Numerical Optimization

Shuai Lu 170742

Homework 2

Exercise 2.1 [3 points] Let V be a linear vector space over R. Consider two vector norms $\|\cdot\|_{(1)}$ and $\|\cdot\|_{(2)}$ on V. These norms are said to be equivalent if there exists $C \in R, C > 0$, such that

$$\forall v \in V \quad \frac{1}{C}\|v\|_{(1)} \leq \|v\|_{(2)} \leq C\|v\|_{(1)}$$

(If V is finite-dimensional then all vector norms on it are equivalent.) Let $\{v_k\}_{k>1}$ be a sequence in V. Prove that this sequence converges w.r.t. one of the equivalent norms iff it converges w.r.t. the other one.

Solution:

If $\|v_k\|_{(p)}$ converges to α, it means $\forall \varepsilon > 0, \exists$N, if $k >$N, then

$$\|v_k - \alpha\|_{(p)} < \varepsilon$$

Because $\exists$C>0, st.

$$\frac{1}{C}\|v_k - \alpha\|_{(q)} \leq \|v_k - \alpha\|_{(p)} \leq C\|v_k - \alpha\|_{(q)}$$

Combine these two inequalities, we have

$$\|v_k - \alpha\|_{(q)} < C\varepsilon$$

So $\|v_k\|_{(q)}$ converges to α.

If $\|v_k - \alpha\|_{(q)} < \varepsilon$, and $\exists$C$_1$ >0, st.

$$\frac{1}{C_1}\|v_k - \alpha\|_{(p)} \leq \|v_k - \alpha\|_{(q)} \leq C_1\|v_k - \alpha\|_{(p)}$$

Combine these two inequalities, we have

$$\|v_k - \alpha\|_{(q)} < C_1\varepsilon$$

So $\|v_k\|_{(p)}$ converges to α.

Exercise 2.2 [1 point] Let $f : R \to R$ be a convex function. Consider a convergent sequence $\{x_k\}_{k\geq 1}, x_k \to x_*$, that satisfies $f(x_{k+1}) < f(x_k)$. Is x_* always a minimizer of f? Prove if yes, or find a counterexample.

Solution:

Let $f(x) := (x+1)^2$, $x_k = \frac{1}{k}$, so $f(x_{k+1}) < f(x_k)$ and $x_k \to x_*$ with $x_* = 0$. However, $f^{'}(x_*) \neq 0$, so $x_* = 0$ is not a minimizer of f.

Exercise 2.3 Consider the unconstrained minimization problem

$$f(x, y) := 5x^2 + 5y^2 - 6xy + 10x + 6y + 5 \to \min, \quad \mathbf{x} := (x, y) \in R^2$$

Find analytically its minimizer $\mathbf{x}_* = (x_*, y_*)$ and the minimum value of$f$. Besides that, for the numerical solution of this problem, implement the following iterative methods (with the exact line search):
(a) [4 points] The coordinate search method: In every iteration $k = 1, 2, \ldots, 20$, to get $\mathbf{x}_k$ from $\mathbf{x}_{k-1}$, first minimize the function only in x, then only in y.

Solution:

Analytically

$$\nabla f(x,y) = \begin{pmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{pmatrix} = \begin{pmatrix} 10x - 6y + 10 \\ 10y - 6x + 6 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \Rightarrow x = -\frac{17}{8}, y = -\frac{15}{8}$$

And

$$\nabla^2 f(x,y) = \begin{pmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{pmatrix} = \begin{pmatrix} 10 & -6 \\ -6 & 10 \end{pmatrix}$$

is positive definite. So $f(x,y)$ is convex and $(-\frac{17}{8}, -\frac{15}{8})$ is the minimizer.
The solution of coordinate search method is shown in Figure 1 and 2.

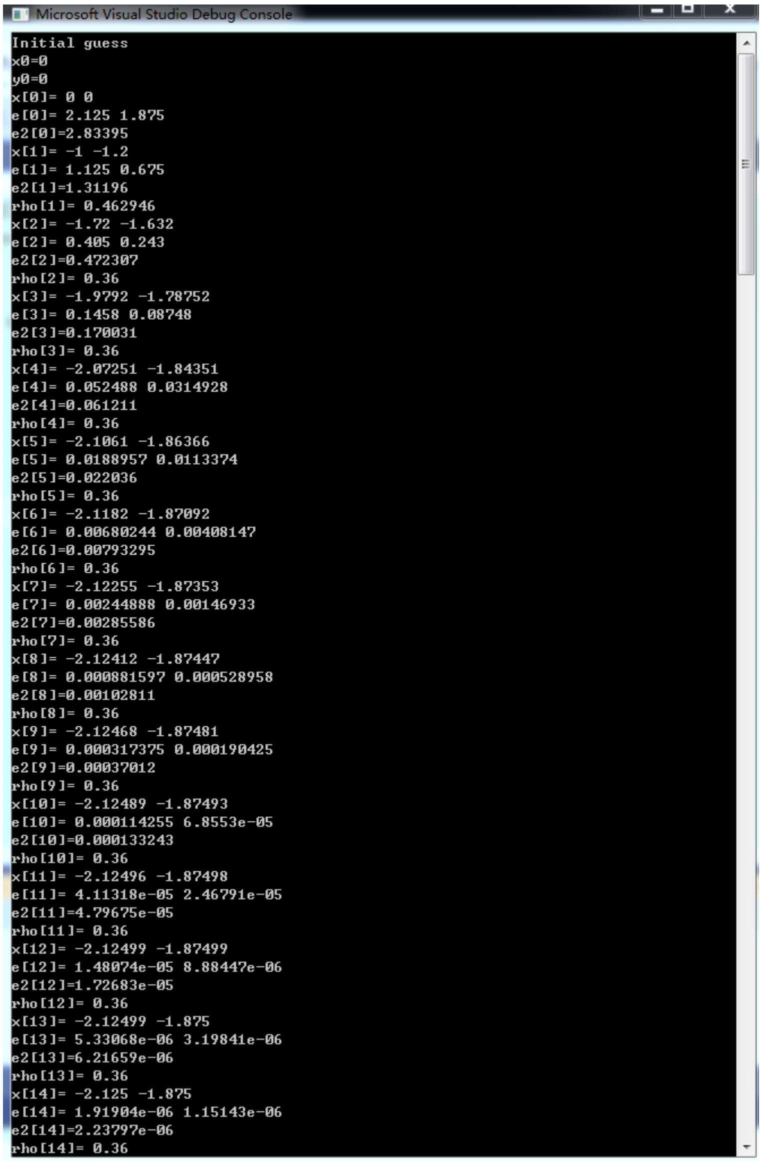


Figure 1: Coordinate search method(step 0 to 14)

(b) [4 points] The gradient method: In every iteration $k = 1, 2, \ldots, 20$, to get $\mathbf{x}_k$ from $\mathbf{x}_{k-1}$, minimize the funtion in the direction of $\mathbf{p}_{k-1} = -\nabla f(\mathbf{x}_{k-1}) \in R^2$. (Compute the gradient using the analytical formula.)
In both cases, set $\mathbf{x}_0 = 0$, and after every iteration k as well as for the initial guess ($k = 0$), print $\mathbf{x}_k, \mathbf{e}_k := \mathbf{x}_k - \mathbf{x}_*$ as well as $\|\mathbf{e}_k\|_2$. For $k \geq 1$ print also $\rho_k := \|\mathbf{e}_k\|_2 / \|\mathbf{e}_{k-1}\|_2$.
Hint: For the minimization of f in one direction, note that f is a scalar quadratic function in one argument (parabola) along every line.

Solution:

```
x[14]= -2.125 -1.875
e[14]= 1.91904e-06 1.15143e-06
e2[14]=2.23797e-06
rho[14]= 0.36
x[15]= -2.125 -1.875
e[15]= 6.90856e-07 4.14514e-07
e2[15]=8.0567e-07
rho[15]= 0.36
x[16]= -2.125 -1.875
e[16]= 2.48708e-07 1.49225e-07
e2[16]=2.90041e-07
rho[16]= 0.36
x[17]= -2.125 -1.875
e[17]= 8.95349e-08 5.3721e-08
e2[17]=1.04415e-07
rho[17]= 0.36
x[18]= -2.125 -1.875
e[18]= 3.22326e-08 1.93395e-08
e2[18]=3.75893e-08
rho[18]= 0.36
x[19]= -2.125 -1.875
e[19]= 1.16037e-08 6.96224e-09
e2[19]=1.35322e-08
rho[19]= 0.36
x[20]= -2.125 -1.875
e[20]= 4.17734e-09 2.50641e-09
e2[20]=4.87158e-09
rho[20]= 0.36
```

Figure 2: Coordinate search method(step 14 to 20)

The solution of gradient method is shown in Figure 3 and 4.

Exercise 2.4

(a) [3 points] Consider a 3 times continuously differentiable function $f : R \to R$. For the parameter (step) $h \in R, h > 0$, let

$$\delta_+ f(x) := \frac{f(x+h) - f(x)}{h}, \quad \delta_0 f(x) := \frac{f(x+h) - f(x-h)}{2h}$$

(the left and the central differences, respectively). Prove that

$$f'(x) = \delta_+ f(x) + O(h) \quad and \quad f'(x) = \delta_0 f(x) + O\left(h^2\right)$$

Compose a table with $\delta_+ f(x) - f'(x)$ and $\delta_0 f(x) - f'(x)$ for $f(x) = \sin x, x = \frac{\pi}{3}$ and $h = 0.1 \cdot 10^{-k}, k = 0, \ldots, 5$ and $h = 10^{-12}$. Use the double-precision floating poing arithmetics. How to the accuracies of the differences compare?

Solution:

The Taylor expansion of $f(x+h)$ at x:

$$f\left(x+h\right) = f\left(x\right) + f'\left(x\right)h + O(h^2)$$

So we can get

$$f'\left(x\right) = \frac{f\left(x+h\right) - f(x)}{h} + O(h) = \delta_+ f(x) + O(h)$$

The Taylor expansion of $f(x+h)$ and $f(x-h)$ at x:

$$f\left(x+h\right) = f\left(x\right) + f'\left(x\right)h + \frac{1}{2}f''\left(x\right)h_2 + O(h^3)$$

$$f\left(x-h\right) = f\left(x\right) - f'\left(x\right)h + \frac{1}{2}f''\left(x\right)h_2 + O(h^3)$$

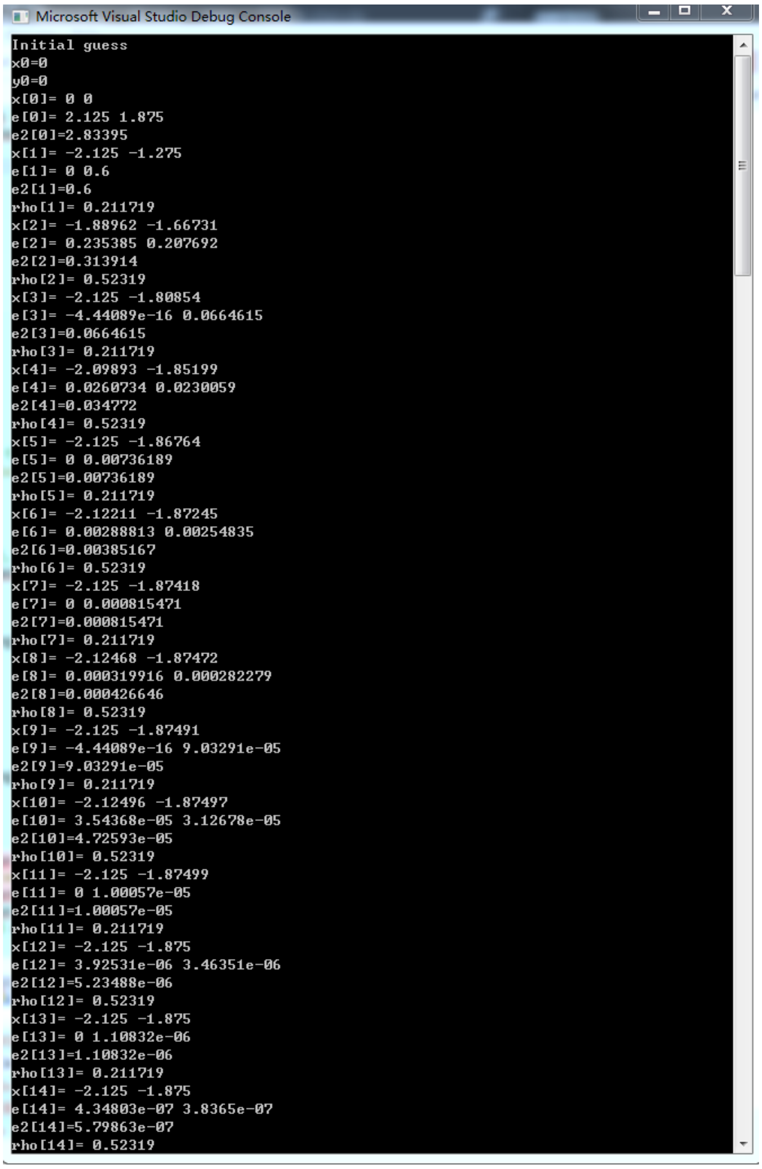
Combine these two equations, we can get

$$f'\left(x\right) = \frac{f\left(x+h\right) - f\left(x-h\right)}{2h} + O(h^2) = \delta_0 f(x) + O(h^2)$$

The error of left and cental differences approximation is shown in Figure 5. The accuracy of the central difference approximation is much higher than that of the left difference approximation.

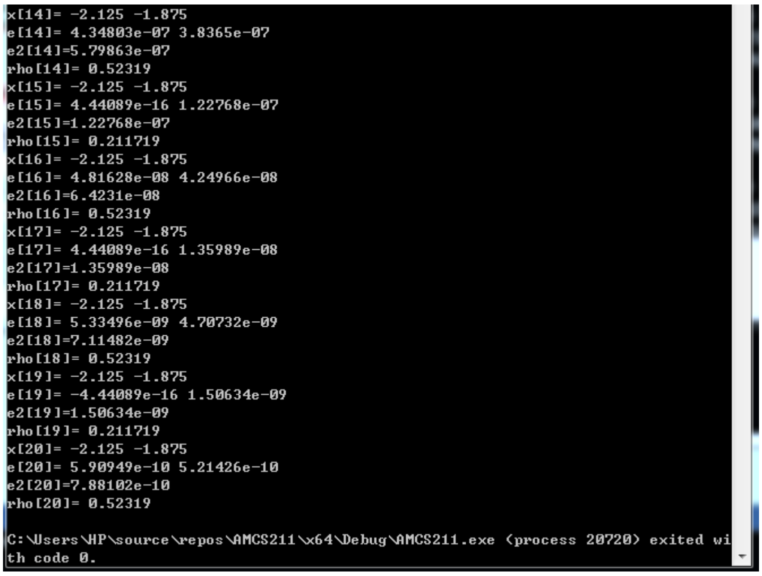
(b) [1 point] How many evaluations of $\mathbf{f} : R^n \to R$ are required to approximate $\nabla \mathbf{f}$ at some $\mathbf{x} \in R^n$ by (1) left differences, (2) central differences.

Solution:



```
Microsoft Visual Studio Debug Console
Initial guess
x0=0
y0=0
x[0]= 0 0
e[0]= 2.125 1.875
e2[0]=2.83395
x[1]= -2.125 -1.275
e[1]= 0 0.6
e2[1]=0.6
rho[1]= 0.211719
x[2]= -1.88962 -1.66731
e[2]= 0.235385 0.207692
e2[2]=0.313914
rho[2]= 0.52319
x[3]= -2.125 -1.80854
e[3]= -4.44089e-16 0.0664615
e2[3]=0.0664615
rho[3]= 0.211719
x[4]= -2.09893 -1.85199
e[4]= 0.0260734 0.0230059
e2[4]=0.034772
rho[4]= 0.52319
x[5]= -2.125 -1.86764
e[5]= 0 0.00736189
e2[5]=0.00736189
rho[5]= 0.211719
x[6]= -2.12211 -1.87245
e[6]= 0.00288813 0.00254835
e2[6]=0.00385167
rho[6]= 0.52319
x[7]= -2.125 -1.87418
e[7]= 0 0.000815471
e2[7]=0.000815471
rho[7]= 0.211719
x[8]= -2.12468 -1.87472
e[8]= 0.000319916 0.000282279
e2[8]=0.000426646
rho[8]= 0.52319
x[9]= -2.125 -1.87491
e[9]= -4.44089e-16 9.03291e-05
e2[9]=9.03291e-05
rho[9]= 0.211719
x[10]= -2.12496 -1.87497
e[10]= 3.54368e-05 3.12678e-05
e2[10]=4.72593e-05
rho[10]= 0.52319
x[11]= -2.125 -1.87499
e[11]= 0 1.00057e-05
e2[11]=1.00057e-05
rho[11]= 0.211719
x[12]= -2.125 -1.875
e[12]= 3.92531e-06 3.46351e-06
e2[12]=5.23488e-06
rho[12]= 0.52319
x[13]= -2.125 -1.875
e[13]= 0 1.10832e-06
e2[13]=1.10832e-06
rho[13]= 0.211719
x[14]= -2.125 -1.875
e[14]= 4.34803e-07 3.8365e-07
e2[14]=5.79863e-07
rho[14]= 0.52319
```

Figure 3: Gradient method(step 0 to 14)



```
x[14]= -2.125 -1.875
e[14]= 4.34803e-07 3.8365e-07
e2[14]=5.79863e-07
rho[14]= 0.52319
x[15]= -2.125 -1.875
e[15]= 4.44089e-16 1.22768e-07
e2[15]=1.22768e-07
rho[15]= 0.211719
x[16]= -2.125 -1.875
e[16]= 4.81628e-08 4.24966e-08
e2[16]=6.4231e-08
rho[16]= 0.52319
x[17]= -2.125 -1.875
e[17]= 4.44089e-16 1.35989e-08
e2[17]=1.35989e-08
rho[17]= 0.211719
x[18]= -2.125 -1.875
e[18]= 5.33496e-09 4.70732e-09
e2[18]=7.11482e-09
rho[18]= 0.52319
x[19]= -2.125 -1.875
e[19]= -4.44089e-16 1.50634e-09
e2[19]=1.50634e-09
rho[19]= 0.211719
x[20]= -2.125 -1.875
e[20]= 5.90949e-10 5.21426e-10
e2[20]=7.88102e-10
rho[20]= 0.52319

C:\Users\HP\source\repos\AMCS211\x64\Debug\AMCS211.exe (process 20720) exited wi
th code 0.
```

Figure 4: Gradient method(step 14 to 20)

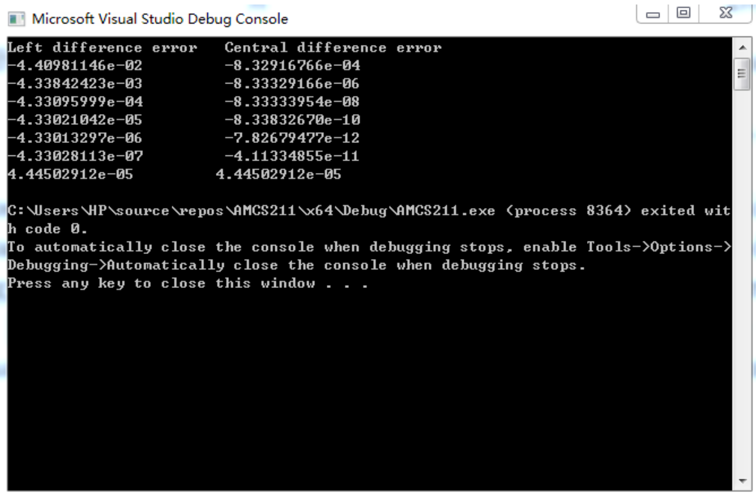


Figure 5: The error of left and cental differences approximation

For left difference approximation:

$$\triangle x^T \nabla f\left(x\right) = f\left(x + \triangle x\right) - f\left(x\right)$$

There are n unknown components of $\nabla \mathbf{f}$ in this equation, so we need n equations to solve out $\nabla \mathbf{f}$. So we need n times evaluation for $f\left(x + \triangle x\right)$ and 1 time for $f(x)$.
For central difference approximation:

$$\triangle x^T \nabla f\left(x\right) = \frac{f\left(x + \triangle x\right) - f\left(x - \triangle x\right)}{2}$$

We need n times evaluation for $f\left(x + \triangle x\right)$ and n times for $f\left(x - \triangle x\right)$.