Univ.-Prof. Dr. Radu Ioan Boţ

Vienna, November 15, 2022
To be discussed on November 22, 2022

Nonlinear Optimization

Exercise session 4

19. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a twice continuously differentiable function and $x^0 \in \mathbb{R}^n$ such that the lower level set $\mathcal{L}(x^0) := \{z \in \mathbb{R}^n : f(z) \leq f(x^0)\}$ is convex and bounded. Prove that ∇f is Lipschitz continuous on $\mathcal{L}(x^0)$. (3 points)

20. Show by means of an example that the boundedness from below of f is indispensable for the well-definiteness of the Wolfe-Powell step size strategy. (3 points)

21. Let $A \in \mathbb{R}^{n\times n}$ by a symmetric and positive definite matrix, $b \in \mathbb{R}^n$ and the quadratic function $f : \mathbb{R}^n \to \mathbb{R}^n, f(x) = (1/2)x^T Ax - b^T x$. Further, let $x, d \in \mathbb{R}^n$ be such that $\nabla f(x)^T d < 0$. Prove that the global minimum t^* of the function $\varphi : \mathbb{R} \to \mathbb{R}, \varphi(t) = f(x + td)$, is a Wolfe-Powell step size, even for $\sigma \leq 1/2$ and $\rho \geq 0$. (3 points)

22. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable function and $x^0 \in \mathbb{R}^n$. The *Curry rule* reads: for $x \in \mathcal{L}(x^0) := \{z \in \mathbb{R}^n : f(z) \leq f(x^0)\}$ and $d \in \mathbb{R}^n$ fulfilling $\nabla f(x)^T d < 0$ choose

$$t_C := \min\{t > 0 : \nabla f(x + td)^T d = 0\}$$

(t_C is the first critical point of f along the half-line $\{x+td : t \geq 0\}$). Prove: if $\mathcal{L}(x^0)$ is compact and the gradient ∇f is Lipschitz continuous on $\mathcal{L}(x^0)$, then the Curry rule is well-defined and efficient. (4 points)

23. Let $X \subseteq \mathbb{R}^n$ be a convex set. A function $f : X \to \mathbb{R}$ is said to be *strongly convex on X with modulus $\mu > 0$* if

$$f(\lambda x + (1 - \lambda)y) + \mu\lambda(1 - \lambda)\|x - y\|^2 \leq \lambda f(x) + (1 - \lambda)f(y)$$

for all $x, y \in X$ and all $\lambda \in [0, 1]$. Prove that the following statements are equivalent:

(a) f is strongly convex on X with modulus $\mu > 0$;

(b) $g : X \to \mathbb{R}$, $g(x) = f(x) - \mu\|x\|^2$, is convex on X.

(2 points)

24. Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable, $x^0 \in \mathbb{R}^n$, the level set $\mathcal{L}(x^0) = \{z \in \mathbb{R}^n : f(z) \leq f(x^0)\}$ be convex and f be strongly convex on $\mathcal{L}(x^0)$ with modulus $\mu > 0$.

(a) Prove that the set $\mathcal{L}(x^0)$ is compact.

(b) Prove that the optimization problem

$$\min_{x \in \mathcal{L}(x^0)} f(x)$$

has a unique optimal solution x^*.

(c) Prove that

$$\mu \|x - x^*\|^2 \leq f(x) - f(x^*) \ \forall x \in \mathcal{L}(x^0).$$

(3 points)

25. Implement the gradient algorithm with Armijo step size rule (Algorithm 6.1 in the lecture notes). Use as input data the starting vector x^0, the parameter for the stopping criterion ε, and the parameters σ and β for the determination of the Armijo step size. The sequence $x^0, x^1, x^2, ...$ containing the iteration history should be returned and the points $(x^0, f(x^0)), (x^1, f(x^1)), (x^2, f(x^2)), ...$ should be plotted on the graph of the function f.

The implemented algorithm should be tested for the following functions and input data values:

(a) $f(x) = \cos(x)$, $x^0 = 0.5$, ε=10^{-3}, σ=10^{-2} and β=0.5.

(b) $f(x_1, x_2) = (x_1^2 + x_2 - 11)^2 + (x_1 + x_2^2 - 7)^2$ (Himmelblau function), ε=10^{-1}, σ=10^{-2}, β=0.5 and $x^0 = (-0.27, -0.91)^T$, $x^0 = (-0.271, -0.91)^T$, $x^0 = (-0.25, -1.1)^T$ and $x^0 = (-0.25, -1)^T$.

(c) $f(x_1, x_2) = 6x_1^2 - 6x_1x_2 + 2x_2^2 + x_1 + x_2 + 1$, x^0=$(1, 2)^T$, ε=10^{-2}, σ=10^{-2} and β=0.5.

(4 points)

26. The gradient algorithm with Armijo step size rule should be employed for the minimization of the Rosenbrock function

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

with x^0=$(-1.2, 1)^T$, σ=10^{-4} and β=0.5. How many iterations are needed to fulfill the stopping criterion

$$\|\nabla f(x^k)\| \leq \varepsilon,$$

when ε takes the values $10^{-1}, ..., 10^{-5}$? Provide a graphical representation of the implementation history. Provide the vector x^{STOP} at which the algorithm stops and the value of the distance from x^{STOP} to the global minimum $x^* = (1, 1)^T$ of f. (3 points)