Univ.-Prof. Dr. Radu Ioan Boţ

Vienna, November 15, 2019
To be discussed on November 22, 2019

Nonlinear Optimization

Exercise session 4

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 $\{t_k\}_{k\geq 0} \subseteq \mathbb{R}$ be a sequence and $t = \liminf_{k\to+\infty} t_k$ its limit inferior. Prove that there exists a subsequence of $\{t_k\}_{k\geq 0}$ which converges to t. (2 points)

24. 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)

25. 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)

26. (Implementation task)

Implement the gradient algorithm with Armijo step size rule (Algorithm 6.1 in the lecture notes) in MATLAB. In this sense, create a file `gradienten_armijo.m` with

```
function X=gradienten_armijo(func, x0, sigma, beta, epsilon).
```

Here, `func` denotes a function handle, `x0` the starting vector, `epsilon` the parameter for the stopping criterion, and `sigma` and `beta` the parameters for the determination of the Armijo step size. A matrix $X = [x0, x1, x2, ...]$ containing the iteration history should be returned.

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

(a) $f(x) = \cos(x)$, `x0`=0.5, `epsilon`=10^{-3}, `sigma`=10^{-2} and `beta`=0.5.

(b) $f(x_1, x_2) = (x_1^2 + x_2 - 11)^2 + (x_1 + x_2^2 - 7)^2$ (Himmelblau function), `epsilon`=10^{-1}, `sigma`=10^{-2}, `beta`=0.5 and `x0` $= (-0.27, -0.91)^T$, `x0` $= (-0.271, -0.91)^T$, `x0` $= (-0.25, -1.1)^T$ and `x0` $= (-0.25, -1)^T$.

(c) $f(x_1, x_2) = 6x_1^2 - 6x_1x_2 + 2x_2^2 + x_1 + x_2 + 1$, `x0`=$(1, 2)^T$, `epsilon`=$10^{-2}$, `sigma`=$10^{-2}$ and `beta`=0.5.

(4 points)

27. (Implementation task)

The gradient algorithm with Armijo step size rule should be employed for minimizing the Rosenbrock function

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

with `x0`=$(-1.2, 1)^T$, `sigma`=10^{-4} and `beta`=0.5. How many iterations are needed in order to fulfill the stopping criterion

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

when `epsilon` 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 distance of x^{STOP} from the global minimum $x^* = (1, 1)^T$ of f. (3 points)

28. (Implementation task)

Implement the CG algorithm for linear systems (Algorithm 7.2 in the lecture notes) in MATLAB. Use as first line

```
function [x,Iter] = con_grad_lin(A, b, x0, epsilon)
```

Here, `A` denotes a symmetric and positive definite matrix, `b` the right-hand vector, `x0` the starting vector of the algorithm, and `epsilon` the parameter for the stopping criterion. The solution `x` und the number of performed iterations `Iter` should be returned.

The implemented algorithm should be tested on the following optimization problems:

(a) Minimize the function

$$f(x_1, x_2) = 2x_1^2 + x_2^2 - 4x_1 - 2x_2 + 3$$

on $\mathbb{R}^2$, for `x0` $= (5, -5)^T$ and `epsilon` $= 10^{-3}$.

(b) Minimize the function

$$f(x_1, x_2, x_3) = x_1^2 + 0.3x_1x_2 + 0.975x_2^2 + 0.01x_1x_3 + x_3^2 + 3x_1 - 4x_2 + x_3$$

on $\mathbb{R}^3$, for `x0` $= (0, 0, 0)^T$ and `epsilon` $= 10^{-8}$.

(4 points)

The generated files should be sent by November 29, 2019 to `radu.bot@univie.ac.at`.