Adrian Willi

Exercise 8

Task 1

The Bazarra-Shetty function is defined by

$$f(x,y) := (x-2)^4 + (x-2y)^2.$$

a) Compute all first and second partial derivatives of the Bazarra-Shetty function. What do you notice when you compare the two second derivatives $\frac{\partial^2 f}{\partial x \partial y}(x,y)$ und $\frac{\partial^2 f}{\partial y \partial x}(x,y)$? Does this always have to be the case? (Look up Schwarz's theorem.)

b) Compute the gradient and the Hessian matrix of the function at the point $(0,0)$.

c) Find the stationary points, i.e. the potential extrema of the function.

d) Check whether the stationary points found are (local, global) minima, (local, global) maxima or saddle points. Is the following theorem applicable?

Theorem: Let $f : \mathbb{R}^2 \to \mathbb{R}$ be a twice continuously differentiable function (i.e., all second partial derivatives exist and are continuous). Furthermore, assume that (x_0, y_0) is a stationary point of f, i.e., we have $\frac{\partial f}{\partial x}(x_0,y_0) = \frac{\partial f}{\partial y}(x_0,y_0) = 0$. With $\delta := \det(H_f(x_0,y_0)) = \frac{\partial^2 f}{\partial x^2}(x_0,y_0) \cdot \frac{\partial^2 f}{\partial y^2}(x_0,y_0) - \left(\frac{\partial^2 f}{\partial x \partial y}(x_0,y_0)\right)^2$, the following statements hold:

- (x_0, y_0) is not an extremum if $\delta < 0$,
- (x_0, y_0) is a (local) minimum if $\delta > 0$ and $\frac{\partial^2 f}{\partial x^2}(x_0,y_0) > 0$,
- (x_0, y_0) is a (local) maximum if $\delta > 0$ und $\frac{\partial^2 f}{\partial x^2}(x_0,y_0) < 0$.

Task 2

Starting from the point $x^0 = (0,0)$, calculate the next iteration point when minimizing the Bazarra-Shetty function using the following methods:

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

b) Gradient descent with successive halving and parabola fitting;

c) Newton's method.

Task 1

$$f(x,y) := (x-2)^4 + (x-2y)^2$$

a)
$$\frac{\partial f}{\partial x}(x,y) = 4(x-2)^3 + 2(x-2y)$$
$$\frac{\partial f}{\partial y}(x,y) = -4(x-2y)$$
$$\frac{\partial^2 f}{\partial x^2}(x,y) = 12(x-2)^2 + 2$$
$$\frac{\partial^2 f}{\partial x \partial y}(x,y) = -4$$
$$\frac{\partial^2 f}{\partial y^2}(x,y) = 8$$
$$\frac{\partial^2 f}{\partial y \partial x}(x,y) = -4$$

Are the same. According to Schwarz's theorem it does not matter whether we first differentiate wrt. x and then wrt. y or vice versa.

b)
$$\nabla f(0,0) = \begin{bmatrix} \frac{\partial f}{\partial x}(0,0) \\ \frac{\partial f}{\partial y}(0,0) \end{bmatrix} = \begin{bmatrix} -32 \\ 0 \end{bmatrix}$$

$$H_f(0,0) = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2}(0,0) & \frac{\partial^2 f}{\partial x \partial y}(0,0) \\ \frac{\partial^2 f}{\partial y^2}(0,0) & \frac{\partial^2 f}{\partial y \partial x}(0,0) \end{bmatrix} = \begin{bmatrix} 50 & -4 \\ -4 & 8 \end{bmatrix}$$

c) To find stationary points we set the first derivatives to zero

$$\frac{\partial f}{\partial x}(x,y) = 4(x-2)^3 + 2(x-2y) = 0$$
$$\frac{\partial f}{\partial y}(x,y) = -4(x-2y) \overset{!}{=} 0 \quad |\text{simplify}$$
$$-4x + 8y \overset{!}{=} 0 \quad | +4x$$
$$8y \overset{!}{=} 4x \quad | :4$$
$$x \overset{!}{=} 2y$$

$\rightarrow x = 2y$

$$4(x-2)^3 + 2(x-x) = 0 \quad |\text{simplify}$$
$$4(x-2)^3 \overset{!}{=} 0$$
$$x \overset{!}{=} 2$$

with $x = 2$
$$-4(2-2y) = 0$$
$$-8 + 8y \overset{!}{=} 0$$
$$8y \overset{!}{=} 8$$
$$y \overset{!}{=} 1$$

There is exactly one stationary point at $(2,1)$

d) With
$$H_f(2,1) = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2}(2,1) & \frac{\partial^2 f}{\partial x \partial y}(2,1) \\ \frac{\partial^2 f}{\partial y \partial x}(2,1) & \frac{\partial^2 f}{\partial y^2}(2,1) \end{bmatrix} = \begin{bmatrix} 2 & -4 \\ -4 & 8 \end{bmatrix}$$

we get $\delta = \det(H_f(2,1)) = 0$. Therefore, we cannot apply the theorem. But, since the function f can only assume non-negative values, we still see the stationary point $(2,1)$ with $f(2,1) = 0$ is both local and global minimum.

Task 2 $\quad f(x,y) := (x-2)^4 + (x-2y)^2$

Start from point $x^0 = (0,0)$

a) Gradient descent with successive halving of the step size

At point $(x_0, y_0) = (0,0)$ we have $f(0,0) = 16$ — $(0-2)^4 + (0-2\cdot 0)^2 = (-2)^4 = 16$

$\nabla f(0,0) = \begin{bmatrix} -32 \\ 0 \end{bmatrix}$

β	$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} - \beta \cdot \nabla f(x_0,y_0) = \begin{pmatrix} 0 \\ 0 \end{pmatrix} - \beta \begin{pmatrix} -32 \\ 0 \end{pmatrix}$	$f(x_1, y_1)$
1	$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} - 1 \begin{pmatrix} -32 \\ 0 \end{pmatrix} = \begin{pmatrix} 32 \\ 0 \end{pmatrix}$	811'024
1/2	$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} - \frac{1}{2} \begin{pmatrix} -32 \\ 0 \end{pmatrix} = \begin{pmatrix} 16 \\ 0 \end{pmatrix}$	38'672
1/4	$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} - \frac{1}{4} \begin{pmatrix} -32 \\ 0 \end{pmatrix} = \begin{pmatrix} 8 \\ 0 \end{pmatrix}$	1'360
1/8	$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} - \frac{1}{8} \begin{pmatrix} -32 \\ 0 \end{pmatrix} = \begin{pmatrix} 4 \\ 0 \end{pmatrix}$	32
1/16	$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} - \frac{1}{16} \begin{pmatrix} -32 \\ 0 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \end{pmatrix}$	$4 < 16$ done!

The next iteration point therefore is $x^1 = (x_1, y_1) = (2,0)$ (when only using successive halving)

b) Using the result 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 $\beta = 1/16$

	t	$P(t) \overset{!}{=} f\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix} - t \begin{pmatrix} -32 \\ 0 \end{pmatrix}\right)$
$P(0)$	0	$P(0) \overset{!}{=} f\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix} - 0 \begin{pmatrix} -32 \\ 0 \end{pmatrix}\right) = f\begin{pmatrix} 0 \\ 0 \end{pmatrix} = 16$
$P(\beta)$	1/16	$P(1/16) \overset{!}{=} f\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix} - 1/16 \begin{pmatrix} -32 \\ 0 \end{pmatrix}\right) = f\begin{pmatrix} 2 \\ 0 \end{pmatrix} = 4$
$P(2\beta)$	1/8	$P(1/8) \overset{!}{=} f\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix} - 1/8 \begin{pmatrix} -32 \\ 0 \end{pmatrix}\right) = f\begin{pmatrix} 4 \\ 0 \end{pmatrix} = 32$

$$\Rightarrow \beta^* = \frac{-b}{2a} = \frac{\beta}{2} \cdot \frac{3 \cdot P(0) - 4 \cdot P(\beta) + P(2\beta)}{P(0) - 2 \cdot P(\beta) + P(2\beta)}$$

$$= \frac{1/16}{2} \cdot \frac{3 \cdot 16 - 4 \cdot 4 + 32}{16 - 2 \cdot 4 + 32} = \frac{1/16}{2} \cdot \frac{64}{40} = \underline{\underline{0.05}} = \beta^*$$

Therefore, we consider

$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} - 1/20 \begin{pmatrix} -32 \\ 0 \end{pmatrix} = \begin{pmatrix} 8/5 \\ 0 \end{pmatrix} = \begin{pmatrix} 1.6 \\ 0 \end{pmatrix}$ as our next iteration point. But first we check the resulting function value

For $\beta = 1/16$ we got $f(2,0) = 4$, for $\beta^* = 1/20$ we get $f(1.6, 0) = 2.5856\ldots$ which is better. The next iteration point therefore is $x^1 = (x_1, y_1) = (1.6, 0)$

c) $x^1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix} - (H_f(0,0))^{-1} \cdot \nabla f(0,0)$ $\qquad (H_f(0,0))^{-1} = \begin{pmatrix} 50 & -4 \\ -4 & 8 \end{pmatrix}^{-1} = \begin{pmatrix} 1/48 & 1/96 \\ 1/96 & 25/192 \end{pmatrix}$

$= \begin{pmatrix} 0 \\ 0 \end{pmatrix} - \begin{pmatrix} 1/48 & 1/96 \\ 1/96 & 25/192 \end{pmatrix} \cdot \begin{pmatrix} -32 \\ 0 \end{pmatrix}$

$= \begin{pmatrix} 2/3 \\ 1/3 \end{pmatrix}$ ← This is the next iteration point

Task 3

Consider the function

$$f(x, y, z) := x^2 + y^2 + z^2.$$

a) Calculate the first 3 iteration points when minimizing the function f using gradient descent and successive halving of the step size, starting from $x^0 = (3, 1, -2)$.

b) Calculate the first 3 iteration points when minimizing the function f using Newton's method, again starting from $x^0 = (3, 1, -2)$.

c) What do you observe in both cases? Would we observe the same thing for different starting points? Would we observe the same thing for other quadratic functions like e.g. $f(x, y, z) := x^2 + 2y^2 + 3z^2$ and arbitrary starting points?

Task 4 (*optional)

a) Study the three implementations/variants of the gradient method available from the homepage (constant step size, successive halving, successive halving and parabola fitting), 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 these methods on your programmable calculator.

b) Determine the global minimum of the following function with the programs from a):

$$f(x_1, x_2, x_3, x_4, x_5) := 2(x_1^2 + x_2^2 - 2x_1 - 2x_2 - x_3 - x_4) + x_3^2 + x_4^2 + \frac{15 + x_5^2}{2} - x_5.$$

Task 3 $\quad f(x,y,z) := x^2 + y^2 + z^2$

a) $\frac{\partial f}{\partial x} = 2x \quad \frac{\partial^2 f}{\partial x^2} = 2 \quad \frac{\partial^2 f}{\partial x \partial y} = 0 \quad \frac{\partial^2 f}{\partial x \partial z} = 0$

$\frac{\partial f}{\partial y} = 2y \quad \frac{\partial^2 f}{\partial y^2} = 2 \quad \frac{\partial^2 f}{\partial y \partial x} = 0 \quad \frac{\partial^2 f}{\partial y \partial z} = 0$

$\frac{\partial f}{\partial z} = 2z \quad \frac{\partial^2 f}{\partial z^2} = 2 \quad \frac{\partial^2 f}{\partial z \partial x} = 0 \quad \frac{\partial^2 f}{\partial z \partial y} = 0$

we start at $x^0 = (3,1,-2)$ these we have $f(3,1,-2) = 14$ and

$$\nabla f(3,1,-2) = \begin{pmatrix} 2\cdot 3 \\ 2\cdot 1 \\ 2\cdot(-2) \end{pmatrix} = \begin{pmatrix} 6 \\ 2 \\ -4 \end{pmatrix}$$

$9 + 1 + 4 \quad 14$

β	$f(x_1 y_1)$
$\begin{pmatrix} 3 \\ 1 \\ -2 \end{pmatrix} - 1 \cdot \begin{pmatrix} 6 \\ 2 \\ -4 \end{pmatrix} = \begin{pmatrix} -3 \\ -1 \\ 2 \end{pmatrix}$	14
$\begin{pmatrix} 3 \\ 1 \\ -2 \end{pmatrix} - \frac{1}{2} \begin{pmatrix} 6 \\ 2 \\ -4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$	0 all subsequent iteration points remain at $(0,0,0)$

b) $$x_1 = \begin{pmatrix} 3 \\ 1 \\ -2 \end{pmatrix} - \begin{pmatrix} 0.5 & 0 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0 & 0.5 \end{pmatrix} \cdot \begin{pmatrix} 6 \\ 2 \\ -4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$H_f(x,y,z) = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

$$(H_f(x,y,z))^{-1} = \begin{bmatrix} 0.5 & 0 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0 & 0.5 \end{bmatrix}$$

Regardless of the starting point we get the same result!
All subsequent iteration points remain at $(0,0,0)$

c) It's just a special case (for symmetry reasons) and we were lucky with choosing $\beta = 1$.

Solutions to Exercise 8

Solution to Task 1

a) We get

$$\begin{aligned}
f(x,y) &= (x-2)^4 + (x-2y)^2, \\
\frac{\partial f}{\partial x}(x,y) &= 4(x-2)^3 + 2(x-2y), \\
\frac{\partial f}{\partial y}(x,y) &= -4(x-2y), \\
\frac{\partial^2 f}{\partial x^2}(x,y) &= 12(x-2)^2 + 2, \\
\frac{\partial^2 f}{\partial x \partial y}(x,y) &= -4, \\
\frac{\partial^2 f}{\partial y \partial x}(x,y) &= -4, \\
\frac{\partial^2 f}{\partial y^2}(x,y) &= 8.
\end{aligned}$$

The partial derivatives $\frac{\partial^2 f}{\partial x \partial y}(x,y)$ and $\frac{\partial^2 f}{\partial y \partial x}(x,y)$ are equal. According to Schwarz's theorem, it does not matter whether we first differentiate with respect to x and then with respect to y, or vice versa.

b) The gradient at the point $(0,0)$ is

$$\nabla f(0,0) = \begin{pmatrix} \frac{\partial f}{\partial x}(0,0) \\ \frac{\partial f}{\partial y}(0,0) \end{pmatrix} = \begin{pmatrix} -32 \\ 0 \end{pmatrix}.$$

The Hessian matrix at the point $(0,0)$ is

$$H_f(0,0) = \begin{pmatrix} \frac{\partial^2 f}{\partial x^2}(0,0) & \frac{\partial^2 f}{\partial x \partial y}(0,0) \\ \frac{\partial^2 f}{\partial y \partial x}(0,0) & \frac{\partial^2 f}{\partial y^2}(0,0) \end{pmatrix} = \begin{pmatrix} 50 & -4 \\ -4 & 8 \end{pmatrix}.$$

c) The stationary points can be found by setting the first derivatives to zero, since the gradient must be equal to the null vector.

$$\begin{aligned}
\frac{\partial f}{\partial x}(x,y) &= 4(x-2)^3 + 2(x-2y) = 0 \\
\frac{\partial f}{\partial y}(x,y) &= -4(x-2y) = 0
\end{aligned}$$

From the second equation it follows that $x = 2y$. Substituting this into the first equation gives

$$\begin{aligned}
& 4(x-2)^3 + 2(x-x) = 0 \\
\Leftrightarrow\ & 4(x-2)^3 = 0 \\
\Leftrightarrow\ & x = 2.
\end{aligned}$$

With $x = 2y$, it follows that $y = 1$. Therefore there is exactly one stationary point: $(2,1)$.

d) With

$$H_f(2,1) = \begin{pmatrix} \frac{\partial^2 f}{\partial x^2}(2,1) & \frac{\partial^2 f}{\partial x \partial y}(2,1) \\ \frac{\partial^2 f}{\partial y \partial x}(2,1) & \frac{\partial^2 f}{\partial y^2}(2,1) \end{pmatrix} = \begin{pmatrix} 2 & -4 \\ -4 & 8 \end{pmatrix}$$

we get $\delta = \det(H_f(2,1)) = 0$. Therefore the theorem cannot be applied. However, since the function f can only assume non-negative values, we still see that the stationary point $(2,1)$ with $f(2,1) = 0$ is both a local and a global minimum.

Remark: The fact that H_f is singular at $(2,1)$ is one of the reasons the Bazarra-Shetty function is difficult to optimize, and therefore used as a testbed for different optimization methods. We will get back to this point in Task 9.3.

Solution to Task 2

a) Using our results from Task 1b), at the point $(x_0, y_0) = (0,0)$ we have

$$f(0,0) = 16,$$
$$\nabla f(0,0) = \begin{pmatrix} -32 \\ 0 \end{pmatrix} \quad .$$

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

β	$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} - \beta \cdot \nabla f(x_0, y_0) = \begin{pmatrix} 0 \\ 0 \end{pmatrix} - \beta \begin{pmatrix} -32 \\ 0 \end{pmatrix}$	$f(x_1, y_1)$
1	$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} - 1 \begin{pmatrix} -32 \\ 0 \end{pmatrix} = \begin{pmatrix} 32 \\ 0 \end{pmatrix}$	811024
$\frac{1}{2}$	$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} - \frac{1}{2} \begin{pmatrix} -32 \\ 0 \end{pmatrix} = \begin{pmatrix} 16 \\ 0 \end{pmatrix}$	38672
$\frac{1}{4}$	$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} - \frac{1}{4} \begin{pmatrix} -32 \\ 0 \end{pmatrix} = \begin{pmatrix} 8 \\ 0 \end{pmatrix}$	1360
$\frac{1}{8}$	$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} - \frac{1}{8} \begin{pmatrix} -32 \\ 0 \end{pmatrix} = \begin{pmatrix} 4 \\ 0 \end{pmatrix}$	32
$\frac{1}{16}$	$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} - \frac{1}{16} \begin{pmatrix} -32 \\ 0 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \end{pmatrix}$	$4 < 16$ done!

The next iteration point therefore is $x^1 = (x_1, y_1) = (2,0)$ (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 $\beta = 1/16$.)

t	$P(t) \overset{!}{=} f\left(\begin{pmatrix}0\\0\end{pmatrix} - t\begin{pmatrix}-32\\0\end{pmatrix}\right)$
0	$P(0) \overset{!}{=} f\left(\begin{pmatrix}0\\0\end{pmatrix} - 0\begin{pmatrix}-32\\0\end{pmatrix}\right) = f\begin{pmatrix}0\\0\end{pmatrix} = 16$
$\frac{1}{16}$	$P(\frac{1}{16}) \overset{!}{=} f\left(\begin{pmatrix}0\\0\end{pmatrix} - \frac{1}{16}\begin{pmatrix}-32\\0\end{pmatrix}\right) = f\begin{pmatrix}2\\0\end{pmatrix} = 4$
$\frac{1}{8}$	$P(\frac{1}{8}) \overset{!}{=} f\left(\begin{pmatrix}0\\0\end{pmatrix} - \frac{1}{8}\begin{pmatrix}-32\\0\end{pmatrix}\right) = f\begin{pmatrix}4\\0\end{pmatrix} = 32$

The vertex of the parabola is therefore at $\beta^* = -\frac{b}{2a} = 1/20 = 0.05$.
Therefore, we consider

$$\begin{pmatrix}x_1\\y_1\end{pmatrix} = \begin{pmatrix}0\\0\end{pmatrix} - \frac{1}{20}\begin{pmatrix}-32\\0\end{pmatrix} = \begin{pmatrix}8/5\\0\end{pmatrix} = \begin{pmatrix}1.6\\0\end{pmatrix}$$

as our next iteration point. Before actually choosing it, we need to check the resulting function value: For $\beta = 1/16$ we got the function value $f(2,0) = 4$ (see a)), for $\beta^* = 1/20$ we get the function value $f(1.6,0) = 2.5856\ldots$, which is better. The next iteration point therefore is $x^1 = (x_1, y_1) = (1.6, 0)$.

c) The next iteration point $x^1 = (x_1, y_1)$ when minimizing the Bazarra-Shetty function with Newton's method starting from $x^0 = (0,0)$ is

$$x^1 = \left(\begin{smallmatrix}0\\0\end{smallmatrix}\right) - \left(H_f(0,0)\right)^{-1} \cdot \nabla f(0,0).$$

In order to calculate the new point, the matrix $H_f(0,0)$ we determined in Task 1b) needs to be inverted. This is most conveniently done with the formula

$$\begin{pmatrix}a & b\\c & d\end{pmatrix}^{-1} = \frac{1}{\det(A)}\begin{pmatrix}d & -b\\-c & a\end{pmatrix} = \frac{1}{ad-bc}\begin{pmatrix}d & -b\\-c & a\end{pmatrix}.$$

We get

$$\left(H_f(0,0)\right)^{-1} = \begin{pmatrix}50 & -4\\-4 & 8\end{pmatrix}^{-1} = \frac{1}{384}\begin{pmatrix}8 & 4\\4 & 50\end{pmatrix} = \begin{pmatrix}\frac{1}{48} & \frac{1}{96}\\\frac{1}{96} & \frac{25}{192}\end{pmatrix},$$

and therefore the next iteration point is

$$\begin{pmatrix}x_1\\y_1\end{pmatrix} = \begin{pmatrix}0\\0\end{pmatrix} - \left(H_f(0,0)\right)^{-1} \cdot \nabla f(0,0) = -\begin{pmatrix}\frac{1}{48} & \frac{1}{96}\\\frac{1}{96} & \frac{25}{192}\end{pmatrix}\begin{pmatrix}-32\\0\end{pmatrix} = \begin{pmatrix}\frac{2}{3}\\\frac{1}{3}\end{pmatrix}.$$

Solution to Task 3

a) We arrive at the minimum $(0,0,0)$ after only one step, regardless of the starting point (x_0, y_0, z_0):

The gradient at a point (x, y, z) is

$$\nabla f(x, y, z) = \begin{pmatrix} 2x \\ 2y \\ 2z \end{pmatrix}.$$

$\beta = 1$ does not result in any improvement in the value of the function:

$$\begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix} = \begin{pmatrix} x_0 \\ y_0 \\ z_0 \end{pmatrix} - 1 \cdot \begin{pmatrix} 2x_0 \\ 2y_0 \\ 2z_0 \end{pmatrix} = \begin{pmatrix} -x_0 \\ -y_0 \\ -z_0 \end{pmatrix},$$

and $f(-x_0, -y_0, -z_0) = f(x_0, y_0, z_0)$.
$\beta = 0.5$ already gives the minimum $(0, 0, 0)$ of the function:

$$\begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix} = \begin{pmatrix} x_0 \\ y_0 \\ z_0 \end{pmatrix} - 0.5 \cdot \begin{pmatrix} 2x_0 \\ 2y_0 \\ 2z_0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Note that all subsequent iteration points remain at $(0, 0, 0)$.

b) In this case also we are already finished after the first step, regardless of the starting point: The Hessian matrix at a point (x, y, z) is

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

with

$$\left(H_f(x, y, z)\right)^{-1} = \begin{pmatrix} 0.5 & 0 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0 & 0.5 \end{pmatrix}.$$

Thus we get

$$\begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix} = \begin{pmatrix} x_0 \\ y_0 \\ z_0 \end{pmatrix} - \begin{pmatrix} 0.5 & 0 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0 & 0.5 \end{pmatrix} \begin{pmatrix} 2x_0 \\ 2y_0 \\ 2z_0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix},$$

regardless of the starting point. All subsequent iteration points remain at $(0, 0, 0)$.

c)
- The phenomenon observed in a) is more or less specific to the example considered. In general, for quadratic functions the gradient does *not* point directly towards the optimum. (Here it happened for symmetry reasons, for a counterexample consider $f(x) = x^2 + 2y^2 + 3z^2$ and the starting point $x^0 = (1, 1, 1)$.) Moreover, we were lucky with the step size – the arbitrary choice of $\beta = 1$ for the first attempt turned out to be ideal here, but would lead to lots of jumping back and forth for the very similar function $f(x, y, z) = 0.99 \cdot (x^2 + y^2 + z^2)$.

- The phenomenon observed in b) occurs for *any* quadratic function f with a unique minimum, like e.g. $f(x, y, z) = x^2 - 2xy + 2y^2 + 3z^2 = (x - y)^2 + y^3 + 3z^2$, and any starting point. The reason for this surprising behavior is that Newton's method approximates the derivative/gradient of f by a linear function, and therefore implicitly f itself by a quadratic function. If f is quadratic to begin with, the 'approximation' turns out to be exact!

Solution to Task 4

a) See R scripts.

b) The global Minimum is at point $(1, 1, 1, 1, 1)$.