DATA130026 Optimization
Assignment 10
Due Time: at the beginning of the class, May. 25, 2021

1. Suppose f is a general (may be nonconvex) continuous differentiable real valued function whose gradient is Lipschitz continuous with Lipschitz constant L. Show the convergence results (with convergence rate) with both exact minimization and Armijo rule line search for $\{\min_{i=0,\ldots,k} \|\nabla f(x_i)\|\}_k$.

2. Show, using the definition, that the sequence $1 + k^{-k}$ converges superlinearly to 1.

   Coding questions: MATLAB codes questions (Q2 and Q3) should be answered with MATLAB publish in a PDF file and submitted on elearning (`https://ww2.mathworks.cn/help/matlab/matlab_prog/publishing-matlab-code.html?lang=en`). Each method should be written as an independent function, e.g., [output] = BFGS(input).

3. Solve the following problems in MATLAB with gradient descent methods and damped Newton's methods, all with Armijo rule line search. Try two different initial points for each method.

   (a) Solve the following problem

   $$\min_{x_1,x_2} f(x_1,x_2) = \exp(x_1+3x_2-0.1)+\exp(x_1-3x_2-0.1)+\exp(-x_1-0.1)+0.1x^T x$$

   You should set the stopping criterion as $\|\nabla f(x)\| \leq$ 1e-7. Plot figures to show the logarithm of the Euclidean norm of the gradient versus integration number. (You may use the semilogy function to plot figures.)

   (b) Solve the following logistic regression problem:

   $$\min_{w\in\mathbb{R}^n, c\in\mathbb{R}} \frac{1}{m}\sum_{i=1}^{m} \log(1+\exp(-b_i(w^T a_i + c))) + 0.01(w^T w + c^2)$$

   where a_i, b_i are given data. Use the following MATLAB code to generate the data:
   `m = 500; n = 1000;`
   `A=[`a_1`, ...,`a_m`] = randn(n,m); b = sign(rand(m,1)-0.5);`
   Terminate your code when the Euclidean norm of the gradient is smaller than 10^{-4}. Plot figures to show the logarithm of the Euclidean norm of the gradient versus integration number.

4. Use BFGS methods **with backtracking line search** and BB-step size gradient method **with backtracking line search** (you need to implement the algorithm with two updates of t_k as given in the course slides) method to solve

   $$\min \frac{1}{2}x^T Ax + b^T x$$

for A and b generated by

```
rc=1:10:1000;A=sprandsym(100,0.1,rc);b=randn(100,1);
```

Use the all one vector (`ones(n,1)`) as the starting point. Terminate the problem after 1000 iterations or the norm of gradient is less than 1e-6. Note that the optimal value f^* can be computed analytically by the first order optimality condition. Compute it and plot the evolution for $\log(f(x^k) - f^*)$.