Exercise 3: Better online Learning **Cenying(Tracy) Yang**
SDS 385 Due Date: 9/*/2016
Prof. James Scott

Line Search

(A) **Solution**
Wolfe's conditions include two conditions: (1) sufficient decrease $f(x_k + \alpha p_k) \leq f(x_k) + c_1 \alpha \nabla f_k^T p_k$, where $c_1 \in (0, 1)$ is some fixed constant; and (2) curvature condition $\nabla f(x_k + \alpha p_k)^T p_k \geq c_2 \nabla f_k^T p_k$, where $c_2 \in (c_1, 1)$ is a fixed constant.
The sufficient decrease condition alone is not enough since it is satisfied for all small enough α. However, if the line search algorithm starts with a large step and gradually decreases the step length until the sufficient decrease condition is satisfied, the curvature condition can be dispensed. The idea leads to backtracking line search algorithm. Pseudocode for backtracking line search is stated as follows,
Choose $\alpha_0 \geq 0, \rho, c_1 \in c(0, 1)$
Set $\alpha \leftarrow \alpha_0$
While $f(x_k + \alpha p_k) \geq f(x_k) + c_1 \alpha \nabla f_k^T p_k$
    $\alpha \leftarrow \rho \alpha_0$
End

Quasi-Newton

(A) **Solution**
The key of quasi-Newton is to find an approximate matrix to Hessian at each iteration. The important idea behind the methods is that two successive iterates x_k and x_{k+1} together with the gradients $\nabla f(x_k)$ and $\nabla f(x_{k+1})$ contain curvature (i.e., Hessian) information, in particular, $\nabla f(x_{k+1}) - \nabla f(x_k) \approx H(x_{k+1})(x_{k+1} - x_k)$
Therefore, at every iteration, we would like to choose B^{k+1} to satisfy

$$B_{k+1} y_k = s_k \tag{1}$$

where $y_k = \nabla f(x_{k+1}) - \nabla f(x_k), s_k = x_{k+1} - x_k$. This is known as the secant condition.
BFGS is one method to update B_{k+1}, where

$$B_{k+1} = B_k + \frac{y_k y_k^T}{y_k^T s_k} - \frac{B_k s_k S_k^T B_k}{s_k^T B_k s_k} \tag{2}$$

Its inverse update is

$$(B_{k+1})^{-1} = (I - \frac{s_k y_k^T}{y_k^T s_k}) B_k^{-1} (I - \frac{y_k s_k^T}{y_k^T s_k}) + \frac{s_k s_k^T}{y_k^T s_k} \tag{3}$$

Pseudocode for BFGS along with backtracking line search is stated as follows,
Choose $x_0, B_0 \succ 0, \alpha_0 \geq 0, \rho, c_1 \in c(0, 1), N$
For k in 0:N
    $p_k = -B_k \nabla f(x_k)$
    Set $\alpha \leftarrow \alpha_0$

While $f(x_k + \alpha p_k) \geq f(x_k) + c_1 \alpha \nabla f_k^T p_k$
$\alpha \leftarrow \rho\alpha$
End
$x_{k+1} = x_k + \alpha p_k$
$(B_{k+1})^{-1} = (I - \frac{s_k y_k^T}{y_k^T s_k}) B_k^{-1} (I - \frac{y_k s_k^T}{y_k^T s_k}) + \frac{s_k s_k^T}{y_k^T s_k}$

End