CS 726 Assignment 7

Ruochen Lin

April 25, 2018

1

For simplicity, we define $b_k = x_k - x^*$, where we have $\lim_{k\to\infty} b_k = 0$. Because

$$\|b_k + p_k\| = o(\|b_k\|)$$

and

$$\left\|b_k + p_k^N\right\| = O(\|b_k\|^2),$$

we have

$$\begin{aligned} p_k - p_k^N &= (b_k + p_k) - (b_k + p_k^N) \\ &= o(\|b_k\|) - O(\|b_k\|^2) = o(\|b_k\|). \end{aligned}$$

In addition, we have

$$\|b_k\| = O(\|p_k\|),$$

because otherwise we would have

$$\begin{aligned} \|b_k + p_k\| &= O\big(\max\{\|b_k\|, \|p_k\|\}\big) \\ &\geq O(\|b_k\|), \end{aligned}$$

which contradicts our condition that $\|b_k + p_k\| = o(\|b_k\|)$.
Thus, we have

$$\left\|p_k - p_k^N\right\| = o(\|p_k\|).$$

This has a straightforward geometric explanation: both $b_k + p_k$ and $b_k + p_k^N$ are in an n-sphere centered at the origin with radius $o(\|b_k\|)$, and $p_k - p_k^N =$

$(b_k+p_k)-(b_k+p_k^N)$ is a vector connecting two points in this sphere; of course the length of $p_k - p_k^N$ cannot be larger than the diameter of the n-sphere, which is also $o(\|b_k\|)$.

p_k^N can be written as

$$p_k^N = -\nabla^2 f_k^{-1} B_k p_k,$$

so, if we multiply $p_k - p_k^N$ by $-\nabla^2 f_k$ from the left, we would have

$$\left\|(B_k - \nabla^2 f_k)p_k\right\| = o(\|p_k\|).$$

Finally, because $\lim_{k\to\infty} x_k = x^*$, we have $\lim_{k\to\infty} \nabla^2 f_k = \nabla^2 f^*$, so that

$$\begin{aligned}
\left\|(\nabla^2 f_k - \nabla^2 f^*)p_k\right\| &= o(\|p_k\|) \\
\Rightarrow \left\|(B_k - \nabla^2 f^*)p_k\right\| &= \left\|(B_k - \nabla^2 f_k)p_k + (\nabla^2 f_k - \nabla^2 f^*)p_k\right\| \\
&\le \left\|(B_k - \nabla^2 f_k)p_k\right\| + \left\|(\nabla^2 f_k - \nabla^2 f^*)p_k\right\| \\
&= o(\|p_k\|).
\end{aligned}$$

2

B_k is symmetric positive definite, so it can be diagonalized as $B_k = Q\Lambda Q^T$, with Q being orthogonal and Λ being diagonal with only positive nonzero entries. Thus, there exists $B_k^{1/2} = Q\Lambda^{1/2}Q^T$, and the same is true for B_k^{-1}. μ_k can be written in the following form:

$$\begin{aligned}
\mu_k &= \frac{(y_k^T B_k^{-1} y_k)(s_k^T B_k s_k)}{(y_k^T s_k)^2} \\
&= \frac{(y_k^T B_k^{-1/2})(B_k^{-1/2} y_k)(s_k^T B_k^{1/2})(B_k^{1/2} s_k)}{(y_k^T s_k)^2} \\
&\ge \frac{(y_k^T B_k^{-1/2} B_k^{1/2} s_k)^2}{(y_k^T s_k)^2} \\
&= \frac{(y_k^T s_k)^2}{(y_k^T s_k)^2} \\
&= 1.
\end{aligned}$$

3

If $y_k \neq B_k s_k$ and $(y_k - B_k s_k)^T s_k = 0$, then if there exists v such that $(B_k + \sigma v v^T) s_k = y_k$, $\sigma = \pm 1$, then we have

$$\sigma(v^T s_k) v = y_k - B_k s_k.$$

If $v^T s_k = 0$, then

$$\begin{aligned} (B_k + \sigma v v^T) s_k &= y_k \\ \Rightarrow B_k s_k &= y_k, \end{aligned}$$

which contradicts the condition $y_k \neq B_k s_k$. On the other hand, if $v^T s_k \neq 0$, then v must be a multiple of $y_k - B_k s_k$, which needs to satisfy $(y_k - B_k s_k)^T s_k = 0$; this is contradictory as well. In conclusion, there is no symmetric rank one update satisfying secant equation given the conditions.

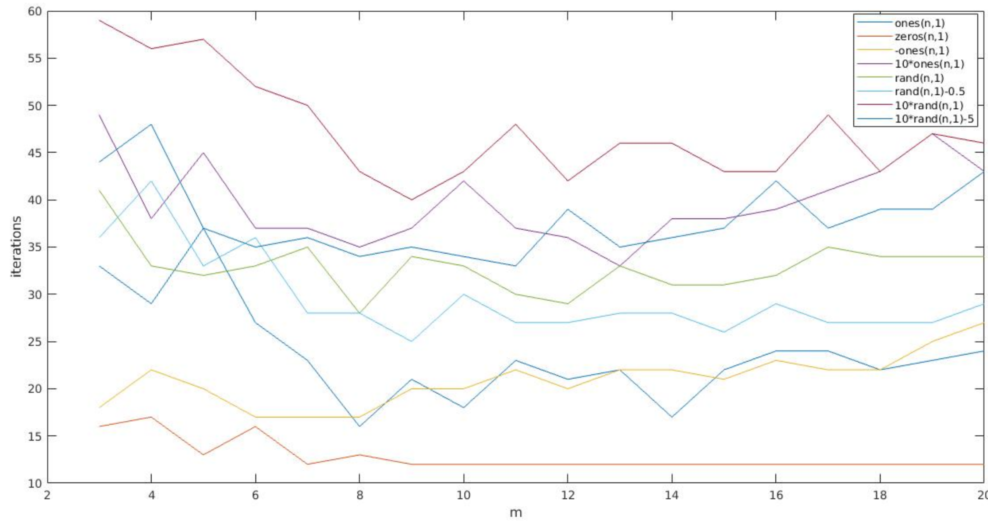

Figure 1: A plot of iter(m) with different starting points.

4

4.1

Here we comment on the change of the number of iterations, with LBFGS, as a function of m with different starting points. We can see from Figure 1 that $\begin{bmatrix} 0 & 0 & \cdots & 0 \end{bmatrix}^T$ leads to the smallest number of iterations, and the number of iterations needed oscillates a bit when $m < 8$, and it flattens out when we have larger ms. For other starting points, it seems that the farther the starting point from the origin is, the more iterations will be needed. We also see an improvement as we increase m in the region of small ms, and then we see an oscillating behavior. Generally, regardless of the starting point, $m = 8 - 10$ will get us satisfactory performance.

4.2

After making minor adjustments so that the nonlinear CG methods and BFGS are using the same set of parameters, EBLS script, and starting point (`ones(100,1)` and `rand(100,1)`), the following is a table of the results: So we see that, with uniform input, BFGS performs much better than the

Table 1: Iterations needed in nonlinear CG and BFGS algorithms to achieve $\left\|\nabla f(x)\right\|_2 \leq 10^{-6}(1+\left|f(x)\right|)$

Algorithm	Iter(`ones`)	Iter(`rand`)
CG-FR	189	343
CG-PRplus	317	472
BFGS	37	173

nonlinear CG algorithms; with random input, BFGS still performs better, but to a lesser extent.