Yugoslav Journal of Operations Research
12 (2002), Number 1, 121-127

ON A SECOND-ORDER STEP-SIZE ALGORITHM*

Nada I. DJURANOVIĆ-MILIČIĆ

Department of Mathematics
Faculty of Technology and Metallurgy
University of Belgrade, Belgrade, Yugoslavia
nmilicic@elab.tmf.bg.ac.yu

Abstract: In this paper we present a modification of the second-order step-size algorithm. This modification is based on the so called forcing functions. It is proved that this modified algorithm is well-defined. It is also proved that every point of accumulation of the sequence generated by this algorithm is a second-order point of the nonlinear programming problem. Two different convergence proofs are given having in mind two interpretations of the presented algorithm.

Keywords: Forcing function, step-size algorithm, second-order conditions.

1. INTRODUCTION

We are concerned with the following problem of the unconstrained optimization:

$$\min\{\varphi(x) \mid x \in D\} \tag{1}$$

where $\varphi : D \subset R^n \to R$ is a twicecontinuously differentiable function on an open set D.

We consider iterative algorithms to find an optimal solution to problem (1) generating sequences of points $\{x_k\}$ of the following form:

$$x_{k+1} = x_k + \alpha_k s_k + \beta_k d_k, \quad k = 0,1,\dots, \tag{2}$$

$$s_k, d_k \neq 0, \quad \langle \nabla\varphi(x_k), s_k \rangle \leq 0, \tag{3}$$

and the steps α_k and β_k are defined by a particular step-size algorithm.

* This research was supported by Science Fund of Serbia, grant number 04M03, through Institute of Mathematics, SANU. AMS Mathematics Subject Classification (1991): 90C30

Before we present the modified algorithm, we shall define the original second-order step-size algorithm.

The original Mc Cormick-Armijo's second order step-size algorithm [4] defines α_k in the following way: $\alpha_k > 0$ is a number satisfying

$$\alpha_k = 2^{-i(k)} ,$$

where $i(k)$ is the smallest integer from $i = 0,1,...$, such that

$$x_{k+1} = x_k + 2^{-i(k)} s_k + 2^{\frac{-i(k)}{2}} d_k \in D$$

and

$$\varphi(x_k) - \varphi(x_{k+1}) \geq \gamma \left[-\langle \nabla \varphi(x_k), s_k \rangle - \frac{1}{2} \langle H(x_k) d_k, d_k \rangle \right] 2^{-i(k)} ,$$

where $0 < \gamma < 1$ is a preassigned constant, $H(x)$ - the Hessian matrix of the function φ at x, s_k, d_k -direction vectors satisfying relations (3).

We begin with the definition which we need in the following text.

Definition (See[**5**]). *A mapping* $\sigma : [0,\infty) \to [0,\infty)$ *is a forcing function if for any sequence* $\{t_k\} \subset [0,\infty)$

$$\lim_{k\to\infty} \sigma(t_k) = 0 \quad \text{implies} \quad \lim_{k\to\infty} t_k = 0$$

and $\sigma(t) > 0$ *for* $t > 0$.

(The concept of the forcing function was introduced first by Elkin in [3].)

2. A MODIFICATION OF THE SECOND-ORDER STEP-SIZE ALGORITHM

The modified algorithm defines α_k in the following way: $\alpha_k > 0$ is a number satisfying

$$\alpha_k = q^{-i(k)}, \quad q > 1,$$

where $i(k)$ is the smallest integer from $i = 0,1,...$, such that

$$x_{k+1} = x_k + q^{-i(k)} s_k + q^{\frac{-i(k)}{2}} d_k \in D \tag{4}$$

and

$$\varphi(x_k) - \varphi(x_{k+1}) \geq q^{-i(k)} \left[\sigma_1(-\langle \nabla \varphi(x_k), s_k \rangle) + \sigma_2(-\frac{1}{2} \langle H(x_k) d_k, d_k \rangle) \right] \tag{5}$$

where $\sigma_1 : [0,\infty) \to [0,\infty)$ and $\sigma_2 : [0,\infty) \to [0,\infty)$ are the forcing functions such that $\delta_1 t \le \sigma_1(t) \le \bar{\delta}_1 t$, $\delta_2 t \le \sigma_2(t) \le \bar{\delta}_2 t$ $\quad 0 < \delta_1 < \bar{\delta}_1 < 1$, $0 < \delta_2 < \bar{\delta}_2 < 1$ and s_k, d_k are the direction vectors satisfying (3) and $\langle H(x_k)d_k, d_k \rangle \le 0$.

In order to have a finite value $i(k)$, it is sufficient that s_k and d_k satisfy (3) and, in addition, that

$$\langle \nabla\varphi(x_k), s_k \rangle < 0 \quad \text{whenever} \quad \nabla\varphi(x_k) \ne 0 \tag{6A}$$

and

$$\langle H(x_k)d_k, d_k \rangle < 0 \quad \text{whenever} \quad \nabla\varphi(x_k) = 0 . \tag{6B}$$

Now we shall prove the first convergence theorem.

Theorem 1. *Let $\varphi : D \subset R^n \to R$ be a twicecontinuously differentiable function on the open set D. Let the sequence $\{x_k\}$ be defined by relations (2), (3), (4),(5),(6A) and (6B). Let $\bar{x}$ be a point of accumulation of $\{x_k\}$ and K_1 a set of indices such that $x_k \to \bar{x}$ for $k \in K_1$.*

Assume that:

1. the sequences $\{s_k\}$ and $\{d_k\}$ are uniformly bounded;
2. $-\langle \nabla\varphi(x_k), s_k \rangle \ge \mu_k(\| \nabla\varphi(x_k) \|)$, $k \in K_1$, where $\mu_k : [0,\infty) \to [0,\infty)$, $k \in K_1$ are forcing functions;
3. there exists a value $\beta > 0$ such that

$$-\langle H(x_k)d_k, d_k \rangle \ge \beta \langle H(x_k)e_k^{\min}, e_k^{\min} \rangle,$$

where $e_k^{\min}$ is an eigenvector of $H(x_k)$ associated with its minimum eigenvalue.

Then $\bar{x}$ is a stationary point, that is

$$\nabla\varphi(\bar{x}) = 0$$

and $H(\bar{x})$ is a positive semidefinite matrix with at least one eigenvalue equal to zero.

Proof: There are two cases to consider.

a) The integers $\{i(k)\}$ for $k \in K_1$ are uniformly bounded from above by some value I.

Because of the descent property it follows that all points of the accumulation have the same function value and

$$(0 \ge)\varphi(x_0) - \varphi(\bar{x}) = \sum_{k \in K_1} [\varphi(x_k) - \varphi(x_{k+1})] \ge$$

$$\ge \sum_{k \in K_1} q^{-i(k)} \left[\sigma_1(-\langle \nabla\varphi(x_k), s_k \rangle) + \sigma_2\left(-\frac{1}{2} \langle H(x_k)d_k, d_k \rangle \right) \right] \ge$$

$$\ge q^{-I}\delta \sum_{k \in K_1} \left[-\langle \nabla\varphi(x_k), s_k \rangle - \frac{1}{2} \langle H(x_k)d_k, d_k \rangle \right], \quad (\delta = \max\{\delta_1, \delta_2\})$$

$$\geq q^{-I}\delta \sum_{k\in K_1}\left[\mu_k(\|\nabla\varphi(x_k)\|)+\frac{1}{2}\beta\left\langle H(x_k)e_k^{\min},e_k^{\min}\right\rangle\right].$$

Since $\varphi(\overline{x})$ is finite and since each term in the brackets is greater than, or equal to zero for each $k\in K_1$, it follows that $\mu_k(\nabla\varphi(x_k))\to 0 \Rightarrow \|\nabla\varphi(x_k)\|\to 0$ (according to the definition of forcing functions) $\Rightarrow \nabla\varphi(\overline{x})=0$ and that $\langle H(\overline{x})\overline{e}_{\min},\overline{e}_{\min}\rangle=0$, where $\overline{e}_{\min}$ is some accumulation point of $\{e_k^{\min}\}$ for $k\in K_1$.

b) There is a subset $K_2\subset K_1$ such that $\lim\limits_{k\to\infty} i(k)=\infty$.

Because of the definition of $i(k)$, then either

$$x_k+q^{-i(k)+1}s_k+q^{\frac{-i(k)+1}{2}}d_k\notin D$$

or

$$\varphi(x_k)-\varphi\left(x_k+q^{-i(k)+1}s_k+q^{\frac{-i(k)+1}{2}}d_k\right)< \tag{7}$$

$$<q^{-i(k)+1}\left[\sigma_1(-\langle\nabla\varphi(x_k),s_k\rangle)+\sigma_2\left(-\frac{1}{2}\langle H(x_k)d_k,d_k\rangle\right)\right].$$

If the former condition held infinitely often, then because

$$x_k+q^{-i(k)+1}s_k+q^{\frac{-i(k)+1}{2}}d_k\to\overline{x},\quad k\in K_2,$$

it would follow that $\overline{x}$ is on the boundary of D. Since D is an open set, $\overline{x}\notin D$, it contradicts the theorem hypothesis. Therefore, without the loss of generality (7) can be considered to hold for all $k\in K_2$.

Since $\varphi\in C^2$, and since the sequences $\{s_k\}$ and $\{d_k\}$ are assumed to be uniformly bounded, the left -hand side of inequality (7) can be written as

$$-q^{-i(k)+1}\langle\nabla\varphi(x_k),s_k\rangle-q^{\frac{-i(k)+1}{2}}\langle\nabla\varphi(x_k),d_k\rangle-$$

$$-\frac{1}{2}\left\langle H(x_k)\left(q^{-i(k)+1}s_k+q^{\frac{-i(k)+1}{2}}d_k\right),q^{-i(k)+1}s_k+q^{\frac{-i(k)+1}{2}}d_k\right\rangle-o(q^{-i(k)+1})<$$

$$<q^{-i(k)+1}\left[\sigma_1(-\langle\nabla\varphi(x_k),s_k\rangle)+\sigma_2\left(-\frac{1}{2}\right)\langle H(x_k)d_k,d_k\rangle\right]<$$

$$<q^{-i(k)+1}\left[-\overline{\delta}_1\langle\nabla\varphi(x_k),s_k\rangle-\overline{\delta}_2\cdot\frac{1}{2}\langle H(x_k)d_k,d_k\rangle\right].$$

Combining terms and incorporating a term where appropriate into $o(q^{-i(k)+1})$ yields (using the fact that $-\langle\nabla\varphi(x_k),s_k\rangle\geq 0$):

$$o(q^{-i(k)+1}) > q^{-i(k)+1}\left[(-1+\overline{\delta}_1)\langle\nabla\varphi(x_k),s_k\rangle-(-\overline{\delta}_2+1)\frac{1}{2}\langle H(x_k)d_k,d_k\rangle\right].$$

Using the theorem hypothesis 3 we obtain

$$o(q^{-i(k)+1}) > q^{-i(k)+1}\left[(-1+\overline{\delta}_1)\langle\nabla\varphi(x_k),s_k\rangle+(-\overline{\delta}_2+1)\frac{\beta}{2}\left\langle H(x_k)e_k^{\min},e_k^{\min}\right\rangle\right].$$

Dividing by $q^{-i(k)+1}$ yields

$$\frac{o(q^{-i(k)+1})}{q^{-i(k)+1}} > (-1+\overline{\delta}_1)\langle\nabla\varphi(x_k),s_k\rangle+(-\overline{\delta}_2+1)\frac{\beta}{2}\left\langle H(x_k)e_k^{\min},e_k^{\min}\right\rangle \geq$$

$$\geq (1-\overline{\delta}_1)\mu_k(\|\nabla\varphi(x_k)\|)+\frac{-\overline{\delta}_2+1}{2}\cdot\beta\cdot\left\langle H(x_k)e_k^{\min},e_k^{\min}\right\rangle.$$

Since each term is, according to the assumptions, greater than or equal to zero, taking the limit as $k\to\infty$ for $k\in K_2$ yields

$$\mu_k(\|\nabla\varphi(x_k)\|)\to 0 \Rightarrow \|\nabla\varphi(x_k)\|\to 0 \Rightarrow \nabla\varphi(\overline{x})=0$$

and

$$\left\langle H(x_k)e_k^{\min},e_k^{\min}\right\rangle \to \langle H(\overline{x})\overline{e}_{\min},\overline{e}_{\min}\rangle = 0.$$

To prove the second convergence theorem we shall follow Y. Amaya [1]. Namely, we are going to show that the trajectory

$$f(t,x_k)=x_k+t^2s_k+t\,d_k \tag{8}$$

proposed by the presented algorithm (i.e. satisfying the relations (2), (3), (4), (5), (6A) and (6B)) and

$$\begin{aligned}&\langle\nabla\varphi(x_k),s_k\rangle<0\\&\langle\nabla\varphi(x_k),d_k\rangle\leq 0\end{aligned} \tag{9}$$

and

$$\langle H(x_k)d_k,d_k\rangle=0$$

if $H(x_k)$ is positive semidefinite, and

$$\begin{aligned}&\langle\nabla\varphi(x_k),s_k\rangle\leq 0\\&\langle\nabla\varphi(x_k),d_k\rangle\leq 0\end{aligned} \tag{10}$$

and

$$\langle H(x_k)d_k,d_k\rangle<0$$

if $H(x_k)$ is not positive semidefinite, has the properties set out in Amaya's paper.

Firstly, we shall briefly present Amaya's algorithm [1].

Let $\varphi : D \subset R^n \to R$ be a twicecontinuously differentiable function on the open set D (i.e. $\varphi \in C^2$) which we want to minimize, and $h : R^+ \times D \to R^n$ is a function such that, for all $x \in D, h(0,x) = x$. We suppose that for every $x \in D, h(t,x)$ is C^2 for $t \geq 0$.

Given $x \in D$, the function $h(t,k)$ describes a trajectory in $D \subset R^n$ originating at x. The minimizing algorithm defines a sequence $\{x_k\}$ in the following way:

$$x_{k+1} = \begin{cases} x_k & \text{if} \quad x_k \in M, \\ h(t_k, x_k) & \text{if} \quad x_k \notin M, \end{cases} \tag{11}$$

where $M = \{x \in D \mid \nabla\varphi(x) = 0 \text{ and } \langle H(x)p, p\rangle \geq 0,\ p \in R^n\}$.

For $x \in D$, we define the C^2 - class function $f_x : R^+ \to R^n$ by

$$f_x(t) = \varphi[h(t,x)],\ t \in R^+.$$

This function is shown to satisfy

$$f'_{x_k}(0) = \left\langle \nabla\varphi(x_k), \dot{h}(0,x_k)\right\rangle \quad \text{and}$$
$$f''_{x_k}(0) = \left\langle H(x_k)\dot{h}(0,x_k), \dot{h}(0,x_k) + \left\langle \nabla\varphi(x_k), \ddot{h}(0,x_k)\right\rangle\right\rangle,$$

where $\dot{h}$ and $\ddot{h}$ denote respectively the first and second derivatives of h with respect to t.

The following assumptions are made:

A1. $L = \{x \in D \mid \varphi(x) \leq \varphi(x_0)\}$ is bounded;

A2. $f'_x(0) \leq 0$ for all $x \notin M$;

A3. if $x \notin M$ and $f'_x(0) = 0$, then $f''_x(0) < 0$.

Amaya in Theorem 3.1 in [1] proves the convergence of a subsequence of points of $\{x_k\}$ defined by (11) to $\bar{x} \in M$, provided that $\varphi \in C^2$ and that assumptions A1, A2, A3 hold.

Now we can present the second convergence theorem for the modified Mc Cormick-Armijo's algoritm.

Theorem 2. *Under assumptions A1, A2 and A3 every point of accumulation* $\bar{x}$ *of the sequence* $\{x_k\}$ *generated by the modified McCormick-Armijo's algorithm and additionally, satisfying (9) and (10) belongs to M, that is, the second-order necessary conditions are satisfied at* $\bar{x}$.

Proof: Let us suppose that $x_k \notin M$ for $k = 0,1,2,\ldots$. From the choice of $t_k = \alpha_k$ by relations (2), (3), (4), (5), (6A) and (6B) we have that $f_{x_k}(t_k) \leq f_{x_k}(0)$, i.e. the sequence $\{\varphi(x_k)\}$ is decreasing; hence $\{x_k\} \subset L$. Due to the assumption A1, the sequence $\{x_k\}$ has a point of accumulation $\bar{x}$.

For the trajectory (8) we have:

$$f_{x_k}^{'}(0)=\left\langle \nabla\varphi(x_k),\dot{h}(0,x_k)\right\rangle,\quad \dot{h}(0,x_k)=d_k,$$

$$f_{x_k}^{''}(0)=\left\langle H(x_k)\dot{h}(0,x_k),\dot{h}(0,x_k)+\left\langle \nabla\varphi(x_k),\ddot{h}(0,x_k)\right\rangle\right\rangle,\quad \ddot{h}(0,x_k)=s_k,\quad \text{i.e.}$$

$$f_{x_k}^{'}(0)=\left\langle \nabla\varphi(x_k),d_k\right\rangle,$$

$$f_{x_k}^{''}(0)=\left\langle H(x_k)d_k,d_k\right\rangle+\left\langle \nabla\varphi(x_k),s_k\right\rangle.$$

From (6A) it follows that the assumption A2 holds. Let us examine the assumption A3. Assuming $f_{x_k}^{'}(0)=0$, we have two cases:

a) if $H(x_k)$ is positive semidefinite, by applying (9) to the relation (11), we obtain
$f_{x_k}^{''}(0)<0;$

b) if $H(x_k)$ is not positive semidefinite, by applying (10) to the relation (11), we obtain
$f_{x_k}^{''}(0)<0.$

Following Amaya's proof of theorem 3.1 in [1] we conclude that $\overline{x}\in M$.

3. CONCLUSION

Because of general assumptions on the objective function φ, the modified algorithm can be used for solving a wide class of unconstrained optimization problems. Also, the choice of forcing functions $\sigma_1(t)$ and $\sigma_2(t)$, with the property $\delta_1 t\le\sigma_1(t)\le\overline{\delta}_1 t,\ \delta_2 t\le\sigma_2(t)\le\overline{\delta}_2 t,\ 0<\delta_1<\overline{\delta}_1<1,\ 0<\delta_2<\overline{\delta}_2<1$ is wide.

Finally, this modified algorithm can be used for solving constrained optimization problems (see [2]) when constraints are adequately considered.

REFERENCES

[1] Amaya, J., "Convergence of curvilinear search algorithms to second order points", *Revista de Matematcas Aplicadas,* 10 (1989) 71-79.

[2] Djuranovic-Miličić, N., "An algorithm in constrained optimization", in: M. Thoma and A. Wyner (eds.), *Lecture Notes in Control and Information Sciences,* Springer-Verlag, Berlin, 1986, 203-208.

[3] Elkin, R., "Convergence theorems for Gauss-Siedel and other minimization algorithms", Doctoral Thesis, University of Maryland, College Park, 1968.

[4] Mc Cormick, G.P., *Nonlinear Programming, Theory, Algorithms and Applications*, Wiley, New York, 1983.

[5] Ortega, J., and Rheinboldt, W., *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York, 1970.