HW3

Yilun Zhang

October 16, 2015

Exercise 1

Because *Toms566* package can't be called on my windows machine, I use matlab instead. `https://people.sc.fsu.edu/~jburkardt/m_src/test_opt/test_opt.html` gives the function, gradient and hessian same as Toms566. But the start value and number of variable (length(x0)) varies.

0.1 ALgorithm

1. input a start value $x0$, the tolerance ε, max of iteration I, and the $F(x)$, gradient $G(x)$, hessian matrix $H(x)$, wether use BFGS. Let $i = 1$, H_0be the inverse.

2. while $||G(x_k)|| > \varepsilon||G(x_0)||$ & $i < I$

   (a)
   $$p_i = -H_iG(x_i)$$

   (b) Let $x_{i+1} = x_i + \alpha_i p_i$, where α_i is loosely computed to meet Armijo condition.

   (c) update $F(x_i), G(x_i)$ to $F(x_{i+1}), G(x_{i+1})$, let H_{i+1} either be the inverse of $H(x_{i+1})$ or computed as (6.17) as in ("Numerical Optimization", 2006).

   (d) i=i+1

3. output number of iteration j, the solution x_j and the optimal $f^* = f(x_j)$

prob ID	f^*	number of iter
1	4.46500818473627	20001
2	0.242677406804875	412
3	1.12793276962393e-08	4
4	0.0032230846252832	20001
5	3.97089247824700e-14	6
6	6.37118858804940e-12	45
7	0.00228767657879530	3076
8	2.49997654797500e-06	59
9	0.489395214700814	16
10	25364766502.5965	20001
11	85822.2019034065	251
12	8.39437213309053e-20	13
13	8.42718143579277e-16	13
14	4.93038065763132e-32	2
15	1.63898347747199e-06	14859
16	1.81352665939531e-08	20001
17	7.79452679362206	20001
18	5.93799296903267e-14	13

Table 1: result of Newton method

0.2 Result

The result is shown at Table 1. The link above also gives the true optimal solution which is 0 2.55185717631309e-32 0.56422337000000e-10 1.45525929500097e-13 0 0 0.00228767006976995 0.0625100000000000 1.04000000000000 0 85822.3541521934 9.49342027102616e-31 0 0 0 0 0 2.55317533551634e-14.

The problem 1, 4, 10, 16, 17 reach the max iteration number. But 4, 16 is very close to the true minimizer. 2, 8 are trapped to the local minimizer. Therefore, **1, 2, 8, 10, 17 can't be solved**.

Exercise 2

Call the function above with

$$f_\theta(x,y) = -\sum_{i=1}^{n}[y_i \log h_\theta(x_i) + (1-y_i)log(1-h_\theta(x_i))]$$

where

$$h_\theta(x_i) = \frac{e^{\theta^T x}}{1+e^{\theta^T x}}$$

$$G_\theta(x,y) = \sum_{i=1}^{n}(h_\theta(x_i) - y_i)x_i$$

$$H(x,y) = \sum_{i=1}^{n}[h_\theta(x_i)(1-h_\theta(x_i))x_i^T x_i]$$

$theta = (a,b,c)$, this three functions above are *neloglik.m, gradient.m, hessian.m* in files. then got the optimal $\hat{\theta} = (-4.949378063, 0.002690684, 0.754686856)$