ACM 113 Introduction to Optimization - Problem Set 1

Ling Li, ling@cs.caltech.edu

April 16, 2001

1.1 Let $x^* = (x_1^*, x_2^*)^T$ be any local minimizer. $f(x) = (x_1 - 2x_2)^2 + x_1^4$ gives

$$\begin{aligned} \nabla f(x) &= \begin{pmatrix} 2(x_1 - 2x_2) + 4x_1^3 \\ -4(x_1 - 2x_2) \end{pmatrix}, \\ \nabla^2 f(x) &= \begin{pmatrix} 2 + 12x_1^2 & -4 \\ -4 & 8 \end{pmatrix}. \end{aligned}$$

Thus the first- and second-order derivatives of f are continuous in $\mathbb{R}^2$. From the first-order necessary condition for a local minimizer, we have

$$\nabla f(x^*) = \begin{pmatrix} 2(x_1^* - 2x_2^*) + 4x_1^{*3} \\ -4(x_1^* - 2x_2^*) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Thus $x_1^* = x_2^* = 0$, i.e., $x^* = (0,0)^T$. Now

$$\nabla^2 f_* = \nabla^2 f(x^*) = \begin{pmatrix} 2 & -4 \\ -4 & 8 \end{pmatrix}$$

is positive semi-definite, since

$$x^T \nabla^2 f_* x = 2(x_1 - 2x_2)^2 \tag{1}$$

is non-negative for any $x = (x_1, x_2)^T \in \mathbb{R}^2$. Though $\nabla^2 f(x^*)$ is not positive definite, since (1) is 0 when $x = (2,1)^T$, we can prove that $x^* = (0,0)^T$ is the only local minimizer, and is also a global minimizer: $f(x) = (x_1 - 2x_2)^2 + x_1^4 \geq 0 = f(x^*)$, with equality iff $x_1 = 2x_2$ and $x_1 = 0$, which is just $x = x^*$. Thus $x^* = (0,0)^T$ is the only (global) minimizer.

1.2 *Rosenbrock function.* $f(x) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$. Let $x^* = (x_1^*, x_2^*)^T$ be any local minimizer.

$$\begin{aligned} \nabla f(x) &= \begin{pmatrix} -400(x_2 - x_1^2)x_1 - 2(1 - x_1) \\ 200(x_2 - x_1^2) \end{pmatrix}, \\ \nabla^2 f(x) &= \begin{pmatrix} 1200x_1^2 - 400x_2 + 2 & -400x_1 \\ -400x_1 & 200 \end{pmatrix}. \end{aligned}$$

The first- and second-order derivatives of f are continuous in $\mathbb{R}^2$. From the first-order necessary condition for a local minimizer, $\nabla f(x^*) = \mathbf{0}$ gives $x_1^* = x_2^* = 1$. Now

$$\nabla^2 f_* = \nabla^2 f(x^*) = \begin{pmatrix} 802 & -400 \\ -400 & 200 \end{pmatrix}$$

is positive definite, since for $x = (x_1, x_2)^T \in \mathbb{R}^2$ and $x \neq \mathbf{0}$,

$$x^T \nabla^2 f_* x = 2x_1^2 + 200(2x_1 - x_2)^2 > 0.$$

Thus by the second-order sufficient condition for local minimizer, we know x^* is a strict local minimizer. And above discussion also shows x^* is the only local minimizer.

1.3 Note that Q is not necessary symmetric. Let $\tilde{Q} = \frac{1}{2}(Q + Q^T)$. From $\tilde{Q}^T = \frac{1}{2}(Q^T + Q) = \tilde{Q}$ and $x^T\tilde{Q}x = \frac{1}{2}(x^TQx + x^TQ^Tx) = \frac{1}{2}(x^TQx + (x^TQx)^T) = x^TQx$, we know $\tilde{Q}$ is symmetric and positive definite. Thus $\tilde{Q}^{-1}$ exists.

$f(x) = \frac{1}{2}x^TQx - c^Tx$. Thus

$$\nabla f(x) = \tilde{Q}x - c, \quad \nabla^2 f = \tilde{Q}\,. \tag{2}$$

From the start point x_0, Newton's method gives

$$p_0^N = -(\nabla^2 f_0)^{-1}\nabla f_0 = -\tilde{Q}^{-1}(\tilde{Q}x_0 - c) = -x_0 + \tilde{Q}^{-1}c\,.$$

Hence $x_1 = x_0 + p_0^N = \tilde{Q}^{-1}c$. It is easy to verify that $\nabla f(x_1) = \mathbf{0}$ and $\nabla^2 f(x_1) = \tilde{Q}$ is positive definite. Thus x_1 is a strict local minimizer of f. (The minimum of f is $-\frac{1}{2}c^T\tilde{Q}^{-1}c$.)

When applying Newton's method to a function f, we first approximate f by the first 3 items of its Taylor series. Then we use the first-order necessary condition to get descent direction p_k. So it is not surprising that when f itself is quadratic, Newton's method would find the local minimum in one step.

1.4 Define $g(\alpha) = f(x_k + \alpha p_k)$, where f is defined in Problem 1.3. From

$$\nabla f(x_k + \alpha p_k) = \tilde{Q}(x_k + \alpha p_k) - c^T = (\tilde{Q}x_k - c^T) + \alpha\tilde{Q}p_k = \nabla f_k + \alpha\tilde{Q}p_k,$$

we get $g'(\alpha) = p_k^T\nabla f(x_k + \alpha p_k) = p_k^T\nabla f_k + \alpha p_k^T\tilde{Q}p_k$. Solving $g'(\alpha_k) = 0$ gives

$$\alpha_k = -\frac{p_k^T\nabla f_k}{p_k^T\tilde{Q}p_k}. \tag{3}$$

Since p_k is a descent direction, we have $p_k \neq \mathbf{0}$ and $p_k^T\nabla f_k < 0$. Since $\tilde{Q}$ is positive definite, we also have $p_k^T\tilde{Q}p_k > 0$. Thus $\alpha_k > 0$. From $g''(\alpha) = p_k^T\tilde{Q}p_k > 0$, we know

$$\alpha_k = \arg\min_{\alpha>0} g(\alpha) = \arg\min_{\alpha>0} f(x_k + \alpha p_k).$$

Thus we proved that the line search should use the step length given by (3).

1.5 *Line search.* Starting from x_k, the deepest descent direction is (see (2))

$$p_k = -\nabla f(x_k) = c - \tilde{Q}x_k\,.$$

Then from (3), we can get α_k and then calculate

$$x_{k+1} = x_k + \alpha_k p_k\,.$$

Here we use $c = (6, 3)^T$, $x_0 = (-4, 5)^T$, and

$$Q_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad Q_2 = \begin{pmatrix} 1 & 0 \\ 0 & 10 \end{pmatrix}, \quad Q_3 = \begin{pmatrix} 1 & 0 \\ 0 & 100 \end{pmatrix}$$

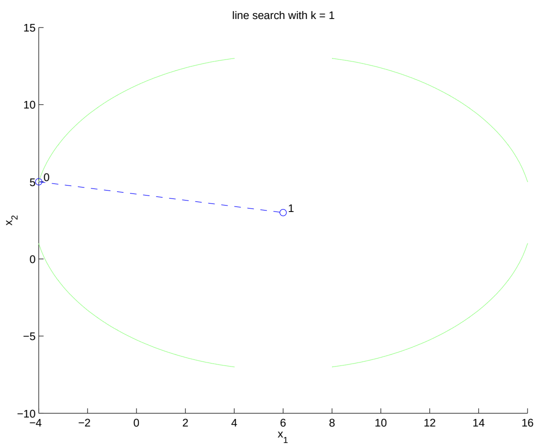
as examples with condition numbers $\kappa \in \{1, 10, 100\}$. The theoretical bound constants computed by

$$C = \left(\frac{\kappa - 1}{\kappa + 1}\right)^2$$

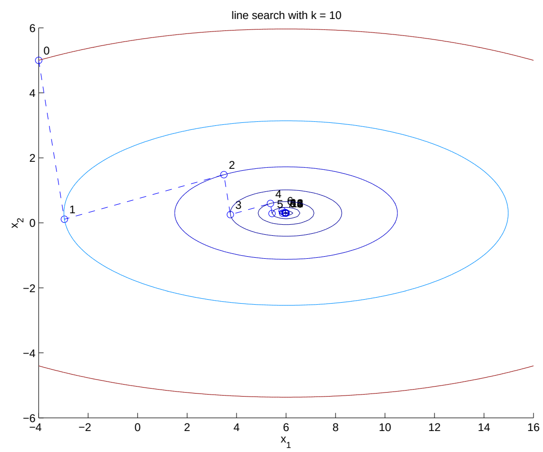
and the estimated constants C_e are

κ	1	10	100
C	0	0.66942	0.96079
C_e	0	0.25128	0.038135

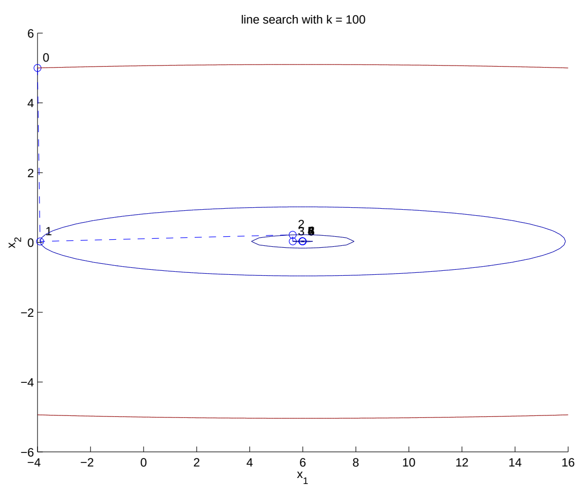
The iteration of x_k is shown in Figure 1.



(a) $\kappa = 1$



(b) $\kappa = 10$



(c) $\kappa = 100$

Figure 1: Line search with $\kappa = 1, 10, 100$. The small circles with numbers are x_k's, and the dashed lines give the descent direction p_k's. Those ellipses are contours of the object function.