Numerical Analysis, 2020 Fall

Homework 7

2018533197 高世杰

Due on 23:59 Dec 2, 2020

Problem1

1\. (a). 对求解非线性等式 $F(x)=0$

在局部用一阶泰勒展开(切线)近似原方程求解，通过迭代向原方程的解逼近，最终求得解

$F(x)\approx F(x^k)+F'(x^k)(x-x^k)=0$

如果 $F'(x^k)$ 可逆，则有 $x=x^k-F'(x^k)^{-1}F(x^k)$

取 $x^{k+1}=x=x^k-F'(x^k)^{-1}F(x^k)$ 为新展开点进行迭代

(b) 在最小二乘问题中，$\min_x F(x)$，$F(x)=\frac{1}{2}\|f(x)\|^2$，$M(x)=\nabla^2F(x)$

由Newton法可得移动方向 $\Delta x^k=x^{k+1}-x^k=-M(x_k)^{-1}f'(x_k)^Tf(x_k)$

Gauss-Newton法对Hessian矩阵 $M(x)$ 作近似

$M(x_k)=\nabla^2F(x_k)=f''(x_k)^Tf(x_k)+f'(x_k)^Tf'(x_k)$

↓ 舍去二阶项，令 $M(x_k)=f'(x_k)^Tf'(x_k)$ 作为近似的Hessian

即有 $\Delta x^k=-(f'(x_k)^Tf'(x_k))^{-1}f'(x_k)^Tf(x_k)$

(C) 二阶收敛，即 $\exists c>0, \exists K>0, \forall k>K, k\in N^*: \|x^{k+1}-x^*\|\le c\|x^k-x^*\|^2$

(d) 由Newton法有式：$\|x^{k+1}-x^*\|\le k\|x^k-x^*\|+\frac{\omega}{2}\|x^k-x^*\|^2$，其中

$\|I-M(x_k)^{-1}F''(x_k)\|=\|I-M(x_k)^{-1}[f''(x_k)^Tf(x_k)+M(x_k)]\|$

$=\|M(x_k)^{-1}f''(x_k)^Tf(x_k)\|$

$\le\|M(x_k)^{-1}\|\|f''(x_k)^T\|\|f(x_k)\|\le k$

∴得 $f''(x)=0 \Rightarrow$ 一步收敛

(e) 线搜索是选择一个目标函数值下降方向，然后计算该方向步长，进行迭代，

也即：$x^{k+1}=x^k+\alpha^kd^k$，为了保证每次函数值均下降，且选择的步长 α^k 不至于太小，

因此增加了Armijo线搜索条件：$f(x^k+\alpha^kd^k)\le f(x^k)+c_1\alpha^k\nabla f(x_k)^Td^k$

这保证了搜索的全局收敛性

e的证明：

由Armijo可求得 $\alpha^k\in(0,\frac{2(1-c_1)}{L})$ 其中L为Lipschitz常数，r为回溯参数

$\exists K>0, \forall k>K, k\in N^*: f(x^{k+1})<f(x^k)-c_1\cdot\frac{2(1-c_1)r}{L}\|\nabla f(x^k)\|^2$

记第一个满足该不等式成立的k为K

$f(x^{K+2})<f(x^{K+1})-c_1\cdot\frac{2(1-c_1)r}{L}\|\nabla f(x^{K+1})\|^2$

$c_1\cdot\frac{2(1-c_1)r}{L}\sum_{i=K}^{\infty}\|\nabla f(x^i)\|^2<\sum_{i=K}^{\infty}f(x^i)-f(x^{i+1})$

$=f(x^K)-\lim_{i\to\infty}f(x^{i+1})$

$\because \forall x\in R: f(x)\ge\underline{f}$

$\therefore <f(x^K)-\underline{f}$

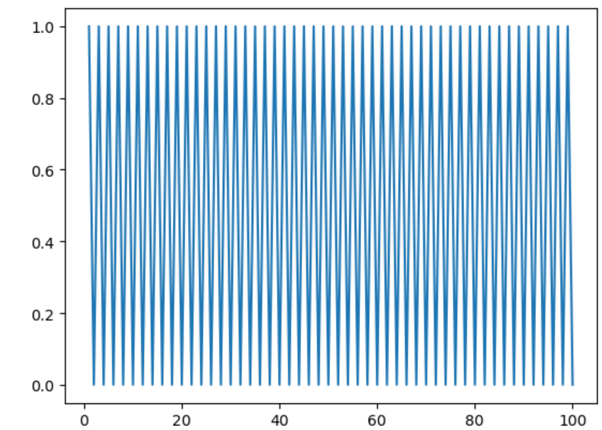
$<+\infty$

$\therefore \sum_{i=K}^{\infty}\|\nabla f(x^i)\|^2<+\infty$

$\therefore \lim_{k\to\infty}\|\nabla f(x^k)\|=0$

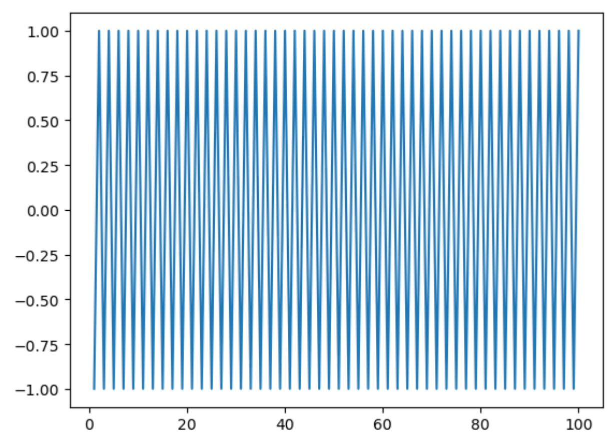
Problem2

The iteration points oscillate between 0 and 1 This is x:



1.0
0.8
0.6
0.4
0.2
0.0
0
20
40
60
80
100

This is d:



1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
0
20
40
60
80
100

1 Problem3

1.1 a.

3.(1) $f(x,y)=\frac{1}{2}(x-1)^2+\frac{1}{2}(10(y-x^2))^2+\frac{1}{2}y^2$

$$\nabla f(x,y)=\begin{pmatrix} x-1+100(y-x^2)\cdot(-2x) \\ 100(y-x^2)+y \end{pmatrix} \qquad \nabla^2 f(x,y)=\begin{pmatrix} 600x^2-200y+1 & -200x \\ -200x & 101 \end{pmatrix}$$

$$=\begin{pmatrix} 200x^3-200xy+x-1 \\ -100x^2+101y \end{pmatrix}$$

$$f(x,y)=\frac{1}{2}\|R(x,y)\|_2^2=\frac{1}{2}R(x,y)^TR(x,y)$$

$$R(x,y)=\begin{pmatrix} x-1 \\ 10(y-x^2) \\ y \end{pmatrix} \quad \nabla R(x,y)=\begin{pmatrix} 1 & 0 \\ -20x & 10 \\ 0 & 1 \end{pmatrix}$$

$$\Delta x^k=-\nabla^2 f(x^k)^{-1}\nabla f(x^k)$$

$$=-\nabla^2 f(x^k)^{-1}\nabla R(x^k)^TR(x^k)$$

$$\nabla^2 f(x^k)=\nabla[\nabla R(x^k)^TR(x^k)]$$

$$=\nabla^2 R(x^k)^TR(x^k)+\nabla R(x^k)^T\nabla R(x^k)$$

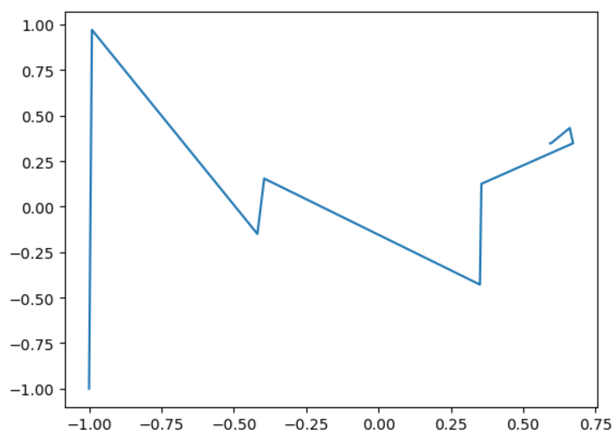
Newton: $\Delta x^k=-[\nabla^2 R(x^k)R(x^k)+\nabla R(x^k)^T\nabla R(x^k)]^{-1}\nabla R(x^k)^TR(x^k)$ $=\begin{pmatrix} 600x^2-200y+1 & -200x \\ -200x & 101 \end{pmatrix}$

Gauss: $\Delta x^k=-[\nabla R(x^k)^T\nabla R(x^k)]^{-1}\nabla R(x^k)^TR(x^k)$ $=\begin{pmatrix} 1+400x^2 & -200x \\ -200x & 101 \end{pmatrix}$

$\therefore$ when $\nabla^2 R(x^k)=0$, $[\nabla^2 R(x^k)R(x^k)+\nabla R(x^k)^T\nabla R(x^k)]$ and $\nabla R(x^k)^T\nabla R(x^k)$ are the same

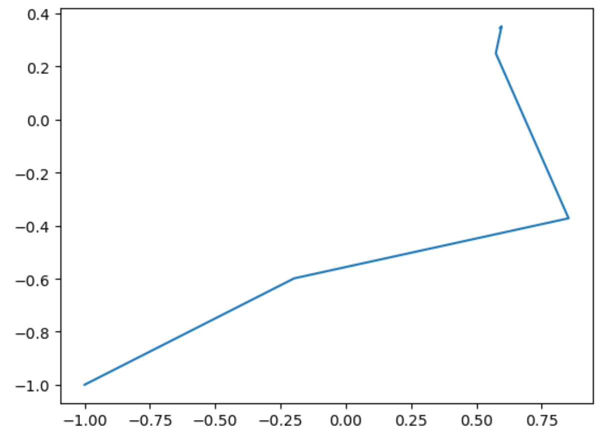
1.2 b.

Iterator times n=10

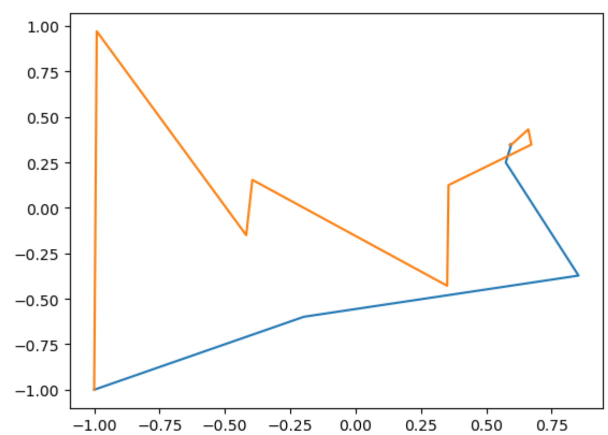


1.3 c.

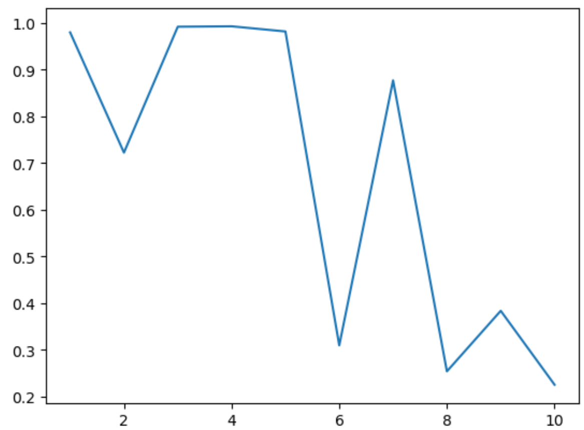
Iterator times n=7



And we plot newton and Gauss-newton together:



And difference:



1.4 d.

When the Gauss-Newton method is operational,Because The Gauss-Newton method approximates the second order information with the first order information, when in the same number of iterations,Gauss-Newton's running time is shorter.In the case of large step size, Gauss Newton's method will not go too far as Newton's method, so the number of iterations will be less, and the iteration path will basically keep walking towards the optimal direction, instead of deviating。