SoSe 2020 Kiel University
Dr. Joscha Reimer Department of Computer Science

Optimization and Data Science
5. Homework exercises

Theoretical exercise 1:
Consider the general descent method applied to a continuous differentiable function $f : \mathbb{R}^n \to \mathbb{R}$*. Let* $(x_k)_{k\in\mathbb{N}_0}$ *be the resulting sequence of iterates.*

a) *Prove that if* x^* *and* $\hat{x}$ *are two accumulation points of* $(x_k)_{k\in\mathbb{N}_0}$*, then* $f(x^*) = f(\hat{x})$*.*

b) *Prove that if* x^* *is an accumulation points of* $(x_k)_{k\in\mathbb{N}_0}$*, then* x^* *is not a strict maximum.*

c) *Find an example where the general descent method stops at a strict maximum.*

Hint: Use the definition of accumulation points and the continuity of f to prove a) and the definition of accumulation points and strict maxima to prove b).
A sequence $(x_k)_{k\in\mathbb{N}_0}$ has an accumulation point x^* if and only if a subsequence $(x_{n_k})_{k\in\mathbb{N}_0}$ exists such that $\lim_{k\to\infty} x_{n_k} = x^*$. $(x_{n_k})_{k\in\mathbb{N}_0}$ is a subsequence of $(x_k)_{k\in\mathbb{N}_0}$ if $(n_k)_{k\in\mathbb{N}_0}$ is a strictly monotonically increasing sequence of natural numbers.

Theoretical exercise 2:
Consider the general descent method applied to

$$f : \mathbb{R}^2 \to \mathbb{R}, x \mapsto \tfrac{1}{2}\|x\|_2^2$$

with $x_0 \in \mathbb{R}^2 \setminus \{0\}$ *and search direction*

$$d_k := s_k - \frac{1}{2^{k+3}} g_k$$

with $g_k := \nabla f(x_k)$ *and* $s_k \in \mathbb{R}^n$ *such that* $s_k \perp g_k$ *and* $\|d_k\|_2 = \|g_k\|_2$ *for all* $k \in \mathbb{N}_0$*. Let* $(x_k)_{k\in\mathbb{N}_0}$ *be the resulting sequence of iterates. Prove that*

a) $\rho_k < 2^{-(k+2)}$ *for all* $k \in \mathbb{N}_0$ *(where* ρ_k *are the step-sizes).*

b) $\|x_0 - x_{k+1}\|_2 \leq \frac{1}{2}\|x_0\|_2$ *for all* $k \in \mathbb{N}_0$

c) $(x_k)_{k\in\mathbb{N}_0}$ *has no accumulation point which is a local minimum of* f*.*

d) d_k *are descent directions for all* $k \in \mathbb{N}_0$*.*

e) $(d_k)_{k\in\mathbb{N}_0}$ *is not gradient-related.*

Hint: A telescoping sum, the triangular inequality and the partial sum formula of the geometric series can be useful for proving b). The definition of descent directions and gradient-related directions might be useful for proving d) and e).
$v \perp w$ for arbitrary $v, w \in \mathbb{R}^n$ means that the vectors are orthogonal, which means $v^T w = 0$.

Programming exercise 1:

a) *Implement the Armijo step-size algorithm.*

b) *Implement the gradient method using the Armijo step-size algorithm.*

c) *Test your implementation using the Roosenbrock and the Bazaraa-Shetty function with different start points.*

The solutions of the theoretical exercises will be discussed on 11. Mai 2020.