MDS 6106 – Introduction to Optimization

Exercise Sheet 3

Assignment A3.1 (Bisection and Golden Section Method): (approx. *20* points)
In this first exercise, we investigate the performance of the bisection and golden section method.

a) We consider the optimization problem

$$\min_x \ f(x) := -\frac{1}{(x-1)^2}\left[\log(x) - \frac{2(x-1)}{x+1}\right] \qquad \text{s.t.} \qquad x \in [1.5, 4.5].$$

Implement the golden section method to solve this problem and output a solution with accuracy at least 10^{-5}.

b) Consider the minimization problem

$$\min_{x\in\mathbb{R}} \ g(x) \qquad \text{s.t.} \qquad x \in [0,1],$$

where g is given by $g(x) := e^{-x} - \cos(x)$. Solve this problem using the bisection and the golden section method. Compare the number of iterations required to recover a solution in $[0,1]$ with accuracy less or equal than 10^{-5}.

Assignment A3.2 (Descent Directions and Lipschitz Continuity): (approx. *20* points)
Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable function and consider $x \in \mathbb{R}^n$ with $\nabla f(x) \neq 0$. Verify the following statements:

a) Set $d = -(\nabla f(x)_j) \cdot e_j = -\frac{\partial f}{\partial x_j}(x) \cdot e_j$, where $e_j \in \mathbb{R}^n$ is the j-th unit vector and $j \in \{1, ..., n\}$ is an index satisfying

$$\left|\frac{\partial f}{\partial x_j}(x)\right| = \max_{1\leq i\leq n} \left|\frac{\partial f}{\partial x_i}(x)\right| = \|\nabla f(x)\|_\infty.$$

Then d is a descent direction of f at x.

b) Let f be twice continuously differentiable and define $d_i = -(\nabla f(x)_i)/\max\{\nabla^2 f(x)_{ii}, \varepsilon\}$ for all $i \in \{1, ..., n\}$ and for some $\varepsilon > 0$. Show that d is well-defined (we do not divide by zero) and that it is a descent direction of f at x.

c) Discuss whether the gradient of the following mappings is Lipschitz continuous or not:

- $f_1 : \mathbb{R}^3 \to \mathbb{R}$, $f_1(x) := \frac{3}{2}x_1^2 + 2x_1x_2 - \frac{1}{2}x_3^2$.
- $f_2 : \mathbb{R}^2 \to \mathbb{R}$, $f_2(x) := \log(1 + x_1^2) + x_2^4$.

Report the corresponding Lipschitz constant L in case ∇f_i, $i = 1, 2$, is Lipschitz continuous.

Assignment A3.3 (Implementing the Gradient Method): (approx. *60* points)
Implement the gradient descent method (Lecture L-06, slide 19) that was presented in the lecture as a function `gradient_method` in `MATLAB` or `Python`.

The following input functions and parameters should be considered:

- `obj`, `grad` – function handles that calculate and return the objective function $f(x)$ and the gradient $\nabla f(x)$ at an input vector $x \in \mathbb{R}^n$. You can treat these handles as functions or fields of a class or structure `f` or use them directly as input. (For example, your function can have the form `gradient_method(obj,grad,...)`).
- x^0 – the initial point.
- `tol` – a tolerance parameter. The method should stop whenever the current iterate x^k satisfies the criterion $\|\nabla f(x^k)\| \leq$ `tol`.

We want to analyze the performance of the gradient method for different step size strategies. In particular, we want to test and compare backtracking, exact line search, and diminishing step sizes. The following parameters will be relevant for these strategies:

- $s > 0$, $\sigma, \gamma \in (0, 1)$ – parameters for backtracking and the Armijo condition.
- `alpha` – a function that returns a pre-defined, diminishing step size $\alpha_k =$ `alpha(k)` satisfying $\alpha_k \to 0$ and $\sum \alpha_k = \infty$.
- You can use the golden section method from Assignment A3.1 to determine the exact step size α_k. The parameters for the golden section method are: `maxit` (maximum number of iterations), `tol` (stopping tolerance), $[0, \mathtt{a}]$ (the interval of the step size).

You can organize the latter parameters in an appropriate `options` class or structure. It is also possible to implement two separate algorithms for backtracking and diminishing step sizes. The method(s) should return the final iterate x^k that satisfies the stopping criterion.

a) Test your implementation on the following problem:

$$\min_{x \in \mathbb{R}^2} \ f(x) = f_1(x)^2 + f_2(x)^2,$$

where $f_1 : \mathbb{R}^2 \to \mathbb{R}$ and $f_2 : \mathbb{R}^2 \to \mathbb{R}$ are given by:

$$f_1(x) := -1 + x_1 + ((5 - x_2)x_2 - 2)x_2,$$
$$f_2(x) := -1 + x_1 + ((x_2 + 1)x_2 - 10)x_2.$$

This problem has the following five stationary points:

$$x_1^* = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad x_2^* = \begin{pmatrix} -11 \\ 1 + \sqrt{5} \end{pmatrix}, \quad x_3^* = \begin{pmatrix} -11 \\ 1 - \sqrt{5} \end{pmatrix}, \quad x_4^* = \begin{pmatrix} -\frac{13}{3} \\ -\frac{2}{3} \end{pmatrix}, \quad x_5^* = \begin{pmatrix} 1 \\ 2 \end{pmatrix}.$$

- Analyze which of the points $x_1^*, ..., x_5^*$ are (global/local) minimizer, (global/local) maximizer, or saddle points.
- Apply the gradient method with backtracking and parameters $(s, \sigma, \gamma) = (1, 0.5, 0.1)$, with diminishing step sizes
$$\alpha_k = \frac{0.02}{\log(k + 10)},$$
and exact line search (`maxit` $= 100$, `tol` $= 10^{-6}$, $\mathtt{a} = 2$) to solve the problem $\min_x f(x)$.

The algorithms should use the stopping tolerance $\texttt{tol} = 10^{-5}$. Test the methods using the initial point $(-5, 0)^\top$ and report the behavior and performance of the methods. In particular, for each of the step size strategies, compare the number of iterations and the point to which each algorithm converged.

b) Let us define the set $\mathcal{X}^0 := \{x \in \mathbb{R}^2 : x_1 \in \{-10, 0\}, x_2 \in [-5, 5]\} \cup \{x \in \mathbb{R}^2 : x_1 \in [-10, 0], x_2 \in \{-5, 5\}\}$ (this is a square). Run the methods:

- Gradient descent method with backtracking and $(s, \sigma, \gamma) = (1, 0.5, 0.1)$,
- Gradient method with exact line search and $\texttt{maxit} = 100$, $\texttt{tol} = 10^{-6}$, $\texttt{a} = 2$,

again with p different initial points selected from the set $\mathcal{X}^0$. The initial points should uniformly cover the different parts of the set $\mathcal{X}^0$ and you can use the tolerance $\texttt{tol} = 10^{-5}$ and $p \in [10, 25] \cap \mathbb{N}$. For each algorithm create a single figure that contains all of the solution paths generated for the different initial points. The initial points and limit points should be clearly visible. Add a contour plot of the function f in the background of each figure.

Bonus: You can also run an additional test with properly adjusted diminishing step sizes (to still guarantee convergence, you likely need to use smaller step sizes or restrict the chosen initial points).

Exercise E3.1 (Multiple Choice – Descent Directions):
Answer the following multiple choice questions and decide whether the statements are true or false. Give short explanations of your answer.

a) Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable, $x^k \in \mathbb{R}^n$, and let d^k be a descent direction of f at x^k. We consider the function $\phi(\alpha) := f(x^k + \alpha d^k) - f(x^k)$, $\alpha \geq 0$.

Which of the following statements is true?

☐ 1. It holds that $\phi(0) < 0$.

☐ 2. It holds that $\phi'(0) = 0$.

☐ 3. We have $\phi(0) = 0$.

☐ 4. It holds that $\phi'(0) = \nabla f(x^k + \alpha d^k)^\top d^k$.

☐ 5. ϕ is a function from $\mathbb{R}^n$ to $\mathbb{R}$.

☐ 6. ϕ is a function from $\mathbb{R}_+$ to $\mathbb{R}$.

☐ 7. We have $\phi'(0) < 0$.

☐ 8. It holds that $\phi'(0) = \nabla f(x^k)^\top d^k$.

b) The derivative f' of the function $f(x) = \sin(x)$ is Lipschitz continuous.

☐ True. ☐ False.

c) Let $x^0 \in \mathbb{R}^n$ be an initial point with $\nabla f(x^0) \neq 0$ and consider the point $x^1 = x^0 - \nabla f(x^0)$. Then, it holds that $f(x^1) < f(x^0)$.

☐ True. ☐ False.

d) We consider the function

$$f(x) = -x_1 + 4x_2 + 10x_1^2 - 9x_1x_2 - 5x_2^2.$$

Let us set $x = (1,1)^\top$. Which of the following statements is true?

- ☐ 1. $d = (0,0)^\top$ is a descent direction of f at x.
- ☐ 2. $d = (0,0)^\top$ is not a descent direction of f at x.
- ☐ 3. $d = (1,1)^\top$ is a descent direction of f at x.
- ☐ 4. $d = (1,1)^\top$ is not a descent direction of f at x.
- ☐ 5. $d = (1.5,1)^\top$ is a descent direction of f at x.
- ☐ 6. $d = (1.5,1)^\top$ is not a descent direction of f at x.
- ☐ 7. $d = (0,1)^\top$ is a descent direction of f at x.
- ☐ 8. $d = (0,1)^\top$ is not a descent direction of f at x.

Exercise E3.2 (Multiple Choice – Step Sizes):
Answer the following multiple choice questions and decide whether the statements are true or false. Try to give short explanations of your answer.

a) We use a general descent method (Lecture L-06, slide 8) to minimize a continuously differentiable function f. We consider the k-th iterate x^k and suppose that $\nabla f(x^k) \neq 0$. We choose the direction $d^k = -3\nabla f(x^k)$. Which of the following statements is true?

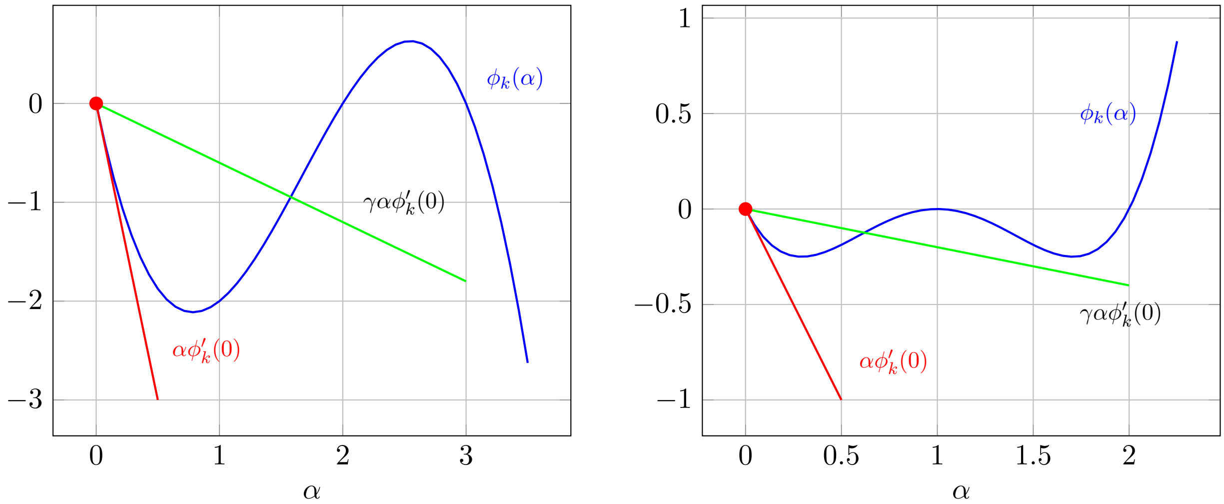
- ☐ 1. d^k is a descent direction.
- ☐ 2. For the step size $\alpha = 1$, we have $f(x^k + \alpha d^k) < f(x^k)$.
- ☐ 3. For the step size $\alpha = \frac{1}{14}$, it holds that $f(x^k + \alpha d^k) < f(x^k)$.
- ☐ 4. For the step size $\alpha = \frac{1}{7}$, it holds that $f(x^k + \alpha d^k) < f(x^k)$.
- ☐ 5. For the step size $\alpha = \frac{1}{3}$, we have $f(x^k + \alpha d^k) < f(x^k)$.
- ☐ 6. We can not ensure the condition $f(x^k + \alpha d^k) < f(x^k)$ for the different step sizes $\alpha \in \{1, \frac{1}{3}, \frac{1}{7}, \frac{1}{14}\}$. This depends on f.
- ☐ 7. There is $\varepsilon > 0$ such that $f(x^k + \alpha d^k) < f(x^k)$ for all $\alpha \in (0, \varepsilon]$.

b) Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable, $x^k \in \mathbb{R}^n$, and let d^k be a descent direction of f at x^k. We again consider the function $\phi(\alpha) := f(x^k + \alpha d^k) - f(x^k)$, $\alpha \geq 0$.

For a given parameter $\gamma \in (0,1)$, a step size $\alpha \geq 0$ is said to satisfy the Armijo condition if $\phi(\alpha) \leq \gamma\alpha \cdot \phi'(0)$. Backtracking or the Armijo line search technique tries to find the largest step size $\alpha \in \{1, \sigma, \sigma^2, \sigma^3, ...\}$, $\sigma \in (0,1)$, that satisfies the Armijo condition. Exact line search determines the step size $\bar{\alpha}$ as global minimizer of ϕ: $\bar{\alpha} = \arg\min_{\alpha \geq 0} \phi(\alpha)$.

Consider Figure 1(a). Which of the following statements is true?

- ☐ 1. Backtracking returns the step size $\alpha = 1$.
- ☐ 2. Suppose that $\sigma = \frac{1}{2}$. Then, backtracking returns the step size $\alpha = 1$.
- ☐ 3. We can not guarantee that Armijo line search returns the step size $\alpha = 1$, this depends on the choice of σ.
- ☐ 4. The Armijo condition is satisfied for all $\alpha \in [0, 1.25]$.
- ☐ 5. The exact step size $\bar{\alpha}$ lies in the interval $[0.5, 1]$.



(a) Example 1

(b) Example 2

Figure 1: Multiple choice & step sizes

☐ 6. We can not ensure that the exact step size $\bar{\alpha}$ lies in the interval $[0.5, 1]$.

☐ 7. The exact step size satisfies $\bar{\alpha} \geq 3$.

Consider Figure 1(b). Which of the following statements is true?

☐ 1. Backtracking returns the step size $\alpha = 1$.

☐ 2. Suppose that $\sigma = \frac{1}{2}$. Then, backtracking returns the step size $\alpha = 1$.

☐ 3. Suppose that $\sigma = \frac{1}{2}$. Then, backtracking returns the step size $\alpha = \frac{1}{2}$.

☐ 4. The Armijo condition is satisfied for all $\alpha \in [0, 0.75]$.

☐ 5. If $\phi(\alpha) \to \infty$ for $\alpha \to \infty$, then the exact step size $\bar{\alpha}$ lies in the interval $[0, 0.5]$.

☐ 6. If $\sigma \leq \frac{1}{2}$, then Armijo line search returns the step size $\alpha = \sigma$.

c) If the gradient descent method with Armijo line search terminates after k steps with $\nabla f(x^k) = 0$, then x^k is a local minimum of f.

☐ True. ☐ False.