FILOMAT (Niš) 7 (1993), 97–104

A MODIFICATION OF AN OPTIMIZATION METHOD

Nada I. Đuranović-Miličić

Abstract. *A modification of the Danilin-Pschenichnij method is presented. Under certain assumptions it is proved that the obtained sequence of points converges to a unique optimal solution to the given problem of unconstrained optimazation and that the rate of convergence is superlinear.*

1. Introduction

In this paper we are concerned with the problem of unconstrained optimization

$$\min\{\varphi(x) \mid x \in R^n\}, \tag{1}$$

where $\varphi : R^n \to R$ is a twice continuously differentiable function.

The aim of this paper is to present a modification of the Danilin-Pschenichnij method where the direction vectors are defined by the Danilin-Pschenichnij method (see [1] and [4] and the step α_k is defined by a new step-size algorithm.

2. Preliminaries

We begin first with a few preliminaries.

Suppose a sequence of points $\{x_k\}$ is given. We correspond to $\{x_k\}$ another sequence $\{y_k\}$ in this way:

$$y_k = x_k + r_k, \tag{2}$$

where the vectors $r_k \in R^n$ satisfy the following conditions:

1. If by D_k we denote the matrix whose columns are the vectors $r_k/||r_k||, \dots,$ $r_{k-n+1}/||r_{k-n+1}||$, then $|det D_k| \geq \epsilon$ for all $k \geq n-1$, where ϵ is some predetermined small positive number.

2. The vectors r_k are chosen arbitarily except for $||r_k|| \to 0$, $k \to \infty$.

We define the matrix A_k by the following system of equations:

$$A_k r_{k-i} = e_{k-i}, \quad i = 0, 1, \dots, n-1; \quad k \geq n-1, \tag{3}$$

Received 18.10.1993
1991 *Mathematics Subject Classification*: 90C30
Supported by Grant 0401A of RFNS through Math. Inst. SANU

where $e_{k-i} = \nabla\varphi(y_{k-i}) - \nabla\varphi(x_{k-i})$ and r_k, y_k are elements of the sequence (2). The Danilin-Pschenichnij method generates the sequence of points $\{x_k\}$ of the from:

$$x_{k+1} = x_k - \alpha_k p_k, \quad k = 0, 1, \dots,$$

where $p_k = A_k^{-1}\nabla\varphi(x_k)$ for $k \geq n-1$, and the matrix A_k is defined by (3) (for $k = 0, 1, \dots, n-2$ we can put, for example $p_k = \nabla\varphi(x_k)$).

The step-size α_k is computed by the following algorithm:

Step 1. Put $\alpha = 1$ and compute $x = x_k - \alpha p_k$.

Step 2. Compute $\varphi(x) = \varphi(x_k - \alpha p_k)$.

Step 3. Check the inequality

$$\varphi(x_k) - \varphi(x) \geq \epsilon\alpha\langle\nabla\varphi(x_k), p_k\rangle, \quad 0 < \epsilon < \frac{1}{2}.$$

Step 4. If the inequality (4) is statisfied, set $\alpha_k = 1$ $(= \alpha)$; otherwise reduce α until the condition (4) is satisfied.

In the literature there already exist the particular cases of this method obtained for special choices of the vector r_k given by Bulanij A.P. and Danilin J.M. (see [1] and [2]).

We will also need the following lemmas.

Lemma 1. (see [4]) *Let $\{x_k\}$ be a bounded sequence of points, $\|x_{k+1} - x_k\| \to 0$, $k \to \infty$ and let for every $k \geq n-1$ the matrix A_k be defined by (3). Then $\lim_{k\to\infty} \|A_k - H(x_k)\| = 0$, where by $H(x)$ we denote the Hessian of the function φ at the point x.*

Lemma 2. (see [4]) *Let $\varphi : R^n \to R$ be a twice continuously differentiable function such that there exist constants m and M, $0 < M < \infty$ with the property*

$$m\|y\|^2 \leq \langle y, H(x)y\rangle \leq M\|y\|^2 \quad \text{for any} \quad x, y \in R^n \tag{5}$$

and let $\{x_k\}$ be a sequence such that $\varphi(x_{k+1}) \leq \varphi(x_k)$ and $\langle\nabla\varphi(x_k), x_{k+1} - x_k\rangle \to 0$, $k \to \infty$. Then $\|x_{k+1} - x_k\| \to 0$, $k \to \infty$.

Lemma 3. (see [3]) *Let $\varphi : R^n \to R$ be a twice continuously differentiable function such that there exists a constant m, $0 < m < \infty$ with the property*

$$\langle y, H(x)y\rangle \geq m\|y\|^2 \quad \text{for any} \quad x, y \in R^n,$$

where $H(x)$ is the Hessian of the function φ at the point x. Then for any point $x_0 \in R^n$ the set $L = \{x \in R^n | \varphi(x) \leq \varphi(x_0)\}$ is convex and bounded and

$$m\|y\|^2 \leq \langle y, H(x)y\rangle \leq M\|y\|^2 \quad \text{for any} \quad x \in L, y \in R^n, 0 < m < M < \infty.$$

Proposition 1. (see [4]) *Let the assumptions of Lemma 2 be valid and the matrix* A_k *be defined by (3). Also suppose the inequality*

$$\langle A_k^{-1}\nabla\varphi(x_k), \nabla\varphi(x_k)\rangle > 0 \tag{6}$$

holds and the step α_k *is computed by (4). Then, indepedent of the initial point* x_0 *the sequence* $\{x_k\}$ *defined by the iteration*

$$x_{k+1} = x_k - \alpha_k A_k^{-1}\nabla\varphi(x_k), \quad \alpha_k > 0$$

has the following properties: 1.$\varphi(x_{k+1}) < \varphi(x_k)$*; 2.* $\|x_k - x^*\| \to 0$, $k \to \infty$*; 3. the rate of convergence of the sequence* $\{x_k\}$ *to the point* x^* *is superlinear.*

3. The modification of the Danilin-Pschenichnij method

Now we can present the modification of the Danilinp-Pschenichnij method. The modified Danilinp-Pschenichnij algorithm is an iterative algorithm for finding an optimal solution to the problem (1) generating sequences of points $\{x_k\}$ of the form:

$$x_{k+1} = x_k - \alpha_k p_k, \quad k = 0, 1, 2, \dots, \tag{7}$$

$$p_k = A_k^{-1}\nabla\varphi(x_k), \quad k \geq n-1, \tag{8}$$

where the matrix A_k is defined by (3) (for $k = 0, 1, \dots, n-2$ we can for example put $p_k = \nabla\varphi(x_k)$).

We define the step α_k by the following step-size algorithm:

Step 1. Put $\alpha = \overline{\alpha}_k = \min\{\frac{d(\langle\nabla\varphi(x_k), p_k\rangle)}{\|p_k\|^3}, 1\}$, where $d : [0, \infty) \to [0, \infty)$ is a function such that $\delta_1 t \leq d(t) \leq \delta_2 t$ for any $t \geq 0$ and some $0 < \delta_1 < \delta_2$.

Step 2. If the inequality

$$\varphi(x_k) - \varphi(x_k - \alpha p_k) \geq \epsilon\alpha^2 d(\langle\nabla\varphi(x_k), p_k\rangle), \quad 0 < \epsilon < \frac{1}{2} \tag{9}$$

is satisfied for $\alpha = \overline{\alpha}_k$, set $\alpha_k = \overline{\alpha}_k$; otherwise reduce $\overline{\alpha}_k$ until the condition (9) is satisfied.

Theorem. *Let* $\varphi : R^n \to R$ *be a twice continuously differentiable function with the property*

$$\langle H(x)y, y\rangle \geq m\|y\|^2 \quad \textit{for any} \quad x, y \in R^n, \quad 0 < m < \infty, \tag{10}$$

let the sequence $\{x_k\}$ *be defined by the relations (7) and (8). If the condition (6) is satisfied and the step* α_k *is defined by the above step-size algorithm, then: 1.* $x_k \in L = \{x \in R^n | \varphi(x) \leq \varphi(x_0)\}$*; 2.* $\varphi(x_{k+1} < \varphi(x_k)$*; 3.* $\|x_k - x^*\| \to 0$, $k \to \infty$, *where* x^* *is the inique optimal solution to the problem (1).*

If, in addition to the above conditions $\delta_1 < \delta_2 \leq \frac{2^{\delta_1}}{2\epsilon\delta_1 + 1}$, $0 < \delta_1 < \frac{1}{2\epsilon}$, *then the rate of convergence is superlinear.*

Proof. Firstly we shall prove that the sequence $\{x_k\}$ satisfies the assumptions of Lemma 1. We proceede by induction and assume that $x_k \in L$ and $\nabla\varphi(x_k) \neq 0$.

From the condition (6) it follows that

$$\langle\nabla\varphi(x_k), p_k\rangle = \langle\nabla\varphi(x_k), A_k^{-1}\nabla\varphi(x_k)\rangle > 0 \quad \text{for any} \quad k \geq n-1 \tag{11}$$

By the Taylor's theorem, we have for $\alpha > 0$:

$$\varphi(x_k) - \varphi(x_k - \alpha p_k) = \alpha\langle\nabla\varphi(x_k), p_k\rangle - \frac{\alpha^2}{2}\langle H(\xi_k)p_k, p_k\rangle =$$

$(\xi_k = x_k - \theta_k \alpha p_k, \quad \theta_k \in (0,1))$

$$\alpha d(\langle\nabla\varphi(x_k), p_k\rangle)\left[\frac{\langle\nabla\varphi(x_k), p_k\rangle}{d(\langle\nabla\varphi(x_k), p_k\rangle)} - \frac{\alpha}{2}\frac{\langle H(\xi_k)p_k, p_k\rangle}{d(\langle\nabla\varphi(x_k), p_k\rangle)}\right].$$

From (11) it follows that $d(\langle\nabla\varphi(x_k), p_k\rangle) > 0$. Consequently, the inequality (9) holds if

$$\frac{\langle\nabla\varphi(x_k), p_k\rangle}{d(\langle\nabla\varphi(x_k), p_k\rangle)} - \frac{\alpha}{2}\frac{\langle H(\xi_k)p_k, p_k\rangle}{d(\langle\nabla\varphi(x_k), p_k\rangle)} \geq \alpha\epsilon$$

that is

$$\frac{1}{\alpha}\frac{\langle\nabla\varphi(x_k), p_k\rangle}{d(\langle\nabla\varphi(x_k), p_k\rangle)} - \frac{1}{2}\frac{\langle H(\xi_k)p_k, p_k\rangle}{d(\langle\nabla\varphi(x_k), p_k\rangle)} \geq \epsilon. \tag{12}$$

Since $\frac{\langle\nabla\varphi(x_k), p_k\rangle}{d(\langle\nabla\varphi(x_k), p_k\rangle)} > 0$ and by the inequality (10) $\frac{\langle H(\xi_k)p_k, p_k\rangle}{d(\langle\nabla\varphi(x_k), p_k\rangle)} > 0$, for some $\alpha = \alpha_k$ the inequality (12) will hold, and consequently the condition (9).

From the relation (9) it follows that $\varphi(x_{k+1}) < \varphi(x_k)$ and $x_{k+1} \in L = \{x \in R^n | \varphi(x) \leq \varphi(x_0)\}$, where L (by Lemma 3) is a compact and convex set. Since the function φ is bounded below on the compact and convex set L, it follows that $\varphi(x_{k+1}) - \varphi(x_k) \to 0$, $k \to \infty$. Hence by (9) it follows that

$$\alpha_k^2 d(\langle\nabla\varphi(x_k), p_k\rangle) \to 0, \quad k \to \infty. \tag{13}$$

Furtheon, since $\alpha_k \leq \overline{\alpha}_k \leq \frac{d(\langle\nabla\varphi(x_k), p_k\rangle)}{\|p_k\|^3}$ we have that $d(\langle\nabla\varphi(x_k), p_k\rangle) \geq \alpha_k\|p_k\|^3$.

From this inequality using (13) we obtain

$$\alpha_k^3\|p_k\|^3 \leq \alpha_k^2 d(\langle\nabla\varphi(x_k), p_k\rangle) \to 0, \quad k \to \infty. \tag{14}$$

From (14) if follows that $||x_{k+1} - x_k|| = \alpha_k ||p_k|| \to 0$, $k \to \infty$. Consequently, all assumptions of Lemma 1 are satisfied. Hence, we have

$$||A_k - H(x_k)|| \to o, \quad k \to \infty. \tag{15}$$

From (15) and Lemma 3, if follows that for any M_1 such that $M_1 \geq M$ and m satisfying $0 < m_1 \leq m$ there exist an integer k_0 such that for $k \geq k_0$ and any $y \in R^n$

$$m_1 ||y|| \leq \langle A_k y, y \rangle \leq M_1 ||y||^2 \tag{16}$$

holds.

This relation implies

$$m_1 ||p_k||^2 \leq \langle A_k p_k, p_k \rangle = \langle \nabla\varphi(x_k), p_k \rangle \leq M_1 ||p_k||^2, \quad k \geq k_0. \tag{17}$$

By (16) we have that for sufficiently large k

$$\frac{d(\langle \nabla\varphi(x_k), p_k \rangle)}{||p_k||^3} \geq \frac{\delta_1 \langle \nabla\varphi(x_k), p_k \rangle}{||p_k||^3} \geq \frac{\delta_1 m_1}{||p_k||}. \tag{18}$$

Since $\nabla\varphi$ is bounded on L, applying (16) we get

$$||p_k|| = ||A_k^{-1} \nabla\varphi(x_k)|| \leq \frac{1}{m_1} K \quad \text{for any} \quad k \geq n-1.$$

From (17) and (18) we obtain

$$\frac{d(\langle \nabla\varphi(x_k), p_k \rangle)}{||p_k||^3} \geq \frac{\delta_1 m_1^2}{K} = \overline{\alpha} > 0.$$

Since $\overline{\alpha}_k = \min\{\frac{d(\langle \nabla\varphi(x_k), p_k \rangle)}{||p_k||^3}, 1\}$, we have that for sufficiently large k $\overline{\alpha}_k \geq \overline{\alpha} > 0$. Furtheon, since, $\delta_1 t \leq d(t) \leq \delta_2 t$, by Lemma 3

$$m||y||^2 \leq \langle H(x)y, y \rangle \leq M||y||^2 \quad \text{for any} \quad x \in L, y \in R^n$$

It is evident that the condition (12) is satisfied if the following is valid:

$$\begin{aligned} &\frac{1}{\alpha} \frac{\langle \nabla\varphi(x_k), p_k \rangle}{d(\langle \nabla\varphi(x_k), p_k \rangle)} - \frac{1}{2} \frac{\langle H(\xi_k) p_k, p_k \rangle}{d(\langle \nabla\varphi(x_k), p_k \rangle)} \geq \\ &\frac{1}{\alpha} \frac{\langle \nabla\varphi(x_k), p_k \rangle}{\delta_2 \langle \nabla\varphi(x_k), p_k \rangle} - \frac{1}{2} \frac{M ||p_k||^2}{\delta_1 \langle \nabla\varphi(x_k), p_k \rangle} \geq \\ &\geq \frac{1}{\alpha\delta_2} - \frac{M}{2\delta_1} \frac{||p_k||^2}{m_1 ||p_k||^2} = \frac{1}{\alpha\delta_2} - \frac{M}{2m_1\delta_1} \geq \epsilon. \end{aligned} \tag{19}$$

From (19) it follows that

$$0<\alpha\le\frac{2\delta_1 m_1}{\delta_2(2\delta_1 m_1\epsilon+M)}. \tag{20}$$

This relation implies that there exists a constant α^*, $0<\alpha^*<\dfrac{2\delta_1 m_1}{\delta_2(2\delta_1 m_1\epsilon+M)}$ such that the inequality (19) is satisfied for any k if $\alpha^*\le\alpha_k\le\dfrac{2\delta_1 m_1}{\delta_2(2\delta_1 m_1\epsilon+M)}$.

Now, since $\alpha_k\ge\alpha^*$ and $\overline{\alpha}_k\ge\overline{\alpha}$ for sufficiently large k it follows that there exist a constant $C>0$, such that $\alpha_k\ge C>0$. Hence,

$$||p_k||=\frac{1}{\alpha_k}||x_{k+1}-x_k||\le\frac{||x_{k+1}-x_k||}{C}\to 0,\quad k\to\infty. \tag{21}$$

From (161) and (21) we get

$$||\nabla\varphi(x_k)||=||A_k p_k||\le M_1||p_k||\to 0,\quad k\to\infty. \tag{22}$$

Since the function φ is stricly convex and $\{x_k\}\subset L$, where L is a copmact convex set, applying (22) we have that the sequence $\{x_k\}$ converges to a unique optimal solution x^* to the problem (1).

Now we shall show that $\alpha_k=1$ for sufficiently large k. The relations (18) and (21) imply the following:

$$\frac{d(\langle\nabla\varphi(x_k),p_k\rangle)}{||p_k||^3}\ge\frac{\delta_1 m_1}{||p_k||^3}\to\infty\quad k\to\infty.$$

From this relation and the definition of $\overline{\alpha}_k$ it follows that, for sufficiently large k, $\overline{\alpha}_k=1$.

Furtheon, since the step-size α_k must satisfy the condition (12), by the assuptions of Theorem we have

$$\begin{aligned}&\frac{1}{\alpha_k}\frac{\langle\nabla\varphi(x_k),p_k\rangle}{d(\langle\nabla\varphi(x_k),p_k\rangle)}-\frac{1}{2}\frac{\langle H(\xi_k)p_k,p_k\rangle}{d(\langle\nabla\varphi(x_k),p_k\rangle)}\ge\cdots\\&\ge\frac{1}{\alpha_k\delta_2}-\frac{1}{2\delta_1}\frac{\langle H(\xi_k)p_k,p_k\rangle}{\langle\nabla\varphi(x_k),p_k\rangle}\ge\epsilon,\end{aligned} \tag{23}$$

$\xi_k=x_k+\theta_k(x_{k+1}-x_k)$, $\theta_k\in(0,1)$.

Since

$$\frac{\langle H(\xi_k)p_k,p_k\rangle}{\langle\nabla\varphi(x_k),p_k\rangle}=\frac{\langle[H(\xi_k)-H(x_k)]p_k,p_k\rangle}{\langle A_k p_k,p_k\rangle}+\frac{\langle H(x_k)p_k,p_k\rangle}{\langle A_k p_k,p_k\rangle},$$

by (15),(21) and uniform continuity of H on L, if $k\to\infty$ then $\dfrac{\langle H(\xi_k)p_k,p_k\rangle}{\langle\nabla\varphi(x_k),p_k\rangle}\to 1$.

Consequently, by (23), for sufficiently large k we obtain:

$$\frac{1}{\alpha_k}\frac{\langle \nabla\varphi(x_k), p_k\rangle}{d(\langle \nabla\varphi(x_k), p_k\rangle)} - \frac{1}{2}\frac{\langle H(\xi_k)p_k, p_k\rangle}{d(\langle \nabla\varphi(x_k), p_k\rangle)} \geq \cdots$$

$$\geq \frac{1}{\alpha_k\delta_2} - \frac{1}{2\delta_1} \geq \epsilon.$$

It is easy to show that the last inequality is satisfied if $\delta_1 < \delta_2 \leq \dfrac{2\delta_1}{1+2\epsilon\delta_1}$, $0 < \delta_1 < \frac{1}{2\epsilon}$, $\alpha_k = 1$.

It means that for sufficiently large k the inequality (9) i also satisfied for $\alpha_k = 1$. Now we can estimate the rate of convergence. We have:

$$\|x_{k+1} - x_k\|^2 = \langle x_{k+1} - x^*, x_{k+1} - x^*\rangle = \langle x_k - x^* - A_k^{-1}\nabla\varphi(x_k), x_{k+1} - x^*\rangle.$$

By the Mean Value Theorem we obtain:

$$\langle A_k^{-1}\nabla\varphi(x_k), x_{k+1} - x^*\rangle = \langle A_k^{-1}[\nabla\varphi(x_k) - \nabla\varphi(x^*)], x_{k+1} - x^*\rangle =$$

$$\langle A_k^{-1}H(\eta_k)(x_k - x^*), x_{k+1} - x^*\rangle, \quad \eta_k = x_k + \nu_k(x_k - x^*), \nu_k \in (0,1).$$

The last two inequalities imply

$$\begin{aligned}\|x_{k+1} - x^*\|^2 &= \langle (I - A_k^{-1}H(\eta_k)(x_k - x^*), (x_{k+1} - x^*)\rangle = \\ &= \langle (A_k^{-1}(A_k - H(\eta_k))(x_k - x^*), (x_{k+1} - x^*)\rangle \leq \\ &\leq \|A_k^{-1}\|\|(A_k - H(\eta_k)\|\|x_k - x^*\|\|x_{k+1} - x^*\| \leq \\ &\leq \frac{1}{m_1}\|(A_k - H(\eta_k)\|\|x_k - x^*\|\|x_{k+1} - x^*\|.\end{aligned}$$

Hence, $\|x_{k+1} - x^*\| \leq \gamma_k\|x_k - x^*\|$, where $\gamma_k = \dfrac{1}{m_1}\|A_k - H(\eta_k)\|$.

By Lemma 1 and uniform continuity of H on L it follows that

$$\|A_k - H(\eta_k)\| \leq \|A_k - H(x_k)\| + \|H(x_k) - H(\eta_k)\| \to 0 \quad k \to \infty.$$

Consequently, $\gamma_k = \dfrac{1}{m_1}\|A_k - H(\eta_k)\| \to 0$, $k \to \infty$.

The last relation implies that the sequence $\{x_k\}$ converges to a unique optimal solution x^* superlinearly.

REFERENCES

[1] A.P. Bulanij and J.M. Danilin, *Kvazinjutonovskije algoritmi minimizacii, osnovanie na postroenii sistem soprjažonnih vektorov*, ŽVM i MF 18, (1978), 877–885.

[2] J.M. Danilin, *Metodi soprjažonnih napravljenij ne trebujuščije rešenija odnomernih zadač minimizacii*, DAN SSSR 218, (1974), 513–516.

[3] E. Polak, *Computational Methods in Optimization, a Unified Approach*, Academic Press, New York, 1971.

[4] B.N. Pšeničnij and J.M. Danilin, *Čislennie metodi v ekstremaljnih zadačah*, Nauka, Moskva 1975.

University of Belgrade
Faculty of Technology and Metallurgy
Department of Mathematics
Karnegijeva 4
11000 Belgrade
Yugoslavia