Norwegian University of Science and Technology
Department of Mathematical Sciences

TMA4180 Optimization Theory Spring 2015

Exercise set 2 solution

$\boxed{1}$ 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

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

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

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

One may note that this implies that Newton's method converges in *one* step for any quadratic minimization problem where Q is SPD, since for any x_0 one has $x_1 = Q^{-1}b$, which is the unique minimizer of the problem. Taking unit steps (i.e. $\alpha = 1$), the condition of sufficient decrease requires that

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

In this case, we have

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

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

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

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

$$-\frac{1}{2}b^T Q^{-1}b \leq \left(\frac{1}{2} - c_1\right) x_k^T Q x_k - \left(\frac{1}{2} - c_1\right) 2b^T x_k - c_1 b^T Q^{-1} b,$$

which is equivalent to:

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

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

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

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

Next, we look at the curvature condition

$$(\nabla f(x_k + p_k))^T p_k \geq c_2 (\nabla f_k)^T p_k,$$

where we can see that

$$\nabla f(x_k + p_k) = \nabla f(Q^{-1} b) = b - b = 0,$$

and with $p_k = -Q^{-1} \nabla f_k$, we see that the condition requires

$$-c_2 (\nabla f_k)^T Q^{-1} \nabla f_k \leq 0,$$

which is satisfied for any $c_2 > 0$.

2 An example implementation is given in the MATLAB file `Newton.m`, which can be found on the course webpage.

3 **Exercise 3.7**

We follow the hints given in the exercise and find first that

$$\begin{aligned}
x_{k+1} &= x_k - \alpha_k \nabla f_k \\
\Rightarrow ||x_{k+1} - x^*||_Q^2 &= ||x_k - x^* - \alpha_k \nabla f_k||_Q^2 \\
&= ||x_k - x^*||_Q^2 - 2\alpha_k (\nabla f_k)^T Q (x_k - x^*) + \alpha_k^2 (\nabla f_k)^T Q \nabla f_k \\
\Rightarrow ||x_k - x^*||_Q^2 - ||x_{k+1} - x^*||_Q^2 &= 2\alpha_k (\nabla f_k)^T Q (x_k - x^*) - \alpha_k^2 (\nabla f_k)^T Q \nabla f_k
\end{aligned}$$

Next, we observe that

$$Q(x_k - x^*) = \nabla f_k \qquad \text{and} \qquad \alpha_k = \frac{(\nabla f_k)^T \nabla f_k}{(\nabla f_k)^T Q \nabla f_k}.$$

Applying this, we get

$$2\alpha_k (\nabla f_k)^T Q (x_k - x^*) - \alpha_k^2 (\nabla f_k)^T Q \nabla f_k = \frac{((\nabla f_k)^T \nabla f_k)^2}{(\nabla f_k)^T Q \nabla f_k}.$$

Finally, we observe that

$$||x_k - x^*||_Q^2 = (x_k - x^*)^T Q (x_k - x^*) = (Q^{-1} \nabla f_k)^T Q Q^{-1} \nabla f_k = (\nabla f_k)^T Q^{-T} \nabla f_k$$

and since Q is SPD, $Q^{-T} = Q^{-1}$ so we get $||x_k - x^*||_Q^2 = (\nabla f_k)^T Q^{-1} \nabla f_k$. Inserting all of this into the first equation, we get

$$\begin{aligned} ||x_{k+1} - x^*||_Q^2 &= ||x_k - x^*||_Q^2 - \frac{((\nabla f_k)^T \nabla f_k)^2}{(\nabla f_k)^T Q \nabla f_k} \\ &= \left[1 - \frac{((\nabla f_k)^T \nabla f_k)^2}{(\nabla f_k)^T Q \nabla f_k (\nabla f_k)^T Q^{-1} \nabla f_k}\right] ||x_k - x^*||_Q^2. \end{aligned}$$

Exercise 3.8 Since Q is SPD, we can diagonalize it, i.e.

$$Q = RDR^T, \quad Q^{-1} = RD^{-1}R^T,$$

where R is an orthonormal matrix and $D = \text{diag}\{\lambda_1, \lambda_2, ..., \lambda_n\}$. Each column of R is an eigenvector of Q and the $\lambda_i > 0$ are the eigenvalues of Q. We may assume without loss of generality that the eigenvalues are ordered, i.e. $\lambda_1 \leq \lambda_2 \leq ... \leq \lambda_n$. Since $RR^T = I$, we can write

$$\beta = \frac{(x^T x)^2}{(x^T Q x)(x^T Q^{-1} x)} = \frac{(x^T R R^T x)^2}{(x^T R D R x)(x^T R D^{-1} R^T x)} = \frac{(d^T d)^2}{(d^T D d)(d^T D^{-1} d)},$$

with $d = R^T x$. We now define $\xi_i = \frac{d_i^2}{d^T d}$. Then, $\xi_i > 0$ and $\sum_i \xi_i = \frac{1}{d^T d} \sum_i d_i^2 = 1$. Using this, we may observe that

$$\frac{d^T d}{d^T D d} = \frac{\sum_i d_i^2}{\sum d_i^2 \lambda_i} = \frac{d^T d \sum_i \xi_i}{d^T d \sum_i \xi_i \lambda_i} = \frac{1}{\sum_i \xi_i \lambda_i},$$

and similarly that

$$\frac{d^T d}{d D^{-1} d} = \frac{1}{\sum_i \frac{\xi_i}{\lambda_i}}.$$

Hence,

$$\beta = \frac{1}{(\sum_i \xi_i \lambda_i)(\sum_i \frac{\xi_i}{\lambda_i})}$$

Furthermore, define $\lambda = \sum_i \xi_i \lambda_i$ and $\bar{\lambda} = \sum_i \frac{\xi_i}{\lambda_i}$. Observe that $\lambda_1 \leq \lambda \leq \lambda_n$. Since the function $\phi(\lambda) = \frac{1}{\lambda}$ is convex, we know that

$$\frac{1}{\lambda_i} \leq \frac{\lambda_n - \lambda_i}{\lambda_n - \lambda_1} \frac{1}{\lambda_1} + \frac{\lambda_i - \lambda_1}{\lambda_n - \lambda_1} \frac{1}{\lambda_n} = \frac{\lambda_n + \lambda_1 - \lambda_i}{\lambda_1 \lambda_n}.$$

Hence,

$$\bar{\lambda} \leq \sum_i \left(\frac{\lambda_n + \lambda_1 - \lambda_i}{\lambda_1 \lambda_n}\right) \xi_i = \frac{\lambda_n + \lambda_1 - \lambda}{\lambda_1 \lambda_n}$$

Finally, we deduce that

$$\beta = \frac{1}{\lambda \bar{\lambda}} \geq \frac{\lambda_1 \lambda_n}{\lambda(\lambda_n + \lambda_1 - \lambda)} \geq \frac{\lambda_1 \lambda_n}{\max_{\alpha \in [\lambda_1, \lambda_n]} \{\alpha(\lambda_n + \lambda_1 - \alpha)\}} = \frac{4\lambda_1 \lambda_n}{(\lambda_1 + \lambda_n)^2},$$

which is the desired inequality. In the final equality, we have used that $\alpha(\lambda_n + \lambda_1 - \alpha)$ attains its maximum at $\alpha = \frac{\lambda_1 + \lambda_n}{2}$, which can easily be verified.

$\boxed{4}$ **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

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

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

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

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.