Univ.-Prof. Dr. Radu Ioan Boţ

Vienna, December 6, 2022
To be discussed on December 13, 2022

Nonlinear Optimization

Exercise session 5

27. Let $x^* \in \mathbb{R}^n$ be an isolated accumulation point of the sequence $\{x^k\}_{k\geq 0} \subseteq \mathbb{R}^n$ with the property that for every subsequence $\{x^{k_l}\}_{l\geq 0}$ which converges to x^* the sequence $\{x^{k_l+1} - x^{k_l}\}_{l\geq 0}$ converges to 0. Prove that the entire sequence $\{x^k\}_{k\geq 0}$ converges to x^*.

    Definition. An accummulation point of a sequence is called isolated if it has an open neighbourhood that does not contain other accumulation points of the sequence. (4 points)

28. Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice continuously differentiable and $\{x^k\}_{k\geq 0} \subseteq \mathbb{R}^n$ a sequence generated by the gradient method (Algorithm 6.1). Prove that if $x^* \in \mathbb{R}^n$ is an accumulation point of $\{x^k\}_{k\geq 0}$ such that $\nabla^2 f(x^*)$ is positively definite, then the entire sequence $\{x^k\}_{k\geq 0}$ converges to x^*.

    Hint. Use Theorem 3.3 and Exercise 27. (3 points)

29. Prove that every strongly convex function is strictly convex and provide an example of a differentiable strictly convex function $f : \mathbb{R}^n \to \mathbb{R}$ that is not strongly convex. (2 points)

30. (Banach-Picard iteration) Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a Lipschitz continuous operator with Lipschitz constant $\beta \in [0, 1)$, $x^0 \in \mathbb{R}^n$ and set

    $$(\forall k \geq 0) \qquad x^{k+1} := T(x^k).$$

    Prove that there exists $x^* \in \mathbb{R}^n$ such that:

    (i) x^* is the unique fixed point of T;

    (ii) $\|x^{k+1} - x^*\| \leq \beta\|x^k - x^*\| \ \forall k \geq 0$;

    (iii) $\|x^k - x^*\| \leq \beta^k\|x^0 - x^*\| \ \forall k \geq 0$, which means that $\{x^k\}_{k\geq 0}$ converges to x^* with rate of convergence $O(\beta^k)$ as $k \to +\infty$.

    (3 points)

31. Let $A \in \mathbb{R}^{n\times n}$ be a symmetric and positive matrix, $b \in \mathbb{R}^n$ and the quadratic function $f : \mathbb{R}^n \to \mathbb{R}$, $f(x) = (1/2)x^T Ax - b^T x$. Further, let be $d^0, d^1, ..., d^{n-1} \in \mathbb{R}^n$ nonzero vectors with the property that

    $$(d^i)^T Ad^j = 0 \text{ for all } i, j = 0, ..., n-1, \ i \neq j.$$

    We consider the following algorithm for the minimization of the function f over $\mathbb{R}^n$:

1: Choose $x^0 \in \mathbb{R}^n$ and set $g^0 := Ax^0 - b$.

2: If $g^k = 0$: STOP. Set $m := k$. Then x^m is the global minimum of f.

3: Set

$$t_k := -\frac{(g^k)^T d^k}{(d^k)^T A d^k}, x^{k+1} := x^k + t_k d^k, g^{k+1} := g^k + t_k A d^k.$$

4: Set $k := k+1$ and go to Step 2.

Show by means of complete induction over k: if $g^0, ..., g^k \neq 0$, then x^{k+1} is an optimal solution of the problem

$$(P_k) \qquad\qquad \text{minimize } f(x), x \in x^0 + \operatorname{span}\{d^0, ..., d^k\}.$$

Here, $\operatorname{span}\{d^0, ..., d^k\}$ denotes the linear subspace generated by the vectors $\{d^0, ..., d^k\}$. Since $x^0 + \operatorname{span}\{d^0, ..., d^{n-1}\} = \mathbb{R}^n$, the algorithm stops after $m = n$ steps at latest and x^m is the global minimizer of f.

Hint. Show that $x^{k+1} \in x^0 + \operatorname{span}\{d^0, ..., d^k\}$ and $(g^{k+1})^T d^i = 0, i = 0, ..., k$. (4 points)

32. Replace in the Fletcher-Reeves algorithm (Algorithm 7.5 in the lecture notes) β_k^{FR} by (*the Myers formula*)

$$\beta_k^M := -\frac{\|\nabla f(x^{k+1})\|^2}{\nabla f(x^k)^T d^k}.$$

Prove that if f is continuously differentiable and bounded from below and for every $k \geq 0$ it holds $\nabla f(x^k) \neq 0$, then the algorithm is well-defined.

Hint. Prove that $\nabla f(x^k)^T d^k < 0$ for all $k \geq 0$. (3 points)

33. Solve the minimization problem

$$\min_{x \in \mathbb{R}^n} f(x),$$

where

$$f : \mathbb{R}^n \to \mathbb{R}, \quad f(x) = \begin{cases} \frac{1}{c}\|x\| - \frac{1}{2c^2}, & \text{if } \|x\| \geq \frac{1}{c}, \\ \frac{1}{2}\|x\|^2, & \text{otherwise,} \end{cases}$$

with the gradient algorithm and the fast gradient algorithm

(i) by considering different values for the dimension $n \in \{1, 10, 50, 500, 5000\}$ and for the parameter $c > 0$;

(ii) by using the Nesterov rule and the Chambolle-Dossal rule for the momentum parameters;

(iii) by using different values for the starting point x^0.

Run the two algorithms with step size $\gamma = \frac{1}{L_{\nabla f}}$ for 200 iterations and display $(\|x^k\|, k = 0, 1, 2, ..., 200)$ and $(f(x^k), k = 0, 1, 2, ..., 200)$ as functions of the number of iterations k. (4 points)

34. Implement the CG algorithm for linear systems (Algorithm 8.2). Use as input data the symmetric and positive definite matrix A, the vector b, the starting vector x^0 and the parameter for the stopping criterion ε. The solution x^* and the number of performed iterations should be returned.

Test the algorithm on the following optimization problems and input data values:

(a) Minimize the function

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

over $\mathbb{R}^2$, for $x^0 = (5, -5)^T$ and $\varepsilon = 10^{-3}$.

(b) Minimize the function

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

over $\mathbb{R}^3$, for $x^0 = (0, 0, 0)^T$ and $\varepsilon = 10^{-8}$.

(4 points)