

## Exercise sheet 7

### Exercise 1 [Affine invariance of Newton's method]

We consider Newton's method for the minimization of the  $C^2$  function  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  defined by

$$\nabla^2 f(x^k) s^k = -\nabla f(x^k), \quad x^{k+1} = x^k + s^k. \quad (1)$$

- a) Show, that Newton's method is invariant to affine linear transformations of the form  $My + v = x$  with  $M \in \mathbb{R}^{n \times n}$  invertible and  $v \in \mathbb{R}^n$ . More precisely this means that the application of Newton's method to  $h(y) := f(My + v)$  with the initial guess  $y^0 = M^{-1}(x^0 - v)$  generates the sequences of points  $y^k = M^{-1}(x^k - v)$ , where  $x^k$  are the iterates of (1) with initial guess  $x^0$ .
- b) Is the same true for the gradient method  $x^{k+1} = x^k - \nabla f(x^k)$ ?

### Exercise 2 [Barzilai Borwein method]

Let  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  be a  $C^2$  function. We set  $s^k = x^k - x^{k-1}$  and  $y^k = \nabla f(x^k) - \nabla f(x^{k-1})$ . Then the Taylor formula implies

$$\nabla f(x^{k-1}) = \nabla f(x^k) - \nabla^2 f(x^k) s^k + o(\|s^k\|)$$

that

$$y^k - \nabla^2 f(x^k) s^k \approx 0 \quad \text{and} \quad \nabla^2 f(x^k)^{-1} y^k - s^k \approx 0.$$

The main idea of this method is the approximation of  $\nabla^2 f(x^k)$  by a diagonal matrix  $\alpha_k I$  in terms of the above two equations. The variable  $\alpha^k$  is chosen in an optimal way, i.e.

$$a_1^k = \operatorname{argmin}_{\alpha \in \mathbb{R}} \|y^k - \alpha s^k\|_2^2, \quad a_2^k = \operatorname{argmin}_{\alpha \in \mathbb{R}} \left\| \frac{1}{\alpha} y^k - s^k \right\|_2^2$$

- a) Show that  $a_1^k = \frac{\langle s^k, y^k \rangle}{\|s^k\|^2}$
- b) Show that  $a_2^k = \frac{\|y^k\|^2}{\langle s^k, y^k \rangle}$

For a given initial guess  $x^0 \in \mathbb{R}^n$ , the first step is calculated as  $x^1 = x^0 - \sigma \nabla f(x^0)$  where  $\sigma$  is calculated by the Armijo rule. The following iterations are calculated by

$$x^{k+1} = x^k - \frac{1}{\alpha_i^k} \nabla f(x^k), \quad i = \{1, 2\}$$

if  $\alpha_i^k > 0$  and  $x^{k+1} = x^k - \sigma \nabla f(x^k)$  else. Again  $\sigma$  is calculated by the Armijo rule. Moreover we consider the following variants:

- $i = 1$  for all  $k$
- $i = 2$  for all  $k$
- $i = 1$  if  $k$  is even and  $i = 2$  if  $k$  is uneven

**Exercise 3** [Programming exercise: Barzilai Borwein method]

Repeat Exercise 4 a) from exercise sheet 4 using the described version of the Barzilai Borwein method for  $\varepsilon = 10^{-3}, 10^{-4}, \dots, 10^{-12}$ . Describe your findings and compare your results with your previous results. Moreover, plot  $f(x^k)$  in a semi logarithmic plot. What do you notice? Which variant performs best?

Hand in by email (philip.trautmann@uni-graz.at) until 02.12.2019, 23:59 o'clock.