - To solve the equation $F(\mathbf{x}) = 0$, we first let $g(\mathbf{x}) = F(\mathbf{x})^T F(\mathbf{x})/2$. Select $\mathbf{x}_0$. For $k = 0, 1, 2, \cdots$, do the following
     – Find α_k that solves the one-dimensional minimization
     $$\min_{\alpha \geq 0} g(x_k - \alpha \nabla g(\mathbf{x}_k))$$
     – Update $\mathbf{x}_{k+1}$ by
     $$\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha_k \nabla g(\mathbf{x}_k)$$
   - For a linear problem, select $\mathbf{x}_0$. For $k = 0, 1, 2, \cdots$, do the following
     – compute
     $$\mathbf{d}_k = \mathbf{b} - A\mathbf{x}_k, \qquad \alpha_k = \frac{\mathbf{d}_k \cdot \mathbf{d}_k}{\mathbf{d}_k \cdot A\mathbf{d}_k}$$
     – Update $\mathbf{x}_{k+1}$ by
     $$\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha_k \mathbf{d}_k$$

2 Exercises:

Please submit solutions of problems with star(*) before 6:30PM on Wednesday and finish the rest by yourself.

1. * Consider the following equation

$$\mathbf{F}(\mathbf{x}) = 0 \tag{1}$$

where $\mathbf{F} : \mathbb{R}^n \to \mathbb{R}^n$ is a non-linear function.

(a) Point out the derivation of the following relationship:

$$A_k = A_{k-1} + \frac{(F(\mathbf{x}_k) - F(\mathbf{x}_{k-1}) - A_{k-1}\mathbf{d}_{k-1}) \otimes \mathbf{d}_{k-1}}{\mathbf{d}_{k-1}^T \mathbf{d}_{k-1}}, \qquad \text{where} \qquad \mathbf{d}_{k-1} = \mathbf{x}_k - \mathbf{x}_{k-1} \tag{2}$$

then write down the Broyden's method using (2) without involving A_k^{-1} in your computaion. This is also called 'good' Broyden's method.

(b) Consider the following system of equations

$$\mathbf{F}(\mathbf{x}) = \mathbf{F}(x, y) = \begin{pmatrix} x - y - 1 \\ x^2 + xy - 6 \end{pmatrix} \in \mathbb{R}^2. \tag{3}$$

Compute the first two iterations using 'good/bad' Broyden's method to solve (3) with initial value $\mathbf{x}_0 = (2, 2)$.

Solution. (a) i. Since we require $A_k d_{k-1} = F(x_k) - F(x_{k-1})$ and $A_k y = A_{k-1} y$ for all $y \cdot d_{k-1} = 0$. So in (5.8), take $D = A_k$, $C = A_{k-1}$, $g = w = d_{k-1}$, $z = F(x_k) - F(x_{k-1})$), then we get (2).

ii. Select $\mathbf{x}_0$ and A_0. For $k = 0, 1, 2, \cdots$, do the following

- Compute $\mathbf{d}_k$ by solving $A_k \mathbf{d}_k = -F(\mathbf{x}_k)$
- Update $\mathbf{x}_{k+1}$ by $\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{d}_k$
- Update A_{k+1} using $A_{k+1} = A_k + \frac{(F(\mathbf{x}_{k+1}) - F(\mathbf{x}_k) - A_k \mathbf{d}_k) \otimes \mathbf{d}_k^T}{\mathbf{d}_k^T \mathbf{d}_k}$

(b) Using good Broyden's method, a direct computation yields

$$F(x_0, y_0) = F(2, 2) = \begin{pmatrix} -1 \\ 2 \end{pmatrix};$$

$$DF(x, y) = \begin{pmatrix} 1 & -1 \\ 2x + y & x \end{pmatrix};$$

and so

$$DF(x_0, y_0) = DF(2, 2) = \begin{pmatrix} 1 & -1 \\ 6 & 2 \end{pmatrix}.$$

Let $A_0 = DF(2, 2)$, we have

$$d_1 = -A_0^{-1} F(x_0) = \frac{1}{8} \begin{pmatrix} 2 & 1 \\ -6 & 1 \end{pmatrix} \begin{pmatrix} -1 \\ 2 \end{pmatrix} = \begin{pmatrix} 0 \\ -1 \end{pmatrix}.$$

We then update x_1 by $x_1 = x_0 + d_1 = \begin{pmatrix} 2 \\ 2 \end{pmatrix} + d_1 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$

Now we update A_1 by $A_1 = A_0 + \frac{(F(x_1,y_1)-F(x_0,y_0)-A_0 d_1)\otimes d_1}{d_1^T d_1}$:

$$A_1 = \begin{pmatrix} 1 & -1 \\ 6 & 2 \end{pmatrix} + \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix} - \begin{pmatrix} -1 \\ 2 \end{pmatrix} - \begin{pmatrix} 1 & -1 \\ 6 & 2 \end{pmatrix}\begin{pmatrix} 0 \\ -1 \end{pmatrix}\right)\begin{pmatrix} 0 \\ -1 \end{pmatrix}^T /1$$

That is,

$$A_1 = \begin{pmatrix} 1 & -1 \\ 6 & 2 \end{pmatrix}$$

Similarly, we have $d_2 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, $x_2 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$.

□

2. * Consider the following equation

$$\mathbf{F}(\mathbf{x}) = 0 \tag{4}$$

where $\mathbf{F} : \mathbb{R}^n \to \mathbb{R}^n$ is a non-linear function. Instead of solving (4), we minimize the following function:

$$g(\mathbf{x}) = \frac{1}{2}\mathbf{F}(\mathbf{x})^T \mathbf{F}(\mathbf{x}) \tag{5}$$

(a) Show that $\nabla g(\mathbf{x}) = D\mathbf{F}(x)^T \mathbf{F}(\mathbf{x})$ by direct calculation..

(b) State the Steepest Descent Method to minimize the function (5).

(c) Consider the special case of $\mathbf{F} : \mathbb{R}^2 \to \mathbb{R}^2$, apply the Steepest Descent Method algorithm to solve:

$$\mathbf{F}(x, y) = \begin{pmatrix} 2x + y \\ x - y + 1 \end{pmatrix} = \mathbf{0}.$$

Give all the detailed computing steps for the first iteration with initial guess $(x_0, y_0)^T = (1, 0)$.

Solution.

(a) Denote

$$F(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_n(\mathbf{x})), \quad \text{and} \quad \mathbf{x} = (x_1, x_2, \ldots, x_n)$$

By definition, we have

$$\nabla g(\mathbf{x}) = \frac{1}{2}\begin{pmatrix} \frac{\partial}{\partial x_1}\left(\sum_{i=1}^n f_i(\mathbf{x}) f_i(\mathbf{x})\right) \\ \frac{\partial}{\partial x_2}\left(\sum_{i=1}^n f_i(\mathbf{x}) f_i(\mathbf{x})\right) \\ \vdots \\ \frac{\partial}{\partial x_n}\left(\sum_{i=1}^n f_i(\mathbf{x}) f_i(\mathbf{x})\right) \end{pmatrix}.$$

W.L.O.G, we consider

$$\begin{aligned}
& \frac{\partial}{\partial x_1}\left(\sum_{i=1}^n f_i(\mathbf{x}) f_i(\mathbf{x})\right) \\
= & \sum_{i=1}^n \frac{\partial}{\partial x_1}\left(f_i(\mathbf{x}) f_i(\mathbf{x})\right) \\
= & \sum_{i=1}^n \frac{\partial f_i(\mathbf{x})}{\partial x_1} f_i(\mathbf{x}) + \sum_{i=1}^n f_i(\mathbf{x}) \frac{\partial f_i(\mathbf{x})}{\partial x_1} \\
= & 2\left(\frac{\partial F(\mathbf{x})}{\partial x_1}\right)^T \cdot F(\mathbf{x}).
\end{aligned}$$

Therefore, we have

$$\nabla g(\mathbf{x}) = \begin{pmatrix} \left(\frac{\partial F(\mathbf{x})}{\partial x_1}\right)^T \cdot F(\mathbf{x}) \\ \left(\frac{\partial F(\mathbf{x})}{\partial x_2}\right)^T \cdot F(\mathbf{x}) \\ \vdots \\ \left(\frac{\partial F(\mathbf{x})}{\partial x_n}\right)^T \cdot F(\mathbf{x}) \end{pmatrix} = DF(\mathbf{x})^T F(\mathbf{x}).$$

(b) To solve $F(\mathbf{x}) = 0$, let $g(\mathbf{x}) = \frac{1}{2} F(\mathbf{x})^T F(\mathbf{x})$ and select $\mathbf{x_0}$. For $k = 0, 1, 2, \ldots$, do the following:

i. Find α_k that solve the one-dimensional minimization

$$\min_{\alpha \geq 0} g\left(\mathbf{x_k} - \alpha \nabla g(\mathbf{x_k})\right)) \tag{6}$$

ii. Update $\mathbf{x_{k+1}}$ by

$$\mathbf{x_{k+1}} = \mathbf{x_k} - \alpha_k \nabla g(\mathbf{x_k}) \tag{7}$$

(c) Since $F(x, y) = \begin{pmatrix} 2x + y \\ x - y + 1 \end{pmatrix}$, we have

$$g(x, y) = ((2x + y)^2 + (x - y + 1)^2)/2$$

and

$$DF(x, y) = \begin{pmatrix} 2 & 1 \\ 1 & -1 \end{pmatrix}$$

as $\nabla g(x, y) = DF(x, y)^T F(x, y)$, we get

$$\nabla g(x, y) = [DF(x, y)]^T F(x, y) = \begin{pmatrix} 5x + y + 1 \\ x + 2y - 1 \end{pmatrix}$$

Therefore $\nabla g(x_0, y_0) = \begin{pmatrix} 6 \\ 0 \end{pmatrix}$

Now we need to solve (6) with $(x_0, y_0)^T = (1, 0)$. We have

$$\begin{aligned}
& g((x_0, y_0) - \alpha \nabla (g(x_0, y_0))) \\
= & g\left(1 - 6\alpha, 0\right) \\
= & \tfrac{1}{2}(4(1 - 6\alpha)^2 + (2 - 6\alpha)^2) \\
= & \tfrac{1}{2}(5(6\alpha)^2 - 12(6\alpha) + 8)
\end{aligned}$$

Therefore $g((x_0, y_0) - \alpha\nabla(g(x_0, y_0)))$ attends its minimum($\alpha > 0$) when

$$6\alpha = \frac{6}{5}.$$

that is $\alpha = \frac{1}{5}$ Now, by (7), we obtain

$$\begin{aligned}(x_1, y_1)^T &= (x_0, y_0)^T - \alpha\nabla g(x_0, y_0) = \begin{pmatrix} 1 \\ 0 \end{pmatrix} - \tfrac{1}{5}\begin{pmatrix} 6 \\ 0 \end{pmatrix} \\ &= \begin{pmatrix} -0.2 \\ 0 \end{pmatrix}\end{aligned}$$

□

3. Consider a linear problem of finding solution $\mathbf{x}$ to $A\mathbf{x} = \mathbf{b}$, where A is symmetric and positive definite. Defining the minimizing function $g(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T A\mathbf{x} - \mathbf{b}^T x$

   (a) * Suppose $\mathbf{x}^*$ solves $A\mathbf{x}^* = \mathbf{b}$, showing that $\mathbf{x}^*$ minimizes $g(\mathbf{x})$.

   (b) * Show that the function $g(\mathbf{x})$ is a convex function.
   A function g is called convex if, for any two points $\mathbf{x}_1$ and $\mathbf{x}_2$ in $\mathbb{R}^n$, and $t \in [0, 1]$, we have

$$g(t\mathbf{x}_1 + (1-t)\mathbf{x}_2) \le tg(\mathbf{x}_1) + (1-t)g(\mathbf{x}_2)$$

   (c) Show that the descent direction in the $k-th$ step is perpendicular to the $(k+1)-th$ step. Try to draw a simple diagram to explain the geometric meaning of this phenomenon.

Solution. (a) Note for any point $z = y + x^*$, we have:

$$\begin{aligned} g(z) &= (y + x^*)^T A(y + x^*)/2 - b^T(y + x^*) \\ &= x^{*T}Ax^*/2 - b^T x^* + x^{*T}Ay/2 + y^T Ax^*/2 + y^T Ay/2 - b^T y \\ &= g(x^*) + y^T Ay/2 + y^T Ax^* - b^T y \\ &= g(x^*) + y^T Ay/2 \\ &\ge g(x^*) \end{aligned}$$

The last inequality uses A is symmetric positive definite.

(b)

$$\begin{aligned} tg(x_1) + (1-t)g(x_2) &= tx_1^T Ax_1/2 + (1-t)x_2^T Ax_2/2 - tx_1^T b - (1-t)x_2^T b \\ g(tx_1 + (1-t)x_2) &= t^2 x_1^T Ax_1/2 + (1-t)^2 x_2^T Ax_2/2 + t(1-t)x_1^T Ax_2 - tx_1^T b - (1-t)x_2^T b \end{aligned}$$

Note

$$0 < (x_1^T - x_2^T)A(x_1 - x_2) = x_1^T Ax_1 + x_2^T Ax_2 - 2x_1^T Ax_2 \tag{8}$$

We have

$$\begin{aligned} g(tx_1 + (1-t)x_2) &\le t^2 x_1^T Ax_1/2 + (1-t)^2 x_2^T Ax_2/2 + t(1-t)(x_1^T Ax_1 + x_2^T Ax_2)/2 - tx_1^T b - (1-t)x_2^T b \\ &\le tx_1^T Ax_1/2 + (1-t)x_2^T Ax_2/2 - tx_1^T b - (1-t)x_2^T b \\ &\le tg(x_1) + (1-t)g(x_2) \end{aligned}$$

(c) Let d_k be the k-th descent direction respectively. Then we have

$$d_k = -\nabla g(x_k) = b - Ax_k.$$

Since $g(x)$ is convex function, $g(x_k - \alpha \nabla g(x_k))$ is also convex function for variable α. To find α_k that solves the one-dimensional minimization

$$\min_{\alpha \geq 0} g(x_k - \alpha \nabla f(x_k)),$$

we only need to find α_k satisfying

$$\frac{\partial g(x_k - \alpha \nabla f(x_k))}{\partial \alpha} = 0$$

Then we have

$$\nabla g(x_k - \alpha_k \nabla g(x_k))^T (-\nabla g(x_k)) = 0$$

Since $x_{k+1} = x_k - \alpha_k \nabla g(x_k)$,

$$\nabla g(x_{k+1})^T (-\nabla g(x_k)) = 0$$

Hence,

$$d_{k+1}^T d_k = 0.$$

□