Advanced Topics in Numerical Analysis: Numerical Optimization
CSCI–GA.2945-003
MATH–GA.2012-003
Instructor: Margaret Wright
Computer Science Department
New York University, Spring 2019

Homework Assignment 4
Assigned Thursday, April 11, 2019
Due 11:59pm, Tuesday, April 23, 2019 (not a lecture day!)

Exercise 4.1. This problem is intended to produce some interesting behavior by a pure Newton method for nonlinear equations in two dimensions. We wish to find $x^* \in \mathbb{R}^2$ such that $F(x^*) = 0$, where

\[ F(x) = \begin{pmatrix} x_1^2 + 4x_2 \\ -\frac{1}{2}x_1 + 8x_2^2 \end{pmatrix}. \]  

(4.1)

Write code that implements two methods: Method 1 is a pure Newton method; Method 2 computes the search direction using the Newton equations, but then executes a backtracking line search to choose a value of $\alpha_k$ that achieves a reduction in the merit function $M(x) = \|F(x)\|$, with line search parameters $\eta_s = .001$ and $\gamma_c = \frac{1}{2}$. (In Notes 8, see (i) the local linear model (9.2) and Algorithm 9.1, on page 142, and (ii) the discussion surrounding (13.12) on page 163, involving the “predicted reduction”.) Your code should terminate after a specified maximum number of iterations maxit or when $\|F_k\|$ is less than ftol. At the kth iteration, print k, both components of $x_k$ and $F_k$, and also $\|F_k\|$, using scientific notation for the latter values and showing at least 6 significant digits. In Method 2, also print each trial value of the scalar $\alpha_k$ (during the line search) and the value of the merit function.

(a) Find (by inspection or other simple means) two distinct zeros of $F(x)$ as given in (4.1).

(b) Run Method 1 starting at $x_0 = (1, -1)$ with maxit = 12 and ftol = $1.0e^{-7}$. Are the iterates converging? To what point? What is the apparent rate of convergence of the merit function to zero?

(c) Starting at $x_0 = (1.99, 0)$ with maxit = 9, run Method 1. Describe qualitatively what happens to the iterates and to the merit function. Can you explain this behavior using properties of the specific $F$ and $J$ in this problem?

(d) Starting at $x_0 = (1.99, 0)$ with maxit = 15, run Method 2. What happens to the iterates compared with case (c)? How many non-unit values of $\alpha_k$ occurred?

(e) Starting at $x_0 = (1, 0)$ and with maxit = 8, run Method 1. Please describe and explain the behavior of the iterates.

(f) Starting at $x_0 = (1, 0)$ and with maxit = 8, run Method 2. Comment on the chosen values of $\{\alpha_k\}$ and the differences from the results of (e).

Exercise 4.2. Given a twice-continuously differentiable function $f : \mathbb{R}^n \to \mathbb{R}$, assume that the point $\tilde{x}$ is not a stationary point of $f$. Show that, if $p$ is any specific descent direction at $\tilde{x}$ (i.e., if $p^T g(\tilde{x}) < 0$), then $p$ is the unique minimizer of a quadratic function whose Hessian is

\[ B = I - \frac{1}{p^T p} pp^T - \frac{1}{g^T p} gg^T, \]  

where $g$ denotes the gradient of $f$ at $\tilde{x}$.

You may wish to use the following result from the book by R. Horn and C. Johnson:\footnote{Matrix Analysis, Cambridge University Press, 1985, page 400} If the symmetric matrix $B$ is positive semidefinite, then $x^T B x = 0$ only when $B x = 0$. 
**Exercise 4.3.** Consider a quasi-Newton method in which the search direction $p_k$ is defined by

$$p_k = -M_k g_k,$$

where $g_k$ is the gradient of $f$ at $x_k$ and the symmetric matrix $M_k$ represents a quasi-Newton approximation to the inverse Hessian of a nonlinear function $f(x)$. Assume that the updated matrix satisfies $M_{k+1} y_k = s_k$, where $y_k = g_{k+1} - g_k$ and $s_k = x_{k+1} - x_k$, and that $M_{k+1}$ is updated using the DFP formula,

$$M_{k+1} = M_k - \frac{1}{y_k^T M_k y_k} M_k y_k y_k^T M_k + \frac{1}{s_k^T y_k} M_k s_k s_k^T.$$

Show that, if $M_k$ is singular, then $M_{k+1}$ must be singular.

**Exercise 4.4.** Consider the function $f(x) = x_1^2 + x_2^2 \cos x_3 - e^{x_2} x_3^2 + 4x_3$.

(a) Write down expressions for the gradient $g(x)$ and Hessian $H(x)$ of $f$.

(b) Compute the eigendecomposition (eigenvalues and eigenvectors) of $H(\bar{x})$ at $\bar{x} = (0,1,0)^T$.

(c) (i) Compute the pure Newton direction $p^N$ at $\bar{x}$.

(ii) Is $p^N$ a descent direction for $f$? Explain why or why not.

(d) Compute a modified Newton direction $p^M$ at $\bar{x}$, obtained by solving a linear system involving a modified Hessian that retains the eigenvectors of the true Hessian, but replaces any negative eigenvalue by its absolute value.

(i) Give an explicit algorithm for computing the modified Hessian.

(ii) Is $p^M$ a descent direction? Explain.

(e) (i) Give a direction of negative curvature at $\bar{x}$.

(ii) Explain how you found this direction.

(iii) Verify the negative curvature property numerically.