HOMEWORK 04

ZHEQING ZHANG

Problem 1 .

The julia file is uploaded to my github: https://github.com/zhangzheqing/Homework-for-258.git

The function I am solving is

$$F(x) = \frac{1}{x} - e \tag{1}$$

We can see if x_0 starts from 0.3, the newton iterative method works and the approximated solution converges to the real one. However, if we start from 1.0, the numerical solution blows up. We can see this from transforming the newton iterative scheme.

$$\begin{aligned} x_{k+1} &= x_k - \frac{F(x_k)}{F'(x_k)} \\ x_{k+1} &= 2x_k - ex_k^2 \\ x_{k+1} &= x_k(2 - ex_k) \end{aligned} \tag{2}$$

By letting $\epsilon_k = x_k - \frac{1}{e}$, we have ,

$$\begin{aligned} \epsilon_{k+1} &= 2\epsilon_k + \frac{1}{e} - ex_k^2, \\ \frac{\epsilon_{k+1}}{\epsilon_k} &= 1 - ex_k, \\ |\epsilon_{k+1}| &= e|\epsilon_k|^2, \end{aligned} \tag{3}$$

We can see that we must set the initial value of x_0 appropriately, such that $e|\epsilon_0|^2 < |\epsilon_0|$. When $x_0 = 0.3$, $\epsilon_0 = -0.06$. We can easily see the sequence $\{\epsilon_k\}$ will tend to zero. When $x_0 = 1.0$, $\epsilon_0 = 0.6$. In this case, $\epsilon_1 = 0.9786$. The sequence $\{\epsilon_k\}$ will be an increasing sequence which goes to ∞.

Problem 2 .

a). By the definition of newton method, we have the following scheme,

$$x_{k+1} = (1 - \frac{1}{q})x_k, |\frac{x_{k+1}}{x_k}| = 1 - \frac{1}{q}, \tag{4}$$

Since q is an integer no less than 2, the right hand side is a positive finite number less than 1. Hence the sequence $\{x_k\}$ converges Q linearly. The convergence ratio is $1 - \frac{1}{q}$.

b) Calculate the gradient and hessian of $f(\vec{x})$,

$$
\begin{aligned}
\nabla f &= \beta\vec{x}\cdot||\vec{x}||^{\beta-2}, \\
\nabla^2 f &= 2(\frac{\beta}{2}-1)\beta||\vec{x}||^{\beta-4}\vec{x}\cdot\vec{x}' + \beta||\vec{x}||^{\beta-2}\cdot Id \\
&= \beta||\vec{x}||^{\beta-4}[(\beta-2)\vec{x}\cdot\vec{x}' + ||\vec{x}||^2 Id] \\
(\nabla^2 f)^{-1} &= \frac{1}{\beta||\vec{x}||^{\beta-2}}(Id - \frac{\beta-2}{\beta-1}\frac{1}{||\vec{x}||}\vec{x}\cdot\vec{x}') \\
x_{k+1} &= x_k - \alpha(\nabla^2 f(x_k))^{-1}\cdot\nabla f(x_k) \\
&= x_k - \alpha(x_k - \frac{\beta-2}{\beta-1}x_k) \\
&= (1-\frac{\alpha}{\beta-1})x_k,
\end{aligned}
\tag{5}
$$

When $\beta > 1$, we can choose α appropriately, such that $(1-\frac{\alpha}{\beta-1}) \in (0,1)$. The numerical solution will converge to optimal solution regardless what x_0 we pick.
When $\beta < 1$, the numerical solution will not converge whatever α, x_0 we choose. Because $f(\vec{x}) = ||\vec{x}||^\beta$ is not convex at any point.

c). See my github.

Problem 3 .

We use induction on k.

1. When $k=0$, $D^1q^0 = p^0$
proof, By substituting $D^1 = D^0 + \frac{y^0 y^{0\prime}}{q^{0\prime}y^0}$, We have

$$
\begin{aligned}
D^1q^0 &= D^0q^0 + \frac{y^0y^{0\prime}q^0}{q^{0\prime}y^0}, \\
&= D^0q^0 + y^0, \\
&= p^0,
\end{aligned}
\tag{6}
$$

The last step is derived from the definition of $y^k = p^k - D^kq^k$.

2. Suppose $D^k q^i = p^i$ holds for $k > i$, $D^{k+1} q^i = p^i$ holds for $k \geq i$.
proof. When $i = k$,

$$\begin{aligned} D^{k+1} q^k &= D^k q^k + \frac{y^k y^{k\prime} q^k}{q^{k\prime} y^k}, \\ &= D^k q^k + y^k, \\ &= p^k, \end{aligned} \tag{7}$$

For $i < k$, since $D^k q^i = p^i$,

$$\begin{aligned} D^{k+1} q^i &= D^k q^i + \frac{y^k (p^k - D^k q^k)' q^i}{q^{k\prime} y^k}, \\ &= p^i + \frac{y^k (p^k q^i - q^{k\prime} p^i)}{q^{k\prime} y^k}, \end{aligned} \tag{8}$$

Since $p^n = (\nabla^2 f(x))^{-1} q^n$, we have $p^{k\prime} q^i = q^{k\prime} p^i$.

$$D^{k+1} q^i = p^i, \tag{9}$$

Notice that $D^{k+1} q^i = p^i$ implies,

$$D^n = [p^0, p^1 ... p^{n-1}][q^0, q^1 ... q^{n-1}]^{-1}, \tag{10}$$

Since q^i are linearly independent, $[q^0, q^1 ... q^{n-1}]$ is invertible.
On the other hand, $(\nabla^2 f(x))^{-1}$ can also be obtained by $[p^0, p^1 ... p^{n-1}][q^0, q^1 ... q^{n-1}]^{-1}$, which implies $D^n = (\nabla^2 f(x))^{-1}$.

Problem 4 .

The table of min value of 18 functions in Toms566 module. The stopping criteria I used for the following problem is that $||\nabla f(x)|| < \epsilon||\nabla f(x_0)||$. ϵ is set to be 10^{-8} by default, otherwise stated. The maximum iterative times is set to be 200. I am using BFGS method and backtracking line search with Wolfe conditions. The two parameters in the wolfe conditions are $c_1 = 0.1, c_2 = 0.98$. (see eq3.6ab in Nocedal and Wright 2006). If the nonpositive matrix appears at the initial step, I add an $-Id \cdot (min(eigen) - 10^{-3})$

Min($f(x)$), $||\nabla f(x)||_2$ of the 18 functions

n	ϵ	min($f(x)$)	$\|\|\nabla f(x)\|\|_2$	iterative time	c_1, c_2
1	-	2.88061545e-15	1.45769843367e-6	27	-
2	-	NaN	NaN	48	-
3	-	1.1279327696189195e-8	7.94778e-13	3	-
4	-	5.474852537910e-19	8.618542118946e-6	147	-
5	-	4.410733134962e-17	9.215294228210e-9	37	-
6	-	32.78750309098244	48.14029321284772	48	-
7	-	1.3997601382183117e-6	3.8433422192e-8	107	-
8	-	0.00085364285	0.0310357128	54	-
9	-	88.0318576	0.000212212	185	-
10	-	3.394752465e-25	1.16529008664e-6	11	-
11	-	85822.20162635631	0.00049625052	23	-
12	-	NaN	NaN	11	-
13	-	3.480970364036307e-6	8.000659462479111e-12	50	-
14	-	7.943360089898714e-8	0.0012423960890	200	-
15	-	5.449487980721794e-6	0.000894531	200	-
16	-	1.0975874461618196e-19	1.5608974971579229e-9	16	-
17	-	9.113232745173565e-14	1.0423314653252328e-5	87	$c_1 = 0.01$
18	-	0.0053864308	9.663039407073988e-5	200	-

TABLE 1. Quasi-Newton BFGS method.

The table of min value of 18 functions in Toms566 module. The stopping criteria I used for the following problem is that $||\nabla f(x)|| < \epsilon||\nabla f(x_0)||$. ϵ is set to be 10^{-9} by default, otherwise stated. The maximum iterative times is set to be 200. For solving newton eq, I used backtracking method with $\alpha = 0.3, \beta = 0.5$. Suppose the nonpositive matrix appears, I add an $-Id \cdot (min(eigen) - 10^{-3})$, so that the Hessian will become positive and the minimum eigenvalue will be 10^{-3}.

Discussion: we can see generally the BGFS is more efficient for minimizing the 18 functions given since the precision is enhanced and number of iterative steps it takes to get satisfactory results are relatively small. Especially, the BGFS works better for the problems which contain hessian with very negative eigenvalues. Such as problem 6, 11. The standard newton method can hardly solve. However, the BGFS can give satisfactory results.

Min($f(x)$), $||\nabla f(x)||_2$ of the 18 functions

n	ϵ	min($f(x)$)	$\|\|\nabla f(x)\|\|_2$	iterative time	min(eig)
1	-	6.599157518022159e-26	2.1949952095092286e-12	14	1.4327
2	-	0.392267911	1.18021155	200	-7.25321
3	-	1.1279327696190214e-8	2.469275154190461e-16	3	0.1396
4	-	1.130024585055066e-6	0.0187245685416	200	5.279e-6
5	-	4.966926975238465e-12	9.560748608590294e-8	16	0.00091
6	-	2219.8508439334	1200.3642	200	-70694.56
7	-	12.68180924	76.3362028486	200	-179.275
8	-	0.00054268	0.008473968098162	17	0.0169
9	-	88.053205818	0.4713919111528313	200	-11.318
10	-	1.3590842481722266e-20	1.23798749051e-5	6	1.9999999
11	-	1.620366948060184e6	491395.37427	200	-3926.871
12	-	2.782877264353357e-6	8.897461275093429e-6	200	1.44263e-5
13	-	2.821674512748804e-7	0.00050963	200	-0.0348
14	-	461.3480	1008.6658645	200	-681.624845
15	-	3015.2466367834	1694.1159810112	200	-673.6739
16	-	4.228933478621659e-24	1.8703681633907943e-11	8	0.301463
17	-	1.8225363181717792e-19	1.2747547318240412e-8	39	0.7195
18	-	0.01050835080158	1.7990083	200	-214.357

TABLE 2. newton method with backtracking line search implemented