

# Numerical Optimization

Shuai Lu 170742

## Homework 3

**Exercise 3.1** (a) [3 points] Given  $c_1 \in (0, 1)$ , prove that for any smooth function  $f : R^n \rightarrow R$ , point  $\mathbf{x} \in R^n$  and a decent direction  $\mathbf{p} \in R^n$  at that point, there is an interval for  $\alpha > 0$  such that the Armijo(or sufficient decrease) condition

$$f(\mathbf{x} + \alpha\mathbf{p}) \leq f(\mathbf{x}) + c_1\alpha[\nabla f(\mathbf{x})]^\top \mathbf{p}$$

holds. (This condition guarantees that  $f$  is actually reduced by the doing the step  $\alpha\mathbf{p}$ . It excludes too long steps: Sufficiently small  $\alpha$  would always satisfy this condition.)

Solution:

Let  $\phi(\alpha) = f(\mathbf{x} + \alpha\mathbf{p})$  and  $l(\alpha) = f(\mathbf{x}) + c_1\alpha[\nabla f(\mathbf{x})]^\top \mathbf{p}$ . Since  $\phi(0) = l(0)$ ,  $\phi'(0) = [\nabla f(\mathbf{x})]^\top \mathbf{p} < c_1[\nabla f(\mathbf{x})]^\top \mathbf{p} = l'(0)$ ,  $\phi(\alpha)$  is bounded below and  $l(\alpha)$  is unbounded below, there is a first intersection point  $\alpha_0$  that  $f(\mathbf{x} + \alpha_0\mathbf{p}) = f(\mathbf{x}) + c_1\alpha_0[\nabla f(\mathbf{x})]^\top \mathbf{p}$ . When  $0 < \alpha \leq \alpha_0$ ,

$$\phi(\alpha) = f(\mathbf{x} + \alpha\mathbf{p}) \leq l(\alpha) = f(\mathbf{x}) + c_1\alpha[\nabla f(\mathbf{x})]^\top \mathbf{p}$$

(b) [1 point] Consider

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

$\mathbf{x} = (0, 0)^\top$ ,  $\mathbf{p} = -\nabla f(\mathbf{x})$  and  $c_1 = 0.01$ . Find largest possible interval where the Armijo condition holds.

Solution:

$$\mathbf{x} = (0, 0)^\top, \mathbf{p} = -\nabla f(\mathbf{x}) = -\begin{pmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{pmatrix} = -\begin{pmatrix} 10 \\ 6 \end{pmatrix}, \text{ Armijo condition is:}$$

$$500\alpha^2 + 180\alpha^2 - 360\alpha^2 - 100\alpha - 36\alpha + 5 \leq 5 - 0.01(100 + 36)\alpha$$

So  $0 < \alpha \leq 0.42925$

**Exercise 3.2** (a) [3 points] Given  $c_2 \in (c_1, 1)$ , prove that for any smooth function  $f : R^n \rightarrow R$ , point  $\mathbf{x} \in R^n$  and a decent direction  $\mathbf{p} \in R^n$  at that point, there is an interval for  $\alpha > 0$  such that the curvature condition

$$[\nabla f(\mathbf{x} + \alpha\mathbf{p})]^\top \mathbf{p} \geq c_2[\nabla f(\mathbf{x})]^\top \mathbf{p}$$

holds. (This condition excludes too short steps so that the steps do not deteriorate during the iteration.)

Solution:

From Exercise 3.1, we got  $\alpha_0$  such that

$$f(\mathbf{x} + \alpha_0\mathbf{p}) = f(\mathbf{x}) + c_1\alpha_0[\nabla f(\mathbf{x})]^\top \mathbf{p}$$

By the mean value theorem, there exists  $\alpha_1 \in (0, \alpha_0)$  such that

$$f(\mathbf{x} + \alpha_0\mathbf{p}) - f(\mathbf{x}) = \alpha_0[\nabla f(\mathbf{x} + \alpha_1\mathbf{p})]^\top \mathbf{p}$$

By combining these two equations, we have

$$[\nabla f(\mathbf{x} + \alpha_1\mathbf{p})]^\top \mathbf{p} = c_1[\nabla f(\mathbf{x})]^\top \mathbf{p} > c_2[\nabla f(\mathbf{x})]^\top \mathbf{p}$$

Since the continuity of  $\nabla f(\mathbf{x})$  there exist a  $\delta$ , when  $\alpha \in U(\alpha_1, \delta)$ ,

$$[\nabla f(\mathbf{x} + \alpha\mathbf{p})]^\top \mathbf{p} \geq c_2[\nabla f(\mathbf{x})]^\top \mathbf{p}$$

(b) [1 point] Consider  $f(x, y)$  from Exercise 3.1(b),  $\mathbf{x} = (0, 0)^\top$ ,  $\mathbf{p} = -\nabla f(\mathbf{x})$  and  $c_2 = 0.9$ . Find largest possible interval where the curvature condition holds.

Solution:

From the curvature condition:

$$[\nabla f(\mathbf{x} + \alpha \mathbf{p})]^\top \mathbf{p} \geq c_2 [\nabla f(\mathbf{x})]^\top \mathbf{p}$$

We have:

$$-\begin{pmatrix} -64\alpha + 10 \\ 6 \end{pmatrix}^\top \begin{pmatrix} 10 \\ 6 \end{pmatrix} \geq -0.9 \begin{pmatrix} 10 \\ 6 \end{pmatrix}^\top \begin{pmatrix} 10 \\ 6 \end{pmatrix}$$

So  $\alpha \geq 0.02125$

**Exercise 3.3** (a) [4 points] Reimplement your code from Exercise 2.3(b) (the gradient method) for the same quadratic function  $f(x, y)$  with the following backtracking linesearch strategy: In the  $k$ -th iteration, having  $\mathbf{x}_k$  and  $\mathbf{p}_k$  compute successively

$$\mathbf{x}_{k+1}^{(i)} = \mathbf{x}_k + s_k^{(i)} \mathbf{p}_k, \quad i = 0, \dots, n_{\text{ls}}$$

with  $s_k^{(i)} = 2^{-i}$  (i.e.  $s_k^{(0)} = 1, s_k^{(i+1)} = \frac{1}{2}s_k^{(i)}$  - backtracking). Choose the smallest  $i =: i^*$  for which the sufficient descent is achieved:

$$f(\mathbf{x}_k + s_k^{(i)} \mathbf{p}_k) \leq f(\mathbf{x}_k) + \sigma s_k^{(i)} [\nabla f(\mathbf{x}_k)]^\top \mathbf{p}_k.$$

(If this condition does not hold for any  $i = 0, \dots, n_{\text{ls}}$ , take  $i^* = n_{\text{ls}}$ .) Then set  $\mathbf{x}_{k+1} = \mathbf{x}_k + s_k^{(i^*)} \mathbf{p}_k$ . Use the parameters  $n_{\text{ls}} = 8$  and  $\sigma = 10^{-3}$ . Start with  $\mathbf{x}_0 = (1, 0.2)^\top$  and compute  $\mathbf{x}_1, \dots, \mathbf{x}_5$ . After every iteration  $k$  as well as for the initial guess ( $k = 0$ ), print  $\mathbf{x}_k, f(\mathbf{x}_k), \|\nabla f(\mathbf{x}_k)\|_2$  as well as  $\|\mathbf{e}_k\|_2$  for  $\mathbf{e}_k := \mathbf{x}_k - \mathbf{x}_*$ . For  $k \geq 1$ , print also  $\rho_k := \|\mathbf{e}_k\|_2 / \|\mathbf{e}_{k-1}\|_2$ .

Solution:

The output of the program is shown in Figure 1 and 2.

(b) [2 points] Extend the method implemented in Exercise 3.3(a) with the following termination criterion: Compute  $\mathbf{x}_1, \dots, \mathbf{x}_K$  until  $\|\nabla f(\mathbf{x}_k)\|_2 \leq \epsilon = 10^{-8}$  but stop after at most  $K_{\text{max}} = 50$  iterations.

Solution:

The output of the program is shown in Figure 3.

**Exercise 3.4** [2 points] Characterize the type of the convergence of the coordinate and gradient methods in Exercise 2.3: Is it linear, superlinear or quadratic? Explain why.

Solution:

$$f(x, y) = \begin{pmatrix} x \\ y \end{pmatrix}^\top \begin{pmatrix} 5 & -3 \\ -3 & 5 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 10 \\ 6 \end{pmatrix}^\top \begin{pmatrix} x \\ y \end{pmatrix} + 5$$

$Q = \begin{pmatrix} 5 & -3 \\ -3 & 5 \end{pmatrix}$ , the eigenvalues of  $Q$  are 8 and 2. For exact line search, we have

$$\|x_{k+1} - x^*\|_Q^2 \leq \left( \frac{\lambda_n - \lambda_1}{\lambda_n + \lambda_1} \right)^2 \|x_k - x^*\|_Q^2 = 0.36 \|x_k - x^*\|_Q^2$$

So it's linear.

```
Microsoft Visual Studio Debug Console

Initial guess
x0=1
y0=0.2
x[0]= 1 0.2
f[0]= 20.2
df2[0]=18.9061
e2[0]=3.75117
x[1]= -1.35 -0.05
f[1]= -0.08
df2[1]=13.9714
e2[1]=1.98274
rho[1]= 0.528566
x[2]= -0.95 -1.75
f[2]= -5.15
df2[2]=12.4354
e2[2]=1.18163
rho[2]= 0.595959
x[3]= -2.325 -1.025
f[3]= -6.4175
df2[3]=12.0208
e2[3]=0.873212
rho[3]= 0.73899
x[4]= -1.4375 -2.2375
f[4]= -6.73437
df2[4]=11.9149
e2[4]=0.777215
rho[4]= 0.890064
x[5]= -2.56875 -1.26875
f[5]= -6.81359
df2[5]=11.8883
e2[5]=0.751301
rho[5]= 0.966658
x[6]= -1.55938 -2.35938
f[6]= -6.8334
df2[6]=11.8816
e2[6]=0.744682
rho[6]= 0.99119
x[7]= -2.09453 -1.84453
f[7]= -11.2463
df2[7]=0.172357
e2[7]=0.0430893
rho[7]= 0.0578627
x[8]= -2.125 -1.875
f[8]= -11.25
df2[8]=0
e2[8]=6.28037e-16
rho[8]= 1.45752e-14
x[9]= -2.125 -1.875
f[9]= -11.25
df2[9]=0
e2[9]=6.28037e-16
rho[9]= 1
x[10]= -2.125 -1.875
f[10]= -11.25
df2[10]=0
e2[10]=6.28037e-16
rho[10]= 1
x[11]= -2.125 -1.875
f[11]= -11.25
df2[11]=0
e2[11]=6.28037e-16
rho[11]= 1
```

Figure 1: The output of backtracking linesearch(step 1 to 11)

```
Microsoft Visual Studio Debug Console
x[11]= -2.125 -1.875
f[11]= -11.25
df2[11]=0
e2[11]=6.28037e-16
rho[11]= 1
x[12]= -2.125 -1.875
f[12]= -11.25
df2[12]=0
e2[12]=6.28037e-16
rho[12]= 1
x[13]= -2.125 -1.875
f[13]= -11.25
df2[13]=0
e2[13]=6.28037e-16
rho[13]= 1
x[14]= -2.125 -1.875
f[14]= -11.25
df2[14]=0
e2[14]=6.28037e-16
rho[14]= 1
x[15]= -2.125 -1.875
f[15]= -11.25
df2[15]=0
e2[15]=6.28037e-16
rho[15]= 1
x[16]= -2.125 -1.875
f[16]= -11.25
df2[16]=0
e2[16]=6.28037e-16
rho[16]= 1
x[17]= -2.125 -1.875
f[17]= -11.25
df2[17]=0
e2[17]=6.28037e-16
rho[17]= 1
x[18]= -2.125 -1.875
f[18]= -11.25
df2[18]=0
e2[18]=6.28037e-16
rho[18]= 1
x[19]= -2.125 -1.875
f[19]= -11.25
df2[19]=0
e2[19]=6.28037e-16
rho[19]= 1
x[20]= -2.125 -1.875
f[20]= -11.25
df2[20]=0
e2[20]=6.28037e-16
rho[20]= 1
```

Figure 2: The output of backtracking linesearch(step 11 to 20)

```
Microsoft Visual Studio Debug Console

Initial guess
x0=1
y0=0.2
x[0]= 1 0.2
f[0]= 20.2
df2[0]=1.891e+01
e2[0]=3.751e+00
x[1]= -1.350e+00 -5.000e-02
f[1]= -8.000e-02
df2[1]=1.397e+01
e2[1]=1.983e+00
rho[1]= 5.286e-01
x[2]= -9.500e-01 -1.750e+00
f[2]= -5.150e+00
df2[2]=1.244e+01
e2[2]=1.182e+00
rho[2]= 5.960e-01
x[3]= -2.325e+00 -1.025e+00
f[3]= -6.417e+00
df2[3]=1.202e+01
e2[3]=8.732e-01
rho[3]= 7.390e-01
x[4]= -1.438e+00 -2.237e+00
f[4]= -6.734e+00
df2[4]=1.191e+01
e2[4]=7.772e-01
rho[4]= 8.901e-01
x[5]= -2.569e+00 -1.269e+00
f[5]= -6.814e+00
df2[5]=1.189e+01
e2[5]=7.513e-01
rho[5]= 9.667e-01
x[6]= -1.559e+00 -2.359e+00
f[6]= -6.833e+00
df2[6]=1.188e+01
e2[6]=7.447e-01
rho[6]= 9.912e-01
x[7]= -2.095e+00 -1.845e+00
f[7]= -1.125e+01
df2[7]=1.724e-01
e2[7]=4.309e-02
rho[7]= 5.786e-02
x[8]= -2.125e+00 -1.875e+00
f[8]= -1.125e+01
df2[8]=0.000e+00
e2[8]=6.280e-16
rho[8]= 1.458e-14
```

Figure 3: The output of backtracking linesearch(step 1 to 8)