Numerical Optimization

Xu Shu
201818000206008
University of Chinese Academy of Sciences

December 22, 2018

1 Homework 1

1.1

$$\begin{aligned}\nabla q(x) &= \frac{1}{2}D(x)Ax + \frac{1}{2}D(Ax)x + D(b^T x)\\ &= \frac{1}{2}EAx + \frac{1}{2}Ax + b\\ &= Ax + b\end{aligned}$$

$$\begin{aligned}\nabla^2 q(x) &= D(\nabla q(x))\\ &= D(Ax + b)\\ &= A\end{aligned}$$

1.2

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

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

$$\nabla^2 f(x^*) = \begin{bmatrix} 802 & -400 \\ -400 & 200 \end{bmatrix}$$

Since $802 > 0$ and $802*200-400^2 > 0$, $\nabla f(x^*)$ is positive definite. And there is only one solution $x^* = (1,1)^T$ such that $\nabla f(x) = 0$. Thus $x^* = (1,1)^T$ is the unique local minimize of $f(x)$.

Given $\hat{\Delta} = 1,\ \Delta_0 = 0.5,\ \eta = 0.125$ and $x_0 = (0,0)^T$, the iterations are as follows:

n	x_1	x_2
0	0	0
1	0.1250	0
2	0.3371	0.0687
3	0.4034	0.1583
4	0.5946	0.3194
5	0.6463	0.4151
6	0.7980	0.6138
7	0.8341	0.6944
8	0.8341	0.6944
9	0.8658	0.7483
10	0.9278	0.8568
11	0.9678	0.9351
12	0.9922	0.9839
13	0.9992	0.9983
14	1.0000	1.0000

Table 1: Trust Region Method with Dogleg

1.3

$$\nabla f(x) = \begin{bmatrix} 10x_1 \\ x_2 \end{bmatrix}, \; p(x) = -\nabla f(x) = \begin{bmatrix} -10x_1 \\ -x_2 \end{bmatrix}$$

We have $\alpha_k = -\frac{p_k^T \nabla f_k}{p_k^T A p^k}$ and $x_{k+1} = x_k + \alpha_k p_k$ for the exact line search. So

$$\alpha_0 = \frac{2}{11}, \;\; p_0 = (-1, -1)^T, \;\; x_1 = (-\frac{9}{110}, \frac{9}{11})^T$$

$$\alpha_1 = \frac{2}{11}, \;\; p_1 = (\frac{9}{11}, -\frac{9}{11})^T, \;\; x_2 = (\frac{81}{1210}, \frac{81}{121})^T.$$

Motivated by the first two iterations, we conjecture that $\{x^k\}$ will vibrate left and right and converge to $(0,0)^T$.

1.4

For convenience, we first reformulate the original problem as follows

$$min \;\; \frac{1}{2} x^T A x - b^T x$$

where

$$A = \begin{bmatrix} 4 & 1 \\ 1 & 3 \end{bmatrix}, \;\; b = \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

After applying CG method with exact line search with argument A, b, we have

$$
\begin{aligned}
&x_0 = (2,1)^T,\quad r0 = (8,3)^T,\quad \alpha_0 = 0.2205\\
&x_1 = (0.2356, 0.3384)^T,\ r_1 = (0.2810, -0.7492)^T,\\
&\beta_1 = 0.0088,\ p_1 = (-0.3511, 0.7229)^T,\ \alpha_1 = 0.4122\\
&x_2 = (0.0909, 0.6364)^T
\end{aligned}
$$

2 Homework 2

2.1

Note that

$$
f(x) = t_1^2 + t_2^2 + t_1^4,\quad \nabla f(x) = \begin{bmatrix} 2t_1 + 4t_1^3 \\ 2t_2 \end{bmatrix},
$$

$$
\nabla^2 f(x) = \begin{bmatrix} 2+12t_1^2 & 0 \\ 0 & 2 \end{bmatrix},\quad \nabla^2 f(x)^{-1} = \begin{bmatrix} \frac{1}{2+12t_1^2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}.
$$

Take the step size to be unit, $x_{k+1} = x_k - \nabla^2 f_k{}^{-1} \nabla f_k = (\frac{4t_{k1}^3}{1+6t_{k1}^2}, 0)^T$. So $x_1 = (\frac{4\epsilon^3}{1+6\epsilon^2}, 0)^T$ and $\|x_1\|_2 = \frac{4\epsilon^3}{1+6\epsilon^2} \leq 4\epsilon^3 = O(\epsilon^3)$.

2.2

Given convergence tolerance $\epsilon = 10^{-4}$, $x_0 = (1,1)^T$ and $H_0 = B_0 = I$, the iterations of BFGS method and DFP method are as follows:

n	x_1	x_2	x_1	x_2
0	1	1	1	1
1	-1	-3	-1	-3
2	-0.5432	0.1358	-0.4967	0.1242
3	0.6433	-0.0469	0.5125	-0.0413
4	-0.2202e-3	-0.7151e-3	-0.0875e-3	-0.2667e-3
5	0.0147e-4	0.1025e-4	0.0281e-5	0.1753e-5

Table 2: BFGS method(left) and DFP method(right)

2.3

A closed form of Cauchy point is

$$
s_k^c = -\tau_k \frac{\Delta_k}{\|g_k\|} g_k
$$

where

$$\tau_k = \begin{cases} 1, & \text{if } g_k^T B g_k \le 0, \\ \min(\frac{\|g_k\|^3}{\Delta_k g_k^T B g_k}, 1), & \text{else.} \end{cases}$$

In this exercise, $g_k = J_k^T r_k$ and $B_k = J_k^T J_k$ and so

$$q_k(0) - q_k(s_k^c) = -\frac{\tau_k^2 \Delta^2}{2\|g^k\|^2} \|J_k g_k\|^2 + \tau_k \Delta_k \|J_k^T r_k\|.$$

$\tau_k = 1$:

- $g_k^T B g_k = \|J_k g_k\|^2 = 0$: $q_k(0) - q_k(s_k^c) = \Delta_k \|J_k^T r_k\| \ge \frac{1}{2}\Delta_k \|J_k^T r_k\|$;
- $\|g_k\|^3 \ge \Delta_k g_k^T B g_k$: $q_k(0) - q_k(s_k^c) \ge -\frac{1}{2}\Delta_k \|g_k\| + \Delta_k \|g_k\| = \frac{1}{2}\Delta_k \|g_k\|$.

$\tau_k = \frac{\|g_k\|^3}{\Delta_k g_k^T B g_k}$:

$$\|g_k\|^3 \le \Delta_k g_k^T B g_k$$

$$q_k(0) - q_k(s_k^c) = \frac{\|g_k\|^4}{2\|J_k J_k^T r_k\|^2} = \frac{1}{2} \frac{\|g_k\|^2}{R(J_k^T J_k, g_k)} \ge \frac{1}{2} \frac{\|g_k\|^2}{\lambda_{max}} \ge \frac{1}{2} \frac{\|g_k\|^2}{\|J_k^T J_k\|} = \frac{1}{2} \|J_k^T r_k\| \frac{\|J_k^T r_k\|}{\|J_k^T J_k\|}.$$

Here the Rayleigh quotient $R(M, x)$ is defined as $\frac{x^* M x}{x^* x}$ and λ_{max} is the largest eigenvalue of $J_k^T J_k$.

So we have $q_k(0) - q_k(s_k^c) \ge \frac{1}{2}\|J_k^T r_k\| \min\{\frac{\|J_k^T r_k\|}{\|J_k^T J_k\|}, \Delta_k\}$.

2.4

Note that $L(x, \lambda) = t_1 + t_2 - \lambda_1(2 - 2t_1^2 - t_2^2) - \lambda_2 t_2$ and then the KKT condition are as follows:

$$\begin{aligned} 1 + 4\lambda_1 t_1 &= 0 \\ 1 + 2\lambda_1 t_2 - \lambda_2 &= 0 \\ 2 - 2t_1^2 - t_2^2 &\ge 0 \\ t_2 &\ge 0 \\ \lambda_1,\ \lambda_2 &\ge 0 \\ \lambda_1(2 - 2t_1^2 - t_2^2) &= 0 \\ \lambda_2 t_2 &= 0 \end{aligned}$$

Since $1 + 4\lambda_1 t_1 = 0$, $\lambda_1 \ne 0$ and so $1 + 2\lambda_1 t_2 \ge 1$ which deduces that $\lambda_2 \ne 0$. Therefore $t_2 = 0$ and $2 - 2t_1^2 - t_2^2 = 0$. Since $t_1 = -\frac{1}{4\lambda_1} \le 0$, there is only one possible KKT point $(-1, 0)$. In fact, it is indeed a KKT point which can be verified when $\lambda_1 = \frac{1}{4}$ and $\lambda_2 = 1$.

2.5

We reformulate the original equality constrained minimization with quadratic penalty method in which BFGS method is applied to find an approximate minimizer of $Q(x;\mu) = f(x) + \frac{\mu}{2}c_1^2(x)$.

Given $\mu_0 = 1,\ \tau_k = 10^{-k},\ x_0 = (-0.2, -0.2)^T$ in quadratic penalty method and convergence tolerance $\epsilon = 10^{-4}$, $H_0 = B_0 = I$ in BFGS method, the minimizer sequence $\{x_k\}$ is listed in the table 3 left, choosing $x_{k+1}^s = x_k,\ \mu_{k+1} = 2\mu_k$.

n	x_1	x_2	x_1	x_2
0	-0.2	-0.2	0	0
1	-1.0574	-1.0574	0.3333	0.3333
2	-1.0298	-1.0298	0.9275	0.9275
3	-1.0153	-1.0153	0.9671	0.9671
4	-1.0077	-1.0077	0.9840	0.9840
5	-1.0039	-1.0039	0.9921	0.9921
6	-1.0019	-1.0019	0.9961	0.9961
7	-1.0010	-1.0010	0.9980	0.9980
8	-1.0005	-1.0005	0.9990	0.9990
9	-1.0002	-1.0002	0.9995	0.9995
10	-1.0001	-1.0001	0.9998	0.9998

Table 3: $x_0 = (-0.2, -0.2)^T$(left) and $x_0 = (0, 0)^T$(right)

If $x_0 = (0, 0)^T$ is given, x_k converges to $(1, 1)^T$ as shown in table 3 right which is acually the maximizer of the original optimization. This unintentional experiment shows that the judicious choice of the starting point plays an significant role in quasi-Newton methods.