Prof. Dr. Dr. h.c. Florian Rupp
Akaki Matcharashvili

The Gradient Descent Method

This exercise sheet consists of two parts: at first problems for the Additional/ Central Exercise Problems class are given. Their solution will be provided and can serve you as further blueprints when solving similar tasks, e.g. for the homework assignment. Then, the actual Homework Assignments are stated that will be discussed during the TTF in the following week. Please, hand-in your results of these assignments through MSTeams at the date and time specified in MSTeams.

Additional/ Central Exercise Problems:

Exercise 3.1: Applying the Gradient Descent Method

a) Consider the Gradient Descent Method. Show with the help of an example that if the step length σ is badly chosen, the gradient descent might diverge.
Hint: The iteration for gradient descent is given by $x^{k+1} = x^k - \sigma_k \nabla f(x^k)$. Consider the 1D case with a quadratic function for instance.

b) Let $n \geq 1$ be an integer and let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix (non necessarily positive definite) for which all of its eigenvalues are non-zero. Let $a \in \mathbb{R}^n$ be a given vector and we consider the function $f : \mathbb{R}^n \to \mathbb{R}$, defined as

$$f(x) \;=\; \tfrac{1}{2}\,(x-a)^T\, A^2\,(x-a)\ ,$$

where $A^2 = AA$.

(i) Using first and second order optimality conditions show that f has a unique global minimizer on $\mathbb{R}^n$ and determine this minimizer. Denote it by $\overline{x}$.

(ii) Write the updates in the Gradient Descent with exact step size selection starting from a point $x^0 \in \mathbb{R}^n$ to approximate the minimizer $\overline{x}$ of f that has been determined in (i). Determine the step size σ_k in each step.

Exercise 3.2: Optimal Positioning of A Broadcasting Tower — Consider the points $P_1 := (0,0)^T$, $P_2 := (1,0)^T$ and $P_3 := (0,1)^T$ in $\mathbb{R}^2$ and the closed triangle $T \subset \mathbb{R}^2$, with corners in these points. We want to find a point $\overline{x} \in T$, that minimizes the sum of the Euclidean distances to the points P_1, P_2 and P_3, i.e. a solution of the problem

$$\min_{x\in T} f(x) \;:=\; \sum_{i=1}^{3} \|x - P_i\|_2. \tag{1}$$

a) Show that (1) has at least one solution $\overline{x} \in T$.

b) Show that each (global) solution of (1) lies in the interior of T.
Hint: It holds that $f(0.25, 0.25) \approx 1.93469 < 2$.

c) Verify that the functions

$$f_i(x) \;:=\; \|x - P_i\|_2\,, \qquad i \;=\; 1\,,\,2\,,\,3\,,$$

are continuously differentiable in the interior of T, that $\|\nabla f_i(x)\|_2 = 1$ holds for all $x \in \operatorname{int}(T)$ and all i, and that $\nabla f_i(x)$ is always parallel to the vector $x - P_i$ for all $x \in \operatorname{int}(T)$ and all i.

d) Compute a global solution $\overline{x}$ of (1). Is this solution unique?
Hint: Use c) and argue cleverly!

Homework Assignment:

Problem 3.1: Directional Derivatives — Compute the partial derivatives and the directional derivative in the direction $d := (1,-1)^T$ in the point $(0,0)^T$ of the following functions mapping $\mathbb{R}^2$ to $\mathbb{R}$:

a) $f(x_1,x_2) = 2x_1^2 + 3x_1x_2 + x_2$.

b) $f(x_1,x_2) = x_1 \cdot \sin(x_2)$.

c) $f(x_1,x_2) = \frac{x_1^2 x_2}{x_1^4 + x_2^2}$ for $(x_1,x_2)^T \neq (0,0)^T$ and $f(x_1,x_2) = 0$ for $(x_1,x_2)^T = (0,0)^T$.

Problem 3.2: Local Optima Along Lines — Let $f : \mathbb{R}^2 \to \mathbb{R}$ be defined by $f(x,y) := y^2 - 3yx^2 + 2x^4$.

a) Compute the stationary points of f, i.e. the points with $\nabla f(x,y) = 0$.

b) Show that for any $d \in \mathbb{R}^2 \setminus \{0\}$ the scalar function $\phi_d(\sigma) := f(\sigma d)$ is bounded from below and has a strict local minimum at $\sigma = 0$. This implies that f starting at $(0,0)^T$ increases (locally) in every direction.

c) Is the point $(0,0)^T$ a local minimum of f?

Hint: Use SciLab to plot the functions and its level sets in order to illustrate the geometric situation.

Problem 3.3: Comparing the Efficiency of Step Size Rules — Consider the quadratic minimization problem

$$\min_{x \in \mathbb{R}^5} x^T A x \,,$$

where $A = (a_{i,j})$ is the 5×5 Hilbert matrix defined by

$$a_{i,j} \;=\; \frac{1}{i+j-1}\,, \qquad \text{for } i,j = 1,2,3,4,5\,.$$

The matrix can be constructed via the SciLab command `A = testmatrix('hilb',5)`. Run the following methods and compare the number of iterations required by each of the methods when the initial vector is $x^0 = (1,2,3,4,5)^T$ to obtain a solution x with accuracy $\varepsilon = \|\nabla f(x)\| \leq 10^{-4}$:

a) The Gradient Descent Method with backtracking step size rule and parameters $\gamma = 0.5$, $\beta = 1$, and $S_0 = 1$.

b) The Gradient Descent Method with backtracking step size rule and parameters $\gamma = 0.1$, $\beta = 0.5$, and $S_0 = 1$.

c) The Gradient Descent Method with exact line search.

Problem 3.4: Quadratic Problems — Let the function $f : \mathbb{R}^n \to \mathbb{R}$ be defined by

$$f(x) \;:=\; \tfrac{1}{2} x^T H x + b^T x + c\,.$$

where $H \in \mathbb{R}^{n\times n}$, $b \in \mathbb{R}^n$ and $c \in \mathbb{R}$.

a) Compute the gradient ∇f and the Hessian H_f for general $H \in \mathbb{R}^{n\times n}$, $b \in \mathbb{R}^n$ and $c \in \mathbb{R}$.

b) Show, that without loss of generality, we can assume that H is symmetric, i.e. without changing the function f we can replace H by a suitable symmetric matrix.

c) Show: If H is symmetric positive definite, then the function f is strictly convex. Moreover, it holds $\lim_{\|x\|\to\infty} f(x) = \infty$ (i.e. f is coercive).

d) Write the function

$$g(x_1,x_2) \;=\; 5x_1^2 + 5x_2^2 + 8x_1x_2 - 4x_1 - 2x_2 + 3$$

in the form $g(x) = \frac{1}{2}x^T H x + b^T x + c$ with a symmetric matrix $H \in \mathbb{R}^{2\times 2}$. Is H positive definite? What can we deduce from this, regarding existence and uniqueness of a minimizer? Compute the minimizer, if possible.

Problem 3.5: Source Localization Problem — In the Source Localization Problem we are given m locations of sensors $a_1, a_2, \dots, a_m \in \mathbb{R}^m$ and approximate distances $d_1, d_2, \dots, d_m \in \mathbb{R}_0^+$ between the sensors and an unknown 'source' located at $x \in \mathbb{R}^n$:

$$d_i \;\approx\; \|x - a_i\|_2 \,.$$

The problem is to find x given the locations and the approximate distances. A natural formulation as an optimization problem is to consider the non-linear least squares problem

$$\min_{x \in \mathbb{R}^n} \left(f(x) \;:=\; \sum_{i=1}^{m} \left(\|x - a_i\|_2 - d_i\right)^2 \right)$$

We will denote the set of sensor positions by $\mathcal{A} := \{a_1, a_2, \dots, a_m\}$. Show that the optimality condition $\nabla f(x) = 0$ (for $x \cancel{\not\ni} \mathcal{A}$) is the same as

$\notin$

$$x \;=\; \frac{1}{m} \left(\sum_{i=1}^{m} a_i + \sum_{i=1}^{m} d_i \frac{x - a_i}{\|x - a_i\|_2} \right) .$$

Problem 3.6: Performance of the Gradient Descent Method — Implement the Gradient Descent Method with the Armijo step size rule (with $S_0 = 1$) according to the following step size parameter combinations

	combination 1	combination 2	combination 3	combination 4	combination 5
γ	0.1	0.3	0.5	0.7	0.9
β	0.1	0.3	0.5	0.7	0.9

in order to apply it to find the minimizer of $f : \mathbb{R}^2 \to \mathbb{R}$ with

$$f(x_1, x_2) \;=\; \mathrm{e}^{x_1 + 3x_2 - 0.1} + \mathrm{e}^{x_1 - 3x_2 - 0.1} + \mathrm{e}^{-x_1 + 0.1} \,.$$

As initial point take a random sample from a Gaussian, i.e. `x0 = grand(2,1,"nor",0,1)`. Run the gradient descent code for 10 different initial values for each set of parameter combinations of γ and β in the step size selection. Plot the average number of required steps in dependency of γ and β and explain the plot.