Adrian Willi

Exercise 9

Task 1

Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$, defined by $f(x, y) = (x^2 - 2xy + x)^2$. We would like to determine a point at which the function f takes on its minimum value. Starting from $x^0 = (x_0, y_0) = (2, 2)$, calculate the next iteration point $x^1 = (x_1, y_1)$ according to each of the following methods:

a) Gradient method with successive halving of the step size;

b) Gradient method with successive halving and parabola fitting;

c) Newton's method;

d) Broyden's method. Compute also the second iteration point $x^2 = (x_2, y_2)$.

Task 2

Calculate the first two terms we get by applying Aitken's acceleratioen method to the zero convergent sequence starting with $100, 10, 2, \frac{1}{2}, \ldots$

Task 3

The function $f : \mathbb{R}^2 \to \mathbb{R}$ defined by

$$f(x, y) = x^4 + y^4$$

clearly attains its minimum at $(0, 0)$. In this task we investigate the behaviour of Newton's method when determining this minimum, starting from an arbitrary point $(x_0, y_0) \neq (0, 0)$.

a) Determine the first iteration point (x_1, y_1) as a function of the starting point (x_0, y_0).

b) Give a general formula for the n-th iteration point (x_n, y_n).

c) What is the convergence speed in this example? Are you surprised?

d) Apply Aitken's acceleration method to the sequence found in b).(Use the starting point $(x_0, y_0) = (1, 1)$ for simplicity. What do you find? Can you explain your result?

Task 1 $\quad f(x,y) = (x^2 - 2xy + x)^2$

$$\frac{\partial f}{\partial x} = 2(x^2 - 2xy + x)(2x - 2y + 1)$$

$$\frac{\partial f}{\partial y} = 2(x^2 - 2xy + x)(-2x)$$

$$\frac{\partial f}{\partial x^2} = (4x^2 - 8xy + 4x) + (4x - 4y + 2)(2x - 2y + 1) \qquad H_f(2,2) = \begin{bmatrix} -6 & 0 \\ 0 & 32 \end{bmatrix}$$

$$\frac{\partial f}{\partial y \partial x} = \frac{\partial f}{\partial x \partial y} = -12x^2 - 8x + 16xy$$

$$\frac{\partial f}{\partial y^2} = 8x^2$$

$$\nabla f(2,2) = \begin{bmatrix} 2(x^2 - 2xy + x)(2x - 2y + 1) \\ 2(x^2 - 2xy + x)(-2x) \end{bmatrix} = \begin{bmatrix} -4 \\ 16 \end{bmatrix}$$

Starting at point $x^0 = (x_0, y_0) = (2,2)$ with $f(2,2) = 4$

a)

$\binom{2}{2} - \beta \binom{-4}{16}$	$f(x_1, y_1)$
$\binom{2}{2} - 1 \binom{-4}{16} = \binom{6}{-14}$	44'100
$\binom{2}{2} - \frac{1}{2} \binom{-4}{16} = \binom{4}{-6}$	41'624
$\binom{2}{2} - \frac{1}{4} \binom{-4}{16} = \binom{3}{-2}$	576
$\binom{2}{2} - \frac{1}{8} \binom{-4}{16} = \binom{2.5}{0}$	76.56
$\binom{2}{2} - \frac{1}{16} \binom{-4}{16} = \binom{2.25}{1}$	7.91
$\binom{2}{2} - \frac{1}{32} \binom{-4}{16} = \binom{2.125}{1.5}$	$0.07056 < 4$ done!

The next iteration step is at $x^1 = (x_1, y_1) = (2.125, 1.5)$

b)

	t	$P(t) \overset{!}{=} f\left(\binom{2}{2} - t\binom{-4}{16}\right)$
$P(0)$	0	$P(0) \overset{!}{=} f\left(\binom{2}{2} - 0 \binom{-4}{16}\right) = f\binom{2}{2} = 4$
$P(\beta)$	$\frac{1}{32}$	$P(\frac{1}{32}) \overset{!}{=} f\left(\binom{2}{2} - \frac{1}{32}\binom{-4}{16}\right) = f\binom{2.125}{1.5} = 0.07056$
$P(2\beta)$	$\frac{1}{16}$	$P(\frac{1}{16}) \overset{!}{=} f\left(\binom{2}{2} - \frac{1}{16}\binom{-4}{16}\right) = f\binom{2.25}{1} = 7.91$

$$\beta^* = -\frac{b}{2a} = \frac{1}{38}$$

Therefore, we consider

$$\binom{x_1}{y_1} = \binom{2}{2} - \frac{1}{38}\binom{-4}{16} = \binom{2.10526}{1.583607}$$

c) $x^1 = \binom{2}{2} - \left(H_f(2,2)\right)^{-1} \cdot \nabla f(2,2)$

$$= \binom{2}{2} - \begin{bmatrix} -\frac{1}{6} & 0 \\ 0 & \frac{1}{32} \end{bmatrix} \cdot \begin{bmatrix} -4 \\ 16 \end{bmatrix} = \begin{bmatrix} 4/3 \\ 3/2 \end{bmatrix}$$

$$H_f(2,2) = \begin{bmatrix} -6 & 0 \\ 0 & 32 \end{bmatrix}$$

$$\left(H_f(2,2)\right)^{-1} = \begin{bmatrix} -\frac{1}{6} & 0 \\ 0 & \frac{1}{32} \end{bmatrix}$$

Therefore, the next iteration point x^1 in Newton's method is given by $x^1 = (x_1, y_1) = (1.\overline{3}, 1.5)$

d) The first iteration step is the same as in Newton's method

First step: $x^1 = x^0 - (A^0)^{-1} \nabla f(x^0)$

$$\binom{x_0}{y_0} = \binom{2}{2}$$

$$\nabla f(x_0, y_0) = \binom{-4}{16}$$

$$(A^0)^{-1} = \left(H_f(x_0, y_0)\right)^{-1} = \begin{pmatrix} -\frac{1}{6} & 0 \\ 0 & \frac{1}{32} \end{pmatrix}$$

$$\binom{x_1}{y_1} = \binom{2}{2} - \begin{pmatrix} -1/6 & 0 \\ 0 & 1/32 \end{pmatrix} \binom{-4}{16} = \binom{1.\overline{3}}{1.5}$$

Second step:

$$\nabla f(x_1, y_1) = \begin{pmatrix} -1.18519 \\ 4.740741 \end{pmatrix}$$

$$g^1 = \nabla f(x_1, y_1) - \nabla f(x_0, y_0) = \begin{pmatrix} -1.18519 \\ 4.740741 \end{pmatrix} - \binom{-4}{16} = \begin{pmatrix} 2.814815 \\ -11.25926 \end{pmatrix}$$

$$d^1 = x^1 - x^0 = \binom{1.\overline{3}}{1.5} - \binom{2}{2} = \binom{-2/3}{-1/2}$$

$$\begin{bmatrix} 0.157531 \\ 0.148148 \end{bmatrix} \begin{bmatrix} \frac{1}{9} & \frac{-1}{64} \end{bmatrix}$$

$$(A^1)^{-1} = \overset{a}{(A^0)^{-1}} - \frac{\left(\overset{a}{(A^0)^{-1}} \overset{g}{g^1} - \overset{d}{d^1}\right) \overset{d}{(d^1)^T} \overset{a}{(A^0)^{-1}}}{\underset{d}{(d^1)^T} \underset{a}{(A^0)^{-1}} \underset{g}{g^1}}$$

$$= \begin{pmatrix} -\frac{1}{6} & 0 \\ 0 & \frac{1}{32} \end{pmatrix} - \frac{\left(\begin{pmatrix} -\frac{1}{6} & 0 \\ 0 & \frac{1}{32} \end{pmatrix} \begin{pmatrix} 2.814815 \\ -11.25926 \end{pmatrix} - \binom{-2/3}{-1/2}\right) \binom{-2/3}{-1/2}^T \begin{pmatrix} -\frac{1}{6} & 0 \\ 0 & \frac{1}{32} \end{pmatrix}}{\binom{-2/3}{-1/2}^T \begin{pmatrix} -\frac{1}{6} & 0 \\ 0 & \frac{1}{32} \end{pmatrix} \begin{pmatrix} 2.814815 \\ -11.25926 \end{pmatrix}}$$

$$= \begin{pmatrix} -0.211579 & 0.006316 \\ -0.033684 & 0.035087 \end{pmatrix}$$

We obtain

$$\binom{x_2}{y_2} = \binom{x_1}{y_1} - (A^1)^{-1} \nabla f(x_1, y_1)$$

$$= \binom{4/3}{3/2} - \begin{pmatrix} -0.211579 & 0.006316 \\ -0.033684 & 0.035087 \end{pmatrix} \begin{pmatrix} -1.18519 \\ 4.740741 \end{pmatrix} = \binom{1.053}{1.289}$$

i.e. the next iteration point for Broyden's method is $x^2 = (x_2, y_2) = (1.053, 1.289)$

Task 2

$100, 10, 2, \frac{1}{2}, \ldots$

$$y^i = x^i - \frac{(x^i - x^{i-1})^2}{x^i - 2 \cdot x^{i-1} + x^{i-2}}$$

i	x^i	Aitken y^i
0	100	–
1	10	–
2	2	$1.21951 = y^2$
3	$1/2$	$0.15384 = y^3$

$$2 - \frac{(2-10)^2}{2 - 2\cdot 10 + 100}$$

Task 3

$$f(x,y) = x^4 + y^4$$

a)

$$\nabla f(x,y) = \begin{bmatrix} 4x^3 \\ 4y^3 \end{bmatrix} \qquad H_f(x,y) = \begin{bmatrix} 12x^2 & 0 \\ 0 & 12y^2 \end{bmatrix}$$

$$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} - \begin{pmatrix} 12(x_0)^2 & 0 \\ 0 & 12(y_0)^2 \end{pmatrix}^{-1} \cdot \begin{pmatrix} 4(x_0)^3 \\ 4(y_0)^3 \end{pmatrix} = \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} - \begin{pmatrix} x_0/3 \\ y_0/3 \end{pmatrix} = \frac{2}{3}\begin{pmatrix} x_0 \\ y_0 \end{pmatrix}$$

b) $$\begin{pmatrix} x_n \\ y_n \end{pmatrix} = \left(\frac{2}{3}\right)^n \begin{pmatrix} x_0 \\ y_0 \end{pmatrix}$$

c) The convergence is only linear, despite the fact that Newtons method "usually" converges quadratically. Does not apply here because gradients and hessian vanishes.

d) $$\Delta x_i = \left(\frac{2}{3}\right)^i - \left(\frac{2}{3}\right)^{i-1} = -\frac{1}{2}\left(\frac{2}{3}\right)^i$$

and therefore

$$\Delta^2 x_i = -\frac{1}{2}\left(\frac{2}{3}\right)^i - \left(-\frac{1}{2}\left(\frac{2}{3}\right)^{i-1}\right) = \frac{1}{4}\left(\frac{2}{3}\right)^i$$

$$\rightarrow \left(\frac{2}{3}\right)^i - \frac{\left(-\frac{1}{2}\left(\frac{2}{3}\right)^i\right)^2}{\frac{1}{4}\left(\frac{2}{3}\right)^i} = 0$$

Task 4

For $f : \mathbb{R}^2 \to \mathbb{R}$ defined by

$$f(x, y) = x^4 + y^4$$

as in Task 3, perform two steps of Broyden's method starting from $(x_0, y_0) = (1, 1)$. Determine the matrix $(A^1)^{-1}$ used as an approximation for the inverse of the Hessian matrix in the second iteration step, and compare it to the exact inverse of the Hessian matrix that would be applicable.

Task 5 (*optional)

Study the three implementations/variants of Newton's method available from the homepage (exact derivatives, approximate derivatives, Broyden's method), and try them out on the Bazarra-Shetty function and other examples. Implementations are available in R.

It might also be a good idea to implement some of these methods on your programmable calculator if you have one.

Task 6 (*optional)

Study the implementation of Aitken's acceleration method available from the homepage, and use it to improve the convergence of the programs from Task 5 and from Task 4 of Exercise 8.

Task 4

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

$$\nabla f(x,y) = \begin{bmatrix} 4x^3 \\ 4y^3 \end{bmatrix} \qquad H_f(x,y) = \begin{bmatrix} 12x^2 & 0 \\ 0 & 12y^2 \end{bmatrix}$$

Start at $(x_0, y_0) = (1,1)$

$$x_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} - \begin{bmatrix} 1/12 & 0 \\ 0 & 1/12 \end{bmatrix} \cdot \begin{bmatrix} 4 \\ 4 \end{bmatrix} = \begin{pmatrix} 2/3 \\ 2/3 \end{pmatrix} = \begin{pmatrix} x_1 \\ y_1 \end{pmatrix}$$

The inverse of the hessian at this is $\left(H_f\left(\frac{2}{3}, \frac{2}{3}\right)\right)^{-1} = \begin{bmatrix} 3/16 & 0 \\ 0 & 3/16 \end{bmatrix}$

Now we calculate the next step using Broyden's method

$$d^1 = x^1 - x^0 = \begin{pmatrix} 2/3 \\ 2/3 \end{pmatrix} - \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} -1/3 \\ -1/3 \end{pmatrix}$$

$$\nabla f(x_1, y_1) = \begin{pmatrix} 32/27 \\ 32/27 \end{pmatrix}$$

$$g^1 = \nabla f(x_1, y_1) - \nabla f(x_0, y_0) = \begin{pmatrix} 32/27 \\ 32/27 \end{pmatrix} - \begin{pmatrix} 4 \\ 4 \end{pmatrix} = \begin{pmatrix} -76/27 \\ -76/27 \end{pmatrix}$$

Our approximation for the inverted hessian matrix is therefore

$$(A^1)^{-1} = (A^0)^{-1} - \frac{((A^0)^{-1} g^1 - d^1)(d^1)^T (A^0)^{-1}}{(d^1)^T (A^0)^{-1} g^1}$$

$$= \begin{pmatrix} 1/12 & 0 \\ 0 & 1/12 \end{pmatrix} - \frac{\left(\begin{pmatrix} 1/12 & 0 \\ 0 & 1/12 \end{pmatrix}\begin{pmatrix} -76/27 \\ -76/27 \end{pmatrix} - \begin{pmatrix} -1/3 \\ -1/3 \end{pmatrix}\right)\left(-1/3, -1/3\right)\begin{pmatrix} 1/12 & 0 \\ 0 & 1/12 \end{pmatrix}}{\left(-1/3, -1/3\right)\begin{pmatrix} 1/12 & 0 \\ 0 & 1/12 \end{pmatrix}\begin{pmatrix} -76/27 \\ -76/27 \end{pmatrix}}$$

$$= \begin{pmatrix} 1/12 & 0 \\ 0 & 1/12 \end{pmatrix} - \frac{\begin{pmatrix} \frac{-76}{(27 \cdot 12)} + \frac{1}{3} \\ \frac{-76}{(27 \cdot 12)} + \frac{1}{3} \end{pmatrix}\left(-\frac{1}{3}, -\frac{1}{3}\right)\begin{pmatrix} 1/12 & 0 \\ 0 & 1/12 \end{pmatrix}}{\frac{38}{243}}$$

$$= \begin{pmatrix} 1/12 & 0 \\ 0 & 1/12 \end{pmatrix} - \frac{\begin{pmatrix} 8/81 \\ 8/81 \end{pmatrix}\left(-\frac{1}{3}, -\frac{1}{3}\right)\begin{pmatrix} 1/12 & 0 \\ 0 & 1/12 \end{pmatrix}}{\frac{38}{243}}$$

$$= \begin{pmatrix} 0.10087719 & 0.01754386 \\ 0.01754386 & 0.10087719 \end{pmatrix}$$

$$\begin{pmatrix} x_2 \\ y_2 \end{pmatrix} = \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} - (A^1)^{-1} \cdot \nabla f(x_1, y_1) = \begin{pmatrix} 0.5263 \\ 0.5263 \end{pmatrix}$$

$$f(x_2, y_2) = 0.153467$$

Solutions to Exercise 9

Solution to Task 1

As a preliminary, calculate the partial derivatives of the function f:

$$\begin{aligned}
&f(x,y) = (x^2-2xy+x)^2\\
&f_x(x,y) = 2(x^2-2xy+x)(2x-2y+1)\\
&f_y(x,y) = 2(x^2-2xy+x)(-2x)\\
&f_{xx}(x,y) = 2(2x-2y+1)(2x-2y+1)+2(x^2-2xy+x)2\\
&f_{xy}(x,y) = -12x^2-8x+16xy\\
&f_{yx}(x,y) = -12x^2-8x+16xy\\
&f_{yy}(x,y) = 8x^2
\end{aligned}$$

Thus the gradient at the point (x,y) is

$$\nabla f(x,y) = \begin{pmatrix} f_x(x,y)\\ f_y(x,y)\end{pmatrix} = \begin{pmatrix} 2(x^2-2xy+x)(2x-2y+1)\\ -4x(x^2-2xy+x)\end{pmatrix} ,$$

and the Hessian matrix at the point (x,y) is

$$H_f(x,y) = \begin{pmatrix} f_{xx}(x,y) & f_{xy}(x,y)\\ f_{yx}(x,y) & f_{yy}(x,y)\end{pmatrix} = \begin{pmatrix} 2(2x-2y+1)^2+4(x^2-2xy+x) & -12x^2-8x+16xy\\ -12x^2-8x+16xy & 8x^2\end{pmatrix}.$$

a) At the point $(x_0,y_0)=(2,2)$, we have

$$\begin{aligned}
&f(2,2) = 4,\\
&\nabla f(2,2) = \begin{pmatrix}-4\\16\end{pmatrix} .
\end{aligned}$$

While doing the successive halving, we encounter the following values:

β	$\begin{pmatrix}x_1\\y_1\end{pmatrix} = \begin{pmatrix}x_0\\y_0\end{pmatrix} - \beta\cdot\nabla f(x_0,y_0) = \begin{pmatrix}2\\2\end{pmatrix} - \beta\begin{pmatrix}-4\\16\end{pmatrix}$	$f(x_1,y_1)$
1	$\begin{pmatrix}x_1\\y_1\end{pmatrix} = \begin{pmatrix}2\\2\end{pmatrix} - 1\begin{pmatrix}-4\\16\end{pmatrix} = \begin{pmatrix}6\\-14\end{pmatrix}$	44100
$\frac{1}{2}$	$\begin{pmatrix}x_1\\y_1\end{pmatrix} = \begin{pmatrix}2\\2\end{pmatrix} - \frac{1}{2}\begin{pmatrix}-4\\16\end{pmatrix} = \begin{pmatrix}4\\-6\end{pmatrix}$	4624
$\frac{1}{4}$	$\begin{pmatrix}x_1\\y_1\end{pmatrix} = \begin{pmatrix}2\\2\end{pmatrix} - \frac{1}{4}\begin{pmatrix}-4\\16\end{pmatrix} = \begin{pmatrix}3\\-2\end{pmatrix}$	576
$\frac{1}{8}$	$\begin{pmatrix}x_1\\y_1\end{pmatrix} = \begin{pmatrix}2\\2\end{pmatrix} - \frac{1}{8}\begin{pmatrix}-4\\16\end{pmatrix} = \begin{pmatrix}2.5\\0\end{pmatrix}$	76.5625
$\frac{1}{16}$	$\begin{pmatrix}x_1\\y_1\end{pmatrix} = \begin{pmatrix}2\\2\end{pmatrix} - \frac{1}{16}\begin{pmatrix}-4\\16\end{pmatrix} = \begin{pmatrix}2.25\\1\end{pmatrix}$	7.910156
$\frac{1}{32}$	$\begin{pmatrix}x_1\\y_1\end{pmatrix} = \begin{pmatrix}2\\2\end{pmatrix} - \frac{1}{32}\begin{pmatrix}-4\\16\end{pmatrix} = \begin{pmatrix}2.125\\1.5\end{pmatrix}$	0.07056 < 4 done!

The next iteration point therefore is $x^1 = (x_1, y_1) = (2.125, 1.5)$ (when only using successive halving).

b) Using the results from a), we first fit a parabola $P(t) = at^2 + bt + c$ through the following three sample points. (Recall from a) that after the successive halving phase we have $\beta = 1/32$.)

t	$P(t) \overset{!}{=} f\left(\begin{pmatrix}2\\2\end{pmatrix} - t\begin{pmatrix}-4\\16\end{pmatrix}\right)$
0	$P(0) \overset{!}{=} f\left(\begin{pmatrix}2\\2\end{pmatrix} - 0\begin{pmatrix}-4\\16\end{pmatrix}\right) = f\begin{pmatrix}2\\2\end{pmatrix} = 4$
$\frac{1}{32}$	$P(\frac{1}{32}) \overset{!}{=} f\left(\begin{pmatrix}2\\2\end{pmatrix} - \frac{1}{32}\begin{pmatrix}-4\\16\end{pmatrix}\right) = f\begin{pmatrix}2.125\\1.5\end{pmatrix} = 0.07056$
$\frac{1}{16}$	$P(\frac{1}{16}) \overset{!}{=} f\left(\begin{pmatrix}2\\2\end{pmatrix} - \frac{1}{16}\begin{pmatrix}-4\\16\end{pmatrix}\right) = f\begin{pmatrix}2.25\\1\end{pmatrix} = 7.910156$

The vertex of the parabola is therefore at $\beta^* = -\frac{b}{2a} = 0.02602457$.
Therefore, we consider

$$\begin{pmatrix}x_1\\y_1\end{pmatrix} = \begin{pmatrix}2\\2\end{pmatrix} - 0.02602457\begin{pmatrix}-4\\16\end{pmatrix} = \begin{pmatrix}2.104098\\1.583607\end{pmatrix}$$

as our next iteration point. But before actually choosing it, we need to check the resulting function value:
For $\beta = 1/32 = 0.03125$ we got the function value $f(2.125, 1.5) = 0.07056$ (see a)), for $\beta^* = 0.02602457$ we get the function value $f(2.104098, 1.583607) = 0.017636$, which is better. The next iteration point therefore is $x^1 = (x_1, y_1) = (2.104098, 1.583607)$.

c) We calculate the gradient and the Hessian matrix at the iteration point $(x_0, y_0) = (2, 2)$:

$$\nabla f(x_0, y_0) = \begin{pmatrix}-4\\16\end{pmatrix}$$

$$H_f(x_0, y_0) = \begin{pmatrix}2(4-4+1)^2 + 4(4-8+2) & -48-16+64\\ -48-16+64 & 32\end{pmatrix} = \begin{pmatrix}-6 & 0\\ 0 & 32\end{pmatrix}$$

We obtain

$$\begin{aligned}
\begin{pmatrix}x_1\\y_1\end{pmatrix} &= \begin{pmatrix}x_0\\y_0\end{pmatrix} - \Big(H_f(x_0, y_0)\Big)^{-1}\nabla f(x_0, y_0)\\
&= \begin{pmatrix}2\\2\end{pmatrix} - \begin{pmatrix}-6 & 0\\ 0 & 32\end{pmatrix}^{-1}\begin{pmatrix}-4\\16\end{pmatrix}\\
&= \begin{pmatrix}2\\2\end{pmatrix} - \begin{pmatrix}-\frac{1}{6} & 0\\ 0 & \frac{1}{32}\end{pmatrix}\begin{pmatrix}-4\\16\end{pmatrix}\\
&= \begin{pmatrix}2\\2\end{pmatrix} - \begin{pmatrix}\frac{4}{6}\\ \frac{16}{32}\end{pmatrix}\\
&= \begin{pmatrix}\frac{4}{3}\\ \frac{3}{2}\end{pmatrix}.
\end{aligned}$$

Therefore the iteration point x^1 in Newton's method is given by $x^1 = (x_1, y_1) = (1.\bar{3}, 1.5)$.

d) The first iteration step (and hence also the iteration point x^1) in Broyden's method is the same as in Newton's method. We therefore use the results from c):

$$\begin{aligned}
\begin{pmatrix} x_0 \\ y_0 \end{pmatrix} &= \begin{pmatrix} 2 \\ 2 \end{pmatrix} \\
\nabla f(x_0, y_0) &= \begin{pmatrix} -4 \\ 16 \end{pmatrix} \\
\left(A^0\right)^{-1} &= \left(H_f(x_0, y_0)\right)^{-1} = \begin{pmatrix} -\frac{1}{6} & 0 \\ 0 & \frac{1}{32} \end{pmatrix} \\
\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} &= \begin{pmatrix} 1.\bar{3} \\ 1.5 \end{pmatrix}
\end{aligned}$$

We now calculate the gradient and an approximation $\left(A^1\right)^{-1}$ for the inverse of the Hessian Matrix at the iteration point $x^1 = (x_1, y_1) = (1.\bar{3}, 1.5)$:

$$\begin{aligned}
\nabla f(x_1, y_1) &= \begin{pmatrix} -1.185185 \\ 4.740741 \end{pmatrix} \\
d^1 = x^1 - x^0 &= \begin{pmatrix} -\frac{2}{3} \\ -\frac{1}{2} \end{pmatrix} \\
g^1 = \nabla f(x_1, y_1) - \nabla f(x_0, y_0) &= \begin{pmatrix} 2.814815 \\ -11.25926 \end{pmatrix} \\
\left(A^1\right)^{-1} = \left(A^0\right)^{-1} - \frac{\left((A^0)^{-1} g^1 - d^1\right)(d^1)^T (A^0)^{-1}}{(d^1)^T (A^0)^{-1} g^1} &= \begin{pmatrix} -0.211579 & 0.006316 \\ -0.033684 & 0.035987 \end{pmatrix}
\end{aligned}$$

We obtain

$$\begin{aligned}
\begin{pmatrix} x_2 \\ y_2 \end{pmatrix} &= \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} - \left(A^1\right)^{-1} \nabla f(x_1, y_1) \\
&= \begin{pmatrix} \frac{4}{3} \\ \frac{3}{2} \end{pmatrix} - \begin{pmatrix} -0.211579 & 0.006316 \\ -0.033684 & 0.035987 \end{pmatrix} \begin{pmatrix} -1.185185 \\ 4.740741 \end{pmatrix} \\
&= \begin{pmatrix} 1.053 \\ 1.289 \end{pmatrix},
\end{aligned}$$

i.e. the next iteration point in Broyden's method is $x^2 = (x_2, y_2) = (1.053, 1.289)$.

For comparison: The exact inverse Hessian at $(x_1, y_1) = (1.\bar{3}, 1.5)$ is $(H_f(x_1, y_1))^{-1} = \begin{pmatrix} -0.375 & 0 \\ 0 & 0.0703125 \end{pmatrix}$, so the approximation by $(A^1)^{-1}$ is not very good here.

Solution to Task 2

$y^2 = 1.2195122$ and $y^3 = 0.1538462$.

Solution to Task 3

a) $\nabla f(x,y) = \begin{pmatrix} 4x^3 \\ 4y^3 \end{pmatrix}$

$H_f(x,y) = \begin{pmatrix} 12x^2 & 0 \\ 0 & 12y^2 \end{pmatrix}$

$$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} - \begin{pmatrix} 12(x_0)^2 & 0 \\ 0 & 12(y_0)^2 \end{pmatrix}^{-1} \cdot \begin{pmatrix} 4(x_0)^3 \\ 4(y_0)^3 \end{pmatrix} = \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} - \begin{pmatrix} x_0/3 \\ y_0/3 \end{pmatrix} = \tfrac{2}{3} \begin{pmatrix} x_0 \\ y_0 \end{pmatrix}$$

b) $\begin{pmatrix} x_n \\ y_n \end{pmatrix} = \left(\tfrac{2}{3}\right)^n \begin{pmatrix} x_0 \\ y_0 \end{pmatrix}$

c) The convergence speed is only linear, despite the fact that Newton's method „usually" converges quadratically. The general statements about convergence speeds discussed in the lecture do *not* apply in this example because at $(0,0)$ not only the gradient (=first derivative), but also the Hessian matrix (=second derivative) vanishes. Note that this happens in particular when trying to find a minimum that is a multiple root of a (univariate) polynomial with Newton's method, like e.g. the unique minimum at $x = 2$ in $f_1(x) = (x-2)^2$ or in $f_2(x) = (x-2)^5(x^2+x+1)$.

d) We only consider the first component x_i, the second component y_i behaves identically. We have

$$\Delta x_i = \left(\frac{2}{3}\right)^i - \left(\frac{2}{3}\right)^{i-1} = -\frac{1}{2}\left(\frac{2}{3}\right)^i$$

and therefore

$$\Delta^2 x_i = -\frac{1}{2}\left(\frac{2}{3}\right)^i - \left(-\frac{1}{2}\left(\frac{2}{3}\right)^{i-1}\right) = \frac{1}{4}\left(\frac{2}{3}\right)^i.$$

The terms of the Aitken sequence therefore evaluate to

$$\left(\frac{2}{3}\right)^i - \frac{\left(-\frac{1}{2}\left(\frac{2}{3}\right)^i\right)^2}{\frac{1}{4}\left(\frac{2}{3}\right)^i} = 0,$$

for any i. As the sequence x_i converges „exactly linear" towards 0, the Aitken convergence improvement works perfectly here and correctly „predicts" 0 as the goal of the convergence already after the first step.

Solution to Task 4

According to Task 9.3 we have $(x_1, y_1) = (2/3, 2/3)$, and the Hessian matrix at this point is $\begin{pmatrix} 12 \cdot (\frac{2}{3})^2 & 0 \\ 0 & 12 \cdot (\frac{2}{3})^2 \end{pmatrix} = \begin{pmatrix} \frac{16}{3} & 0 \\ 0 & \frac{16}{3} \end{pmatrix}$.The matrix $(A^1)^{-1}$ we need to compute should therefore be understood as an approximation for $\begin{pmatrix} \frac{3}{16} & 0 \\ 0 & \frac{3}{16} \end{pmatrix}$.

Further we have $(A^0)^{-1} = (H_f(x_0, y_0))^{-1} = \begin{pmatrix} \frac{1}{12} & 0 \\ 0 & \frac{1}{12} \end{pmatrix}$

$d^1 = \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} - \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} = \begin{pmatrix} 2/3 \\ 2/3 \end{pmatrix} - \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} -1/3 \\ -1/3 \end{pmatrix}$,

and with $\nabla f(x_1, y_1) = \nabla f(2/3, 2/3) = \begin{pmatrix} 4 \cdot (\frac{2}{3})^3 \\ 4 \cdot (\frac{2}{3})^3 \end{pmatrix} = \begin{pmatrix} \frac{32}{27} \\ \frac{32}{27} \end{pmatrix}$ we get

$g^1 = \nabla f(x_1, y_1) - \nabla f(x_0, y_0) = \begin{pmatrix} \frac{32}{27} \\ \frac{32}{27} \end{pmatrix} - \begin{pmatrix} 4 \\ 4 \end{pmatrix} = \begin{pmatrix} -\frac{76}{27} \\ -\frac{76}{27} \end{pmatrix}$.

Our approximation for the inverted Hessian matrix therefore is

$$
\begin{aligned}
(A^1)^{-1} &= (A^0)^{-1} - \frac{\left((A^0)^{-1} g^1 - d^1\right)(d^1)^T (A^0)^{-1}}{(d^1)^T (A^0)^{-1} g^1} \\
&= \begin{pmatrix} \frac{1}{12} & 0 \\ 0 & \frac{1}{12} \end{pmatrix} - \frac{\left(\begin{pmatrix} \frac{1}{12} & 0 \\ 0 & \frac{1}{12} \end{pmatrix}\begin{pmatrix} -76/27 \\ -76/27 \end{pmatrix} - \begin{pmatrix} -1/3 \\ -1/3 \end{pmatrix}\right)(-1/3, -1/3)\begin{pmatrix} \frac{1}{12} & 0 \\ 0 & \frac{1}{12} \end{pmatrix}}{(-1/3, -1/3)\begin{pmatrix} \frac{1}{12} & 0 \\ 0 & \frac{1}{12} \end{pmatrix}\begin{pmatrix} -76/27 \\ -76/27 \end{pmatrix}} \\
&= \begin{pmatrix} \frac{1}{12} & 0 \\ 0 & \frac{1}{12} \end{pmatrix} - \frac{\begin{pmatrix} -76/(27 \cdot 12) + 1/3 \\ -76/(27 \cdot 12) + 1/3 \end{pmatrix}(-1/3, -1/3)\begin{pmatrix} \frac{1}{12} & 0 \\ 0 & \frac{1}{12} \end{pmatrix}}{2 \cdot \frac{1}{3} \cdot \frac{1}{12} \cdot \frac{76}{27}} \\
&= \begin{pmatrix} \frac{1}{12} & 0 \\ 0 & \frac{1}{12} \end{pmatrix} - \frac{\begin{pmatrix} 8/81 \\ 8/81 \end{pmatrix}(-1/3, -1/3)\begin{pmatrix} \frac{1}{12} & 0 \\ 0 & \frac{1}{12} \end{pmatrix}}{2 \cdot \frac{1}{3} \cdot \frac{1}{12} \cdot \frac{76}{27}} \\
&= \begin{pmatrix} \frac{1}{12} & 0 \\ 0 & \frac{1}{12} \end{pmatrix} + \frac{\frac{8}{81} \cdot \frac{1}{3} \cdot \frac{1}{12}}{2 \cdot \frac{1}{3} \cdot \frac{1}{12} \cdot \frac{76}{27}} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \\
&= \begin{pmatrix} \frac{1}{12} & 0 \\ 0 & \frac{1}{12} \end{pmatrix} + \frac{1}{57} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \\
&= \begin{pmatrix} \frac{23}{228} & \frac{1}{57} \\ \frac{1}{57} & \frac{23}{228} \end{pmatrix} \approx \begin{pmatrix} 0.1009 & 0.0175 \\ 0.0175 & 0.1009 \end{pmatrix}
\end{aligned}
$$

Solution to Task 5 and Task 6

See R files on Moodle.