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 $f : R^n \to R$, 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 α 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 α_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

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

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

Solution:
$\mathbf{x} = (0, 0)^\top, \mathbf{p} = -\nabla f(\mathbf{x}) = -\begin{pmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{pmatrix} = -\begin{pmatrix} 10 \\ 6 \end{pmatrix}$, 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 $f : R^n \to R$, 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 α_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 δ, 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 methdo) 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

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

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:

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

(If this condition does not hold for any $i=0,\ldots,n_{\mathrm{ls}}$, take $i^*=n_{\mathrm{ls}}$.) Then set $\mathbf{x}_{k+1}=\mathbf{x}_k+s_k^{(i^*)}\mathbf{p}_k$. Use the parameters $n_{\mathrm{ls}}=8$ and $\sigma=10^{-3}$. Start with $\mathbf{x}_0=(1,0.2)^\top$ and compute $\mathbf{x}_1,\ldots,\mathbf{x}_5$. 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 $\mathbf{x}_1,\ldots,\mathbf{x}_K$ until $\|\nabla f(\mathbf{x}_k)\|_2 \leq \epsilon = 10^{-8}$ but stop after at most $K_{\max}=50$ 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.

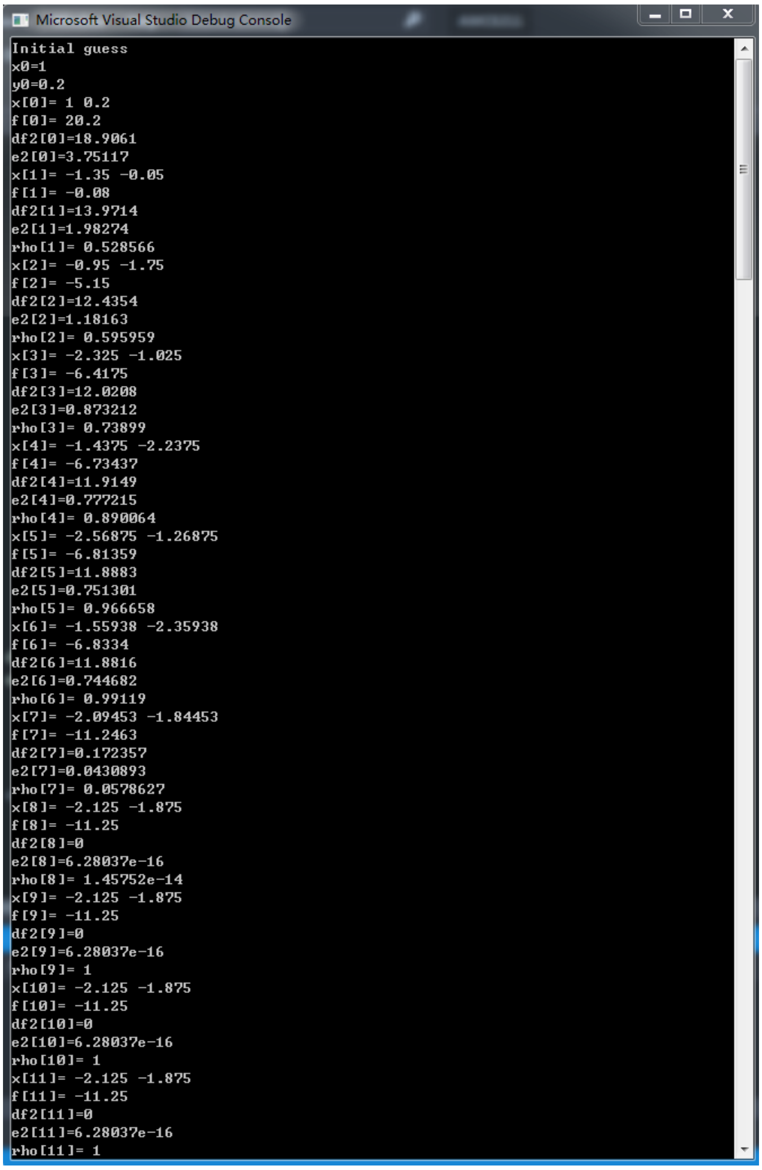

Figure 1: The output of backtracking linesearch(step 1 to 11)

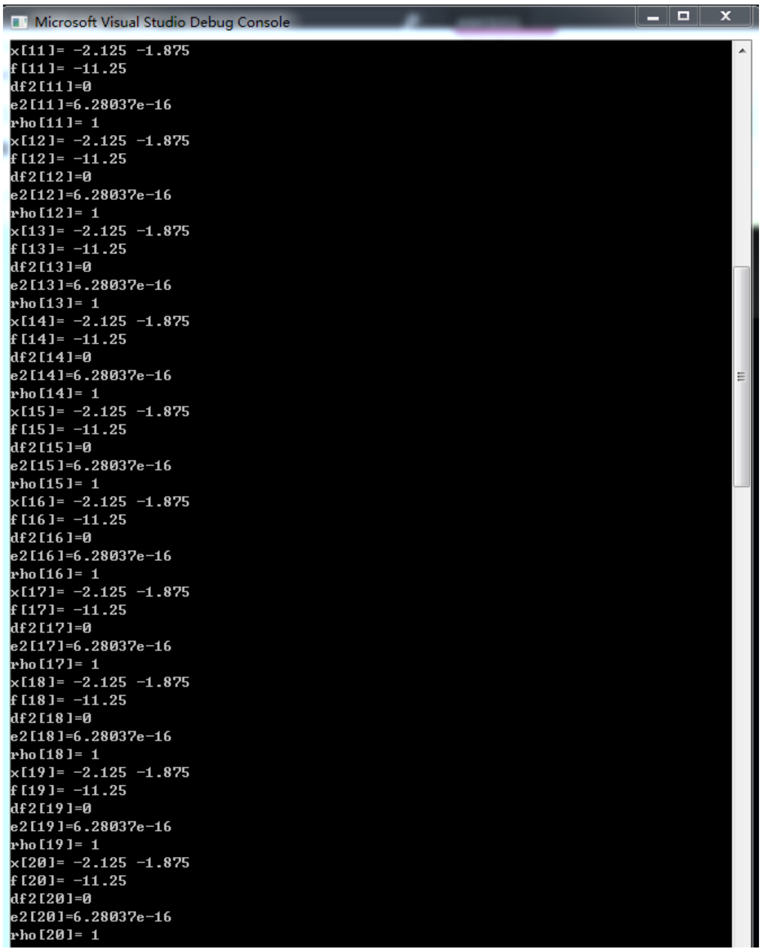

Figure 2: The output of backtracking linesearch(step 11 to 20)

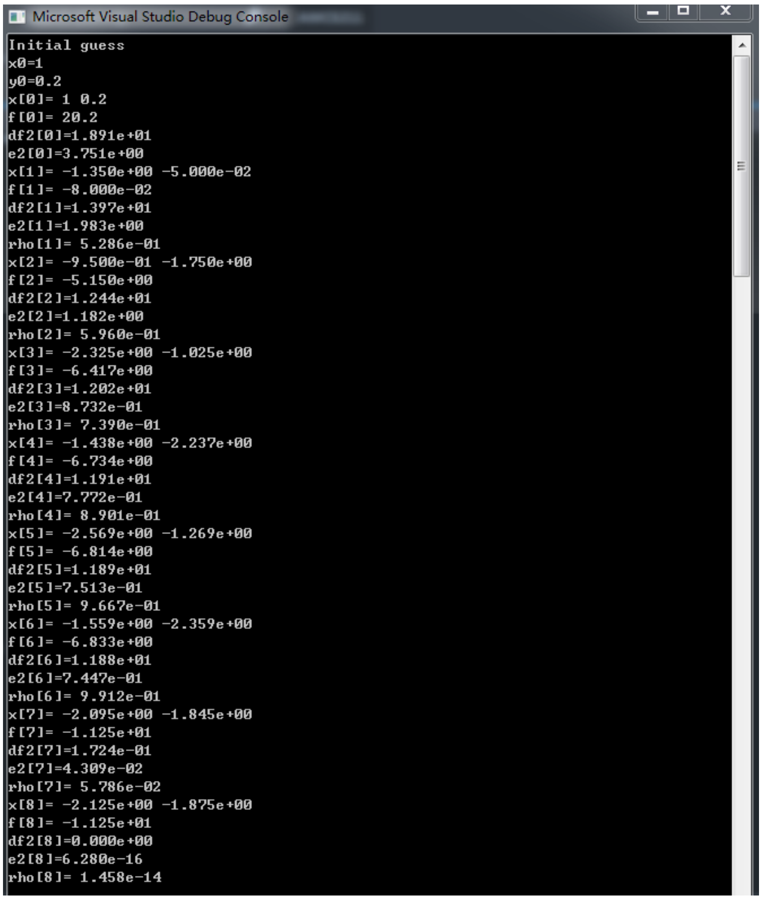

Figure 3: The output of backtracking linesearch(step 1 to 8)