DATA130026.01 Optimization
Solution of Assignment 10

1. Show the convergence results (with convergence rate) with both exact minimization and Armijo rule line search for the following cases. Suppose f is a function whose gradient is Lipschitz continuous with Lipschitz constant L. Suppose the step of exact minimization is given by

$$x_{k+1} = \text{argmin}_{0\le t\le 1/L} f(x_k - t\nabla f(x_k)).$$

Suppose the initial step size of Armijo rule is 1/L. $\{\min_{i=0,\ldots,k} \|\nabla f(x_i)\|\}_k$ for general (may be nonconvex) f.

Solution. Since gradient is Lipschitz continuous with constant L, we have

$$f(y) \le f(x) + \nabla f(x)^T(y-x) + \frac{L}{2}\|y-x\|_2^2. \tag{1}$$

Setting $y = x - t\nabla f(x)$, we have

$$f(x - t\nabla f(x)) \le f(x) - t(1-\frac{L}{2}t)\|\nabla f(x)\|_2^2, \qquad \forall t\in\mathbb{R}. \tag{2}$$

(a) **Exact line search**
We have

$$f(x_{k+1}) = \min_t f(x_k - t\nabla f(x_k)) \le f(x_k - t\nabla f(x_k)), \quad \forall t\in\mathbb{R}. \tag{3}$$

$$\overset{(2)}{\le} f(x_k) - t(1-\frac{L}{2}t)\|\nabla f(x_k)\|_2^2, \quad \forall t\in\mathbb{R}. \tag{4}$$

Minimizing the right part of (4) w.r.t. t, we have

$$\begin{aligned} f(x_{k+1}) &\le \min_t \left(f(x_k) - t(1-\frac{L}{2}t)\|\nabla f(x_k)\|_2^2 \right) \\ &\overset{(t=\frac{1}{L})}{=} f(x_k) - \frac{1}{2L}\|\nabla f(x_k)\|_2^2. \end{aligned} \tag{5}$$

Summing up (5) from $i=0$ to k, we have

$$\begin{aligned} f(x_{k+1}) &\le f(x_0) - \frac{1}{2L}\sum_{i=0}^{k}\|\nabla f(x_i)\|_2^2 \\ &\le f(x_0) - \frac{k+1}{2L}\min_{i=0,\ldots,k}\|\nabla f(x_i)\|_2^2. \end{aligned} \tag{6}$$

Combining (6) with $f(x_{k+1}) \ge f(x^*)$, we have

$$\min_{i=0,\ldots,k}\|\nabla f(x_i)\| \le \sqrt{\frac{2L(f(x_0)-f(x^*))}{k+1}} = O(\frac{1}{\sqrt{k}}). \tag{7}$$

(b) **Backtracking line search**

Fact: if $t \in \left(0, \frac{2}{L}(1-\alpha)\right]$, then we have

$$-t(1-\frac{L}{2}t) \leq -\alpha t. \tag{8}$$

At each round k, the Armijo rule search step size t_k satisfying:

$$f(x_k - t_k \nabla f(x_k)) \leq f(x_k) - \alpha t_k \|\nabla f(x_k)\|_2^2. \tag{9}$$

Recall that we have (2), i.e.,

$$f(x_k - t_k \nabla f(x_k)) \leq f(x_k) - t_k(1-\frac{L}{2}t_k)\|\nabla f(x_k)\|_2^2. \tag{10}$$

Then set $\hat{t}$ as initialization.

- When $\hat{t} \in \left(0, \frac{2}{L}(1-\alpha)\right]$, combining (8) and (10),

$$f(x_k - \hat{t} \nabla f(x_k)) \leq f(x_k) - \alpha \hat{t} \|\nabla f(x_k)\|_2^2, \tag{11}$$

  i.e., $t_k = \hat{t}$.
- When $\hat{t} \in \left(\frac{2}{L}(1-\alpha), +\infty\right)$, then run the backtracking line search, as $t_k = \hat{t}\beta^s$ has the smallest s that satisfies the Armijo rule (9), we have

$$f(x_k - t_k \nabla f(x_k)) \leq f(x_k) - \alpha t_k \|\nabla f(x_k)\|_2^2, \tag{12}$$

  while the former step $\frac{t_k}{\beta}$ doesn't satisfies (9):

$$f(x_k - \frac{t_k}{\beta} \nabla f(x_k)) > f(x_k) - \alpha \frac{t_k}{\beta} \|\nabla f(x_k)\|_2^2. \tag{13}$$

  Combining (13) and (2), we have

$$-\frac{t_k}{\beta}(1-\frac{L}{2}\frac{t_k}{\beta}) > -\alpha \frac{t_k}{\beta}, \tag{14}$$

  i.e., $t_k > \frac{2(1-\alpha)\beta}{L}$.

In conclusion,

$$t_k \geq \min\left\{\hat{t}, \frac{2(1-\alpha)\beta}{L}\right\} \triangleq \tilde{t}. \tag{15}$$

Again using (9), we have

$$f(x_{k+1}) \leq f(x_k) - \alpha \tilde{t} \|\nabla f(x_k)\|_2^2. \tag{16}$$

Then we have

$$\min_{i=0,\ldots,k} \|\nabla f(x_i)\|_2^2 \leq \sqrt{\frac{f(x_0) - f(x^*)}{\alpha \tilde{t}(k+1)}} = O(\frac{1}{\sqrt{k}}). \tag{17}$$

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

Solution. Let $x^k = 1 + k^{-k}$, then $x^* = 1$. Since

$$\lim_{k\to\infty} \frac{x^{k+1} - x^*}{x^k - x^*} = \lim_{k\to\infty} \frac{k^k}{(k+1)^{k+1}} = \lim_{k\to\infty} \frac{1}{k+1}(\frac{k}{k+1})^k = \frac{1}{e} \lim_{k\to\infty} \frac{1}{k+1} = 0$$

which means x^k converges superlinearly to 1.

3. Solve the following problems in MATLAB with gradient descent method and damped Newton's method, 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 iteration 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 iteration number.

Solution.

   (a) Please run `codes/problem_3a.m`.

   (b) Please run `codes/problem_3b.m`.

4. Use BFGS method 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) 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 $\log\left(|f(x^k) - f^*|\right)$.

Solution. Please run `codes/problem_4.m`.