Solutions to Exercise Set 11

Problem 1

Steepest descent method with a step obtained by exact minimization

(a) The steepest descent direction of f in (x_0,y_0) is given by:

$$\mathbf{d} = -\begin{pmatrix} 6x_0 \\ 6y_0 \end{pmatrix} = \begin{pmatrix} -6 \\ -6 \end{pmatrix}$$

Computation of the step α_{min}:

$$\alpha_{min} = \underset{\alpha \geq 0}{\operatorname{argmin}}\, g(\alpha) = \underset{\alpha \geq 0}{\operatorname{argmin}}\, f((x_0\ y_0)^T + \alpha \mathbf{d})$$

We get that $g(\alpha) = f((1-6\alpha\ 1-6\alpha)^T) = 6(1-6\alpha)^2$. Moreover, as the function is strictly convex, the step is obtained by setting $g^{'}(\alpha) = 0$:

$$1 - 6\alpha = 0 \Rightarrow \alpha_{min} = \frac{1}{6}$$

The new iterate is:

$$(x_1\ y_1)^T = (x_0\ y_0)^T + \alpha_{min}\mathbf{d} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

We find the minimum of f in one iteration.

(b) From a theoretical point of view, there is no result that gives the number of iterations necessary to converge in the general case for this method.

Newton's Method

(a) The Newton's direction is given by:

$$\mathbf{d} = -\begin{pmatrix} 6 & 0 \\ 0 & 6 \end{pmatrix}^{-1} \begin{pmatrix} 6 \\ 6 \end{pmatrix} = \begin{pmatrix} -1 \\ -1 \end{pmatrix}$$

The new iterate is:

$$(x_1\ y_1)^T = (x_0\ y_0)^T + \mathbf{d} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

We find the minimum in one iteration.

(b) Newton's method converges in one iteration for strictly convex quadratic problems. Newton's direction is obtained by minimizing a quadratic function explaining why it converges in one iteration.

Conjuguate gradient method

(a) $\mathbf{Q}$ is a symmetric positive definite matrix given by:

$$\begin{pmatrix} 6 & 0 \\ 0 & 6 \end{pmatrix},$$

$\mathbf{b} = (0\ 0)^T$ and $c = 0$. Let's set $\mathbf{x} = (x_0\ y_0)^T$.

The direction is given by:

$$\mathbf{d} = -\mathbf{Q}\mathbf{x} - \mathbf{b} = \begin{pmatrix} -6 \\ -6 \end{pmatrix}$$

The step is:

$$\alpha = -\frac{\mathbf{d}^T(\mathbf{Q}\mathbf{x} + \mathbf{b})}{\mathbf{d}^T\mathbf{Q}\mathbf{d}} = \frac{1}{6}$$

The new iterate:

$$(x_1\ y_1)^T = \mathbf{x} + \alpha\mathbf{d} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

We find the minimum in one iteration.

(b) The maximal number of iterations for this method is given by the dimension of the problem, i.e. 2 in this example.

Problem 2

a) The gradient and the hessian of f are:

$$\nabla f(\mathbf{x}) = \begin{pmatrix} 4(x-2)^3 + 2(x-2)y^2 \\ 2(x-2)^2 y + 2(y+1) \end{pmatrix} \qquad \forall \mathbf{x} \in \mathbb{R}^2$$

$$\nabla^2 f(\mathbf{x}) = \begin{pmatrix} 12(x-2)^2 + 2y^2 & 4(x-2)y \\ 4(x-2)y & 2(x-2)^2 + 2 \end{pmatrix} \qquad \forall \mathbf{x} \in \mathbb{R}^2$$

b) The new iterate $\mathbf{x}_{k+1}$ is given by

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{s}_k$$

The step $\mathbf{s}_k$ is given by solving the following linear system at each iteration:

$$\nabla^2 f(\mathbf{x}_k)\mathbf{s}_k = -\nabla f(\mathbf{x}_k)$$

We get the following values:

k	$\mathbf{x}_k$	$\mathbf{s}_k$	$f(\mathbf{x}_k)$
0	$(1\ \ 1)^T$	$(0\ \ -1.5)^T$	6
1	$(1.0\ \ -0.5)^T$	$(0.3913\ \ -0.1956)^T$	1.5
2	$(1.3913\ \ -0.6956)^T$	$(0.3546\ \ -0.2532)^T$	$4.092 \cdot 10^{-1}$
3	$(1.7459\ \ -0.9491)^T$	$(0.2403\ \ -0.0991)^T$	$6.49 \cdot 10^{-2}$
4	$(1.9862\ \ -1.0482)^T$	–	$2.5 \cdot 10^{-3}$

Problem 3

We compute the gradient:

$$\nabla f(\mathbf{x}) = \begin{pmatrix} 2x_1 + 2x_2 \\ 2x_1 + 4x_2 \end{pmatrix},$$

the hessian:

$$H(\mathbf{x}) = \begin{pmatrix} 2 & 2 \\ 2 & 4 \end{pmatrix}$$

and its inverse:

$$H^{-1}(\mathbf{x}) = \begin{pmatrix} 1 & -1/2 \\ -1/2 & 1/2 \end{pmatrix}.$$

We note that

$$H^{-1}(\mathbf{x})\nabla f(\mathbf{x}) = \begin{pmatrix} 1 & -1/2 \\ -1/2 & 1/2 \end{pmatrix} \begin{pmatrix} 2x_1 + 2x_2 \\ 2x_1 + 4x_2 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \mathbf{x}$$

For any $\mathbf{x}^{(0)} \in \mathbb{R}^2$, we have that:

$$\mathbf{x}^{(1)} = \mathbf{x}^{(0)} - H^{-1}(\mathbf{x}^{(0)})\nabla f(\mathbf{x}^{(0)}) = \mathbf{x}^{(0)} - \mathbf{x}^{(0)} = 0$$

Consequently, the Newton's method converges in one iteration independently from the starting point. This result is true for any positive definite quadratic function.

Problem 4

1. $\nabla f(x,y) = \begin{pmatrix} 4x^3 - 4x \\ 3y^2 - 3 \end{pmatrix}$.
   $\nabla f^2(x,y) = \begin{pmatrix} 12x^2 - 4 & 0 \\ 0 & 6y \end{pmatrix}$.
   (2,2) is not minimun since $\nabla f(2,2) = \begin{pmatrix} 24 \\ 9 \end{pmatrix}$.
   $(-1,1)$ is a minimum since $\nabla f(-1,1) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ and $\nabla f^2(-1,1) = \begin{pmatrix} 8 & 0 \\ 0 & 6 \end{pmatrix}$ is positive definite.
   $(0,-1)$ is a maximum since $\nabla f(0,-1) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ and $\nabla f^2(0,-1) = \begin{pmatrix} -4 & 0 \\ 0 & -6 \end{pmatrix}$ is negative definite.
2. $x_0 = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$
   $x_1 = x_0 - (\nabla^2 f(2,2))^{-1}\nabla f(2,2) = \begin{pmatrix} 2 \\ 2 \end{pmatrix} - \begin{pmatrix} \frac{1}{44} & 0 \\ 0 & \frac{1}{12} \end{pmatrix} \begin{pmatrix} 24 \\ 9 \end{pmatrix} = \begin{pmatrix} \frac{16}{11} \\ \frac{5}{4} \end{pmatrix}$.
   The method is not applicable to the other points since $\nabla f(x,y)$ is null at these points.
3. The Armijo rule: $f\begin{pmatrix} \frac{16}{11} \\ \frac{5}{4} \end{pmatrix} \simeq -1.552 \leq f\begin{pmatrix} 2 \\ 2 \end{pmatrix} - 0.1(24 \;\; 9)\begin{pmatrix} \frac{1}{44} & 0 \\ 0 & \frac{1}{12} \end{pmatrix}\begin{pmatrix} 24 \\ 9 \end{pmatrix} \simeq 8.016$
   is satisfied.

Problem 5

(a) the hessian of f is given by $\mathbf{Q}$ for all $\mathbf{x} \in \mathbb{R}^3$. Q is positive definite and f is strictly convex over $\mathbb{R}^3$. The unique minimum of f over $\mathbb{R}^3$ is given by the unique solution of the system of equations $\mathbf{Qx} = -\mathbf{b}$ that can be rewritten as:

$$\begin{cases} x_1 = -1 \\ 5x_2 = -1 \\ 25x_3 = -1 \end{cases}$$

The solution is given by:

$$\begin{cases} x_1^* = -1 \\ x_2^* = -\frac{1}{5} \\ x_3^* = -\frac{1}{25} \end{cases}$$

This is the unique minimum.

(b) The gradient of f is given by:

$$\nabla f(\mathbf{x}) = \begin{pmatrix} x_1 + 1 \\ 5x_2 + 1 \\ 25x_3 + 1 \end{pmatrix} \qquad \forall x \in \mathbb{R}^3$$

The steepest descent direction for f at $\mathbf{x}^0$ is:

$$\mathbf{d}^0 = -\nabla f(\mathbf{x}^0) = \begin{pmatrix} -1 \\ -1 \\ -1 \end{pmatrix}$$

Problem 6

(a) The gradient of f is given by:

$$\nabla f(x,y) = \begin{pmatrix} 2x^3 - 2xy + x - 1 \\ -x^2 + y \end{pmatrix} \; \forall (x,y) \in \mathbb{R}^2$$

Candidates to be a local minimum are given by the solutions of the system $\nabla f(x,y) = 0$:

$$\begin{cases} 2x^3 - 2xy + x - 1 = 0 \\ -x^2 + y = 0 \end{cases}$$

We get $x^2 = y$ from the second equation and if we replace y by x^2 in the first equation, we immediately obtain that $x = 1$. So the solution is given by:

$$\begin{cases} x = 1 \\ y = 1 \end{cases}$$

Let's have a look at the hessian matrix at (1,1):

$$\nabla^2 f(x,y) = \begin{pmatrix} 6x^2 - 2y + 1 & -2x \\ -2x & 1 \end{pmatrix}$$

and we get:

$$\nabla^2 f(1,1) = \begin{pmatrix} 5 & -2 \\ -2 & 1 \end{pmatrix}.$$

This matrix is positive definite. Indeed $(x \;\; y)\nabla^2 f(1,1)(x \;\; y)^T = 5x^2 - 4xy + y^2 = (2x - y)^2 + x^2 \geq 0$ with equality only when $x = y = 0$. So we conclude that this is the unique minimizer of f over $\mathbb{R}^2$.

(b) At $(x^0,y^0) = (2,2)$, we have that:

$\nabla f(x^0,y^0) = \begin{pmatrix} 9 \\ -2 \end{pmatrix}$ and $\nabla^2 f(x^0,y^0) = \begin{pmatrix} 21 & -4 \\ -4 & 1 \end{pmatrix}$

Newton's direction at (x^0,y^0) is:

$$\mathbf{d}^0 = -[\nabla^2 f(x^0,y^0)]^{-1} \nabla f(x^0,y^0) = -\begin{pmatrix} 0.2 & 0.8 \\ 0.8 & 4.2 \end{pmatrix} \begin{pmatrix} 9 \\ -2 \end{pmatrix} = \begin{pmatrix} -0.2 \\ 1.2 \end{pmatrix}$$

Consequently:

$$\begin{pmatrix} x^1 \\ y^1 \end{pmatrix} = \begin{pmatrix} x^0 \\ y^0 \end{pmatrix} + \mathbf{d}^0 = \begin{pmatrix} 1.8 \\ 3.2 \end{pmatrix}$$

The new iterate is not closer to the optimal value but the objective function has decresed in value of approximately 2.18. So the new iterate is a better point even though it is not closer to the optimal solution !

Newton's method needs 5 iterations to converge to (1,1). The different iterates are given below:

x^k	y^k
2	2
1.8	3.2
1.0593	0.5733
1.0310	1.0622
1	0.9991
1	1