Exercise sheet 4

Exercise 1 [Decay condition]

We consider a strongly convex function $f : \mathbb{R}^n \to \mathbb{R}$, with associated parameter $m > 0$. We also assume that ∇f is Lipschitz with modulus $L > 0$. We consider a steepest gradient descent method of the form:

$$\forall k \in \mathbb{N}, \quad x_{k+1} = x_k - \alpha_k \nabla f(x_k),$$

where the step length α_k is computed in such a way that for all k,

$$-c_1 \|\nabla f(x_k)\|^2 \leq \langle \nabla f(x_{k+1}), \nabla f(x_k) \rangle \leq c_2 \|\nabla f(x_k)\|^2,$$

for some constants c_1 and $c_2 > 0$.

a) Prove that $\alpha_k \leq (1 + c_1)/m$.

b) We set $\phi_k(\alpha) = f(x_k - \alpha \nabla f(x_k))$. Prove that ϕ_k' is Lipschitz-continuous, compute $\phi_k'(0)$ and $\phi_k'(\alpha_k)$, and deduce that $\alpha_k \geq (1 - c_2)/L$.

c) Prove that there exist values of c_1 and $c_2 > 0$ such that there exists $\theta > 0$ such that for all k, the decay condition is satisfied:

$$f(x_k) - f(x_{k+1}) \geq \theta \|\nabla f(x_k)\|^2.$$

d) Prove that if the decay condition is satisfied, then the sequence $(x_k)_{k \in \mathbb{N}}$ converges to the unique minimizer of f.

Exercise 2 [Minimizing step-size rule]

Let $f(x) := \frac{1}{2}\langle Ax, x \rangle + \langle b, x \rangle$, $b \in \mathbb{R}^n$, $A \in \mathbb{R}^{n \times n}$ be symmetric and positive definite. Moreover let $s \in \mathbb{R}^n$ be a descent direction of f in the point $x \in \mathbb{R}^n$ and

$$\sigma^* := \arg\min_{\sigma \geq 0} f(x + \sigma s) = \phi(\sigma)$$

denotes the step-size calculated by the minimizing rule.

a) Explain why $\sigma^* > 0$ holds.

b) Show that σ^* exists and is unique. Calculate it.

c) Show that $\sigma = \sigma^*$ satisfies for all $\gamma \in (0, 1/2]$ the Armijo-condition

$$f(x + \sigma s) - f(x) \leq \sigma \gamma \nabla f(x)^\top s,$$

but not for all $\gamma > 1/2$.

d) Draw the graph of ϕ and visualize the results of c).

Exercise 3 [Curry-step-size-rule]

Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable and choose a $x^0 \in \mathbb{R}^n$, such that the level-set

$$N_f(x^0) = \{x \in \mathbb{R}^n |\, f(x) \leq f(x^0)\}$$

is compact. Moreover let ∇f be Lipschitz-continuous on $N_f(x^0)$, $x \in N_f(x^0)$ and $s \in \mathbb{R}^n$ a descent direction of f in x. The *Curry-step-size-rule* chooses the step-size $\sigma^* > 0$ given by the smallest positive stationary point of $\phi(\sigma) := f(x + \sigma s)$:

$$\sigma^* := \min\{\sigma > 0 \,|\, \nabla f(x + \sigma s)^\top s = 0\}.$$

a) Prove the existence and uniqueness of σ^*.
Hint: Assume that no positive stationary point of ϕ exists and use the Mean value theorem to derive a contradiction to the compactness of $N_f(x^0)$.

b) Prove using the Intermediate value theorem, that there exists a smallest $0 < \hat{\sigma} < \sigma^*$ such that $\nabla f(x + \hat{\sigma} s)^\top s = \frac{1}{2} \nabla f(x)^\top s$.

c) Prove: $f(x + \sigma^* s) - f(x) \leq f(x + \hat{\sigma} s) - f(x) \leq \frac{\hat{\sigma}}{2} \nabla f(x)^\top s = -\frac{\hat{\sigma}}{2} |\nabla f(x)^\top s|$.

(d) Use the Lipschitz-continuity of ∇f with the constant L and the Cauchy-Schwarz inequality, in order to prove that:

$$\frac{1}{2} |\nabla f(x)^\top s| \leq \hat{\sigma} L \|s\|^2.$$

e) Use c) and d) in order to prove that

$$f(x + \sigma^* s) \leq f(x) - \theta \left(\frac{\nabla f(x)^\top s}{\|s\|} \right)^2$$

with a constant $\theta > 0$ independent of x and s holds.

Exercise 4 [Programming exercise: Method of steepest descent]

Implement the method of steepest descent with Armijo line search using backtracking (Algorithm 2.4.1) from the lecture in Matlab or Octave. Divide the method into two functions:

1. The function `Armijo` calculates for given input a step-size `sigma` as its output:

```
function[sigma] = Armijo(f, gfx, x, s, beta, gamma)
```

with inputs and output:

f: A function-handle of the function f (see `feval` in the Matlab or Octave documentation)

gfx: The gradient of f evaluated at the point `x`

x: The current iterate

s: The current search direction

beta, gamma: The parameters of the Armijo-rule.

sigma: Step-size

2. A funktion `SteepDescent`, which uses the function `Armijo` and which has the following form:

```
function[x, steps] = SteepDescent(f, gf, x0, eps, maxsteps)
```

with inputs and outputs:

f, gf: function-handles of f and ∇f

x0: Initial guess x^0

eps: Stopping-tolerance for $\|\nabla f(x^k)\| \leq \varepsilon$

maxsteps: Maximal number of iterates

steps: Number of iterations

x: Approximate solution

a) Test your program with the *Rosenbrock-function*

$$f(x) := 100(x_2 - x_1^2)^2 + (1 - x_1)^2$$

with the initial guess $x^0 = (-1.2, 1)^T$ and parameters maxsteps $= 10000$, $\beta = 0.5$ and $\gamma = 10^{-4}$. Test different values of $\varepsilon = 10^{-1},\ 10^{-2},\ 10^{-3},\ 10^{-4},\ 10^{-5}$.

b) Formulate based on your results a conjecture about the point to which the method converges.

   - Is this point stationary?
   - Prove that this point is a local minimum. Is it local or global?
   - Exist other local or global minima?

(c) Plot the contour-lines of f. (Matlab-function `contour`). Explain why the method is so inefficient.

Hand in by email (philip.trautmann@uni-graz.at) until 28.10.2019, 23:59 o'clock.