Homework 7 for SI211: Numerical Analysis

Mao Song
SIST
ShanghaiTech University
Student ID: 2019233134
maosong@shanghaitech.edu.cn

Abstract

This is the solution for Homework 7 of SI211: Numerical Analysis, which is taught by Boris Houska.

1 Problem 1

Provide short answers to the following questions

1. What is the main idea of Newtons method?
2. What is the main idea of Gauss-Newton methods?
3. What is the local convergence rate of the exact Newton method?
4. Under which conditions Newton methods converges in one step?
5. What is the Armijo line search condition and what is it good for?

Solution 1.

1. In order to solve the non-linear equation $f(x) = 0$, we start with an initial guess x_0 and solve the linear equation systems
$$f(x_k) + M(x_k)(x_{k+1} - x_k) = 0, k = \{0, 1, 2, \dots\} \tag{1}$$
Here, the matrix $M(x_k)$ is chosen in such a way that
$$f(x_k) + M(x_k)(x - x_k) \approx f(x) \tag{2}$$
is an approximation of the function f.

2. In order to solve the nonlinear least-squares optimization problem $\min_x \|f(x)\|_2^2$, we start with an initial guess x_0 and solve the standard least-squares problem
$$\min_{\Delta x_k} 1/2\|f(x_k) + f'(x_k)\Delta x_k\|_2^2, k = \{0, 1, 2, \dots\} \tag{3}$$
and update $x_{k+1} = x_k + \alpha_k \Delta x_k$, where $\alpha \in (0, 1]$ is a line search parameter.

3. Under appropriate assumptions, the exact Newton method has Q-quadratic convergence rate.

4. If the objective function is a quadratic, then Newton methods converge in one step.

5. Armijo condition: $F(x_k + \alpha_k \Delta x_k) \leq F(x_k) + c\alpha_k F'(x_k)\Delta x_k$, where $c << 1$ is a constant. This condition ensures that the line search parameter is not excessively large.

2 Problem 2

Consider the function $f(x) = x^3 + 2x + 2$. What happens if you apply Newton's method to the equation $f(x) = 0$ starting with $x_0 = 0$? Work out the iterates of Newton's method explicitly.

Solution 2. The Newton method is given as follows:

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}, k = 0, 1, \dots \tag{4}$$

where $f(x) = x^3 + 2x + 2$ and $f'(x) = 3x^2 + 2$. The iteration result is shown as follows

k	x_k	$f(x_k)$	$f'(x_k)$
0	0	2	2
1	-1	-1	5
2	-0.8	-0.112	3.92
3	-0.7714	-0.0019	3.7853
4	-0.7709	-0.0000	3.7829
5	-0.7709	-0.0000	3.7829
6	-0.7709	0.0000	3.7829
7	-0.7709	0.0000	3.7829

Table 1: Solution to Problem2

From the table, we can see that the Newton method converges to the root of the equation $f(x) = 0$.

3 Problem 3

The goal of this exercise is to implement and exact Newton methods with Gauss-Newton methods by minimizing the nonlinear function

$$f(x, y) = \frac{1}{2}(x-1)^2 + \frac{1}{2}\left(10(y - x^2)\right)^2 + \frac{1}{2}y^2 \tag{5}$$

1. Derive the gradient and Hessian matrix of the function in 5. Then, re-write it in the form $f(x, y) = \frac{1}{2}\|R(x, y)\|_2^2$, where $R : \mathbb{R}^2 \to \mathbb{R}^3$ is the residual function. Derive the Gauss-Newton Hessian approximation and compare it with the exact one. When do the two matrices coincide?
2. Implement your own Newton's Method with exact Hessian information and full steps. Start from the initial point $(x_0, y_0) = (-1, -1)$ and use as termination condition $\|\nabla f(x_k, y_k)\|_2^2 \leq 10^{-3}$. Keep track of the iterates (x_k, y_k) and plot the results.
3. Update your code to use the Gauss-Newton Hessian approximation instead. Plot the difference between exact and approximate Hessian as a function of the iterations.
4. Compare the performance of the implemented methods. Consider the iteration path (x_k, y_k), the number of iterations and the run time.

Solution 3.

1. The gradient of $f(x, y)$ is

$$\nabla f(x, y) = \begin{bmatrix} x - 1 - 200x(y - x^2) \\ 100(y - x^2) + y \end{bmatrix} \tag{6}$$

the Hessian of $f(x, y)$ is thus given by

$$\nabla^2 f(x, y) = \begin{bmatrix} 600x^2 - 200y + 1 & -200x \\ -200x & 101 \end{bmatrix} \tag{7}$$

$f(x, y)$ can be rewritten as

$$f(x, y) = \frac{1}{2}\left\| \begin{bmatrix} x - 1 \\ 10(y - x^2) \\ y \end{bmatrix} \right\|_2^2 \tag{8}$$

thus

$$R(x, y) = \begin{bmatrix} x - 1 \\ 10(y - x^2) \\ y \end{bmatrix} \tag{9}$$

and

$$\nabla R(x,y) = \begin{bmatrix} 1 & 0 \\ -20x & 10 \\ 0 & 1 \end{bmatrix} \tag{10}$$

Thus the Gauss-Newton Hessian approximation is given by

$$M(x,y) = \nabla R(x,y)^T \nabla R(x,y) = \begin{bmatrix} 1+400x^2 & -200x \\ -200x & 101 \end{bmatrix} \tag{11}$$

from the definition of two Hessian matrices, we can see that when $M = \nabla^2 f(x,y)$ if and only if $y = x^2$.

2. The update of Newton's method is given by

$$\begin{bmatrix} x_{k+1} \\ y_{k+1} \end{bmatrix} = \begin{bmatrix} x_k \\ y_k \end{bmatrix} - \left[\nabla^2 f(x_k, y_k)\right]^{-1} \nabla f(x_k, y_k) \tag{12}$$

the result is shown in figure 2, as we can see, the exact Newton method converges to the optimal solution in 10 iterations.

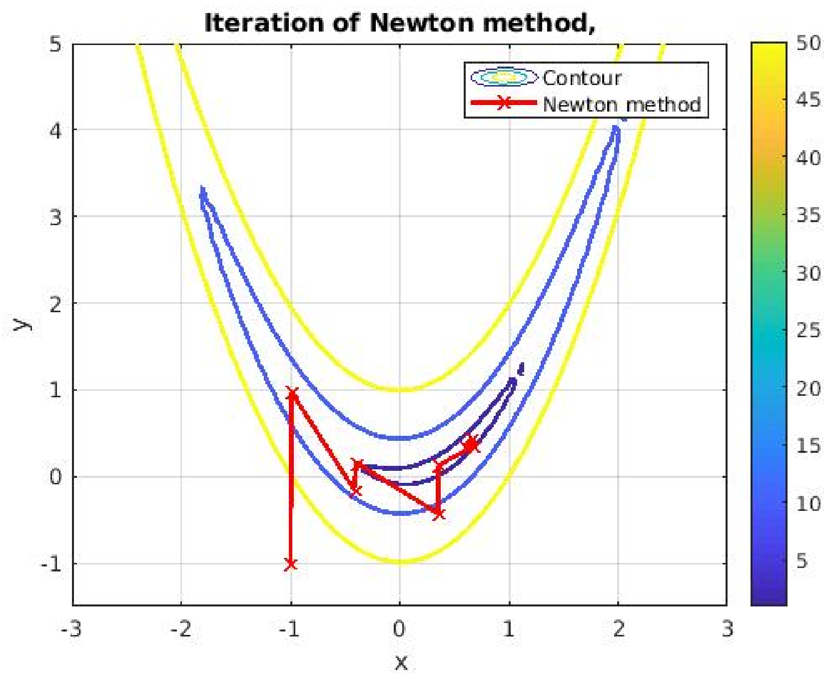

Figure 1: Newton method

3. The update of Gauss Newton's method is given by

$$\begin{bmatrix} x_{k+1} \\ y_{k+1} \end{bmatrix} = \begin{bmatrix} x_k \\ y_k \end{bmatrix} - M(x_k, y_k)^{-1} \nabla f(x_k, y_k) \tag{13}$$

the result is shown in figure 3 The algorithm converges to the optimal solution in 2295 steps,

4. The run time of Gauss-Newton method is 0.025359 seconds, while the run time of exact Newton method is 0.009419 seconds. The exact Newton method is far more faster than the Gauss-Newton method, since it makes use of the exact Hessian, and it's locally convergent with Q-quadratic rate of convergence.

   The Gauss-Newton method is more stable than the Newton method, since the latter one depends on the Hessian matrix of the objection matrix, which might be close to singular. The former one only depends on the first order information of the objective function, and is always positive semi-definite. As we can see from the two figures, though Gauss-Newton converges slowly, it avoids the Zig-Zag phenomemnon.

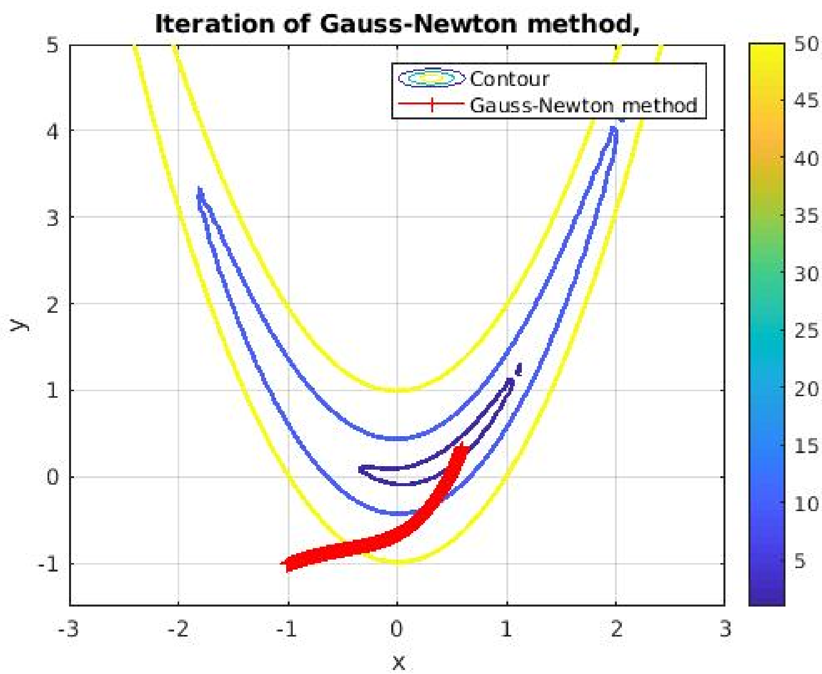

Figure 2: Gauss Newton method