Norwegian University of Science and Technology
Department of Mathematical Sciences

TMA4180 Optimization Theory Spring 2016

Exercise set 2 solution

1 **a)** We begin by recalling that if Q is symmetric and positive definite, then so is Q^{-1}. This will be useful later. Next, we observe that

$$\begin{aligned} f(x) &= \frac{1}{2}x^TQx - b^Tx \\ \nabla f(x) &= Qx - b \\ \nabla^2 f(x) &= Q. \end{aligned}$$

Since the Hessian of f is positive definite, f is strictly convex (see Remark 8, lecture note no. 2). From exercise set 1, we know that a strictly convex function has a unique global minimizer, and from Corollary 6 in note no. 2, we know that this is obtained at the point x^* where $\nabla f(x^*) = 0$, i.e. the global minimizer of f is

$$x^* = Q^{-1}b.$$

b) When using Newton's method, the search direction is $p_k = -(\nabla^2 f_k)^{-1}\nabla f_k$, where we have written $\nabla f_k = \nabla f(x_k)$ and $\nabla^2 f_k = \nabla^2 f(x_k)$ for short. In our case, this gives

$$p_k = -x_k + Q^{-1}b.$$

Taking one step of Newton's method without line search amounts to using a step size of 1, i.e. we take

$$x_1 = x_0 + p_0 = x_0 - x_0 + Q^{-1}b = Q^{-1}b,$$

which, as we have seen, is the global minimizer of f.

c) Taking unit steps (i.e. $\alpha = 1$), the condition of sufficient decrease in the back-tracking algorithm (Algorithm 3.1 in N&W) requires that

$$f(x_k + p_k) \le f(x_k) + c_1(\nabla f_k)^T p_k. \tag{1}$$

In our case, we have

$$f(x_k + p_k) = \frac{1}{2}(Q^{-1}b)^TQQ^{-1}b - b^TQ^{-1}b = \frac{1}{2}b^TQ^{-T}b - b^TQ^{-1}b = -\frac{1}{2}b^TQ^{-1}b,$$

since $Q^{-T} = Q^{-1}$ by symmetry. We also find, after some calculation, that

$$\begin{aligned} f(x_k) + c_1(\nabla f_k)^T p_k &= \frac{1}{2}x_k^TQx_k - b^Tx_k + c_1(Qx_k - b)^T(-x_k + Q^{-1}b) \\ &= \left(\frac{1}{2} - c_1\right)x_k^TQx_k - \left(\frac{1}{2} - c_1\right)2b^Tx_k - c_1b^TQ^{-1}b. \end{aligned}$$

Inserting the two quantities into (1), we find that the condition of sufficient decrease can be stated as

$$-\frac{1}{2}b^TQ^{-1}b \leq \left(\frac{1}{2}-c_1\right)x_k^TQx_k - \left(\frac{1}{2}-c_1\right)2b^Tx_k - c_1b^TQ^{-1}b,$$

which is equivalent to:

$$-\left(\frac{1}{2}-c_1\right)\left[x_k^TQx_k - 2b^Tx_k + b^TQ^{-1}b\right] \leq 0.$$

We now observe that the second factor is equal to $(\nabla f_k)^TQ^{-1}\nabla f_k$, and therefore that the condition of sufficient decrease is equivalent to requiring that

$$\left(c_1 - \frac{1}{2}\right)(\nabla f_k)^TQ^{-1}\nabla f_k \leq 0.$$

Since Q^{-1} is positive definite, meaning $(\nabla f_k)^TQ^{-1}\nabla f_k > 0$ unless the stationary point with $\nabla f_k = 0$ is reached, this condition will hold if and only if $c_1 \leq 1/2$.

2 **a)** We begin by finding

$$\begin{aligned} f(x,y) &= 2x^2 + y^2 - 2xy + 2x^3 + x^4 \\ \nabla f(x,y) &= \begin{bmatrix} 4x - 2y + 6x^2 + 4x^3 \\ 2y - 2x \end{bmatrix}, \\ \nabla^2 f(x,y) &= \begin{bmatrix} 4 + 12x + 12x^2 & -2 \\ -2 & 2 \end{bmatrix}. \end{aligned}$$

From this, we find the stationary points of f to be (0,0), ($-\frac{1}{2}$,$-\frac{1}{2}$) and (-1,-1). To characterize the points, we check the definiteness of the Hessian at each one by observing the eigenvalues of the Hessian. At both (0,0) and (-1,-1) we find the eigenvalues $\lambda = 3 \pm \sqrt{5} > 0$, meaning the Hessian is positive definite and the points are minimizers of f. At ($-\frac{1}{2}$,$-\frac{1}{2}$) we find the eigenvalues to be $\lambda = \frac{3\pm\sqrt{13}}{2}$. One of these is positive while the other is negative, so the Hessian is indefinite at this point, meaning ($-\frac{1}{2}$,$-\frac{1}{2}$) is a saddle point.

Next, we check the function values and see that $f(0,0) = f(-1,-1) = 0$. In addition, we can see that $f(x,y) \to \infty$ as $x^2 + y^2 \to \infty$, and thus we may conclude that (0,0) and (-1,-1) are both global minimizers.

b) The gradient descent method uses iterations of the form

$$\begin{aligned} x_{k+1} &= x_k + \alpha_k p_k \\ p_k &= -\nabla f(x_k). \end{aligned}$$

Starting at $x_0 = (-1,0)$, we find $f(x_0) = 1$, $\nabla f(x_0) = [-2,2]^T$ and $p_0 = [2,-2]^T$. With $c = 1/4$, we find that the acceptance criterion for step lengths becomes

$$\begin{aligned} & f(x_0 + \alpha_0 p_0) \leq f(x_0) + c\alpha_0(\nabla f(x_0))^T p_0 \\ \Rightarrow & f(x_0 + \alpha_0 p_0) \leq 1 - 2\alpha_0. \end{aligned}$$

With $\alpha_0 = 1$, we find $x_0 + p_0 = [1, -2]^T$ and $f(x_0 + p_0) = 13$. This exceeds the acceptance criterion $f(x_0 + p_0) \leq -1$, so we consider instead $\alpha_0 = 1/2$.

With $\alpha_0 = 1/2$, we find $x_0 + p_0/2 = [0, -1]^T$ and $f(x_0 + p_0/2) = 1$. This still exceeds the acceptance criterion $f(x_0 + p_0/2) \leq 0$, so we consider instead $\alpha_0 = 1/4$.

With $\alpha_0 = 1/4$, we find $x_0 + p_0/4 = [-1/2, -1/2]^T$ and $f(x_0 + p_0/4) = 1/16$. This satisfies the acceptance criterion $f(x_0 + p_0/4) \leq 1/2$, so we accept the step length and take $x_1 = (-1/2 - 1/2)$. From the previous exercise, we know this is a saddle point, and thus a stationary point, such that $\nabla f(x_1) = [0, 0]$, and thus the iterations will stop. However, it is not a minimum, so the iterations do not converge to a minimum.

c) Newton's method uses iterations of the form

$$x_{k+1} = x_k + \alpha_k p_k$$
$$p_k = -(\nabla^2 f(x_k))^{-1} \nabla f(x_k).$$

Starting at $x_0 = (-1, 0)$, we find $f(x_0) = 1, \nabla f(x_0) = [-2, 2]^T$ and $p_0 = [0, -1]^T$. With $c = 1/4$, we find that the acceptance criterion for step lengths becomes

$$f(x_0 + \alpha_0 p_0) \leq f(x_0) + c\alpha_0 (\nabla f(x_0))^T p_0$$
$$\Rightarrow f(x_0 + \alpha_0 p_0) \leq 1 - \alpha_0/2.$$

With $\alpha_0 = 1$, we find $x_0 + p_0 = [-1 - 1]^T$ and $f(x_0 + p_0) = 0$. This satisfies the acceptance criterion $f(x_0 + p_0) \leq 1/2$, so we accept the step length and take $x_1 = (-1 - 1)$. This is one of the global minimizers, and we see that the method converges in just one iteration, a lucky coincidence.

3 Example implementations are given in the MATLAB files `Newton.m` and `GradDesc.m`, which can be found on the course webpage.