FULL CONVERGENCE OF THE STEEPEST DESCENT METHOD WITH INEXACT LINE SEARCHES

Regina Burachik

Luis Mauricio Graña Drummond

Alfredo N. Iusem*

and

B.F. Svaiter

Instituto de Matemática Pura e Aplicada

Estrada Dona Castorina, 110

Rio de Janeiro, RJ, CEP 22460-320

Brazil

Abstract

Several finite procedures for determining the step size of the steepest descent method for unconstrained optimization, without performing exact onedimensional minimizations, have been considered in the literature. The convergence analysis of these methods requires that the objective function have bounded level sets and that its gradient satisfy a Lipschitz condition, in order to establish just stationarity of all cluster points. We consider two of such procedures and prove, for a convex objective, convergence of the whole sequence to a minimizer without any level set boundedness assumption and, for one of them, without any Lipschitz condition.

Keywords: Convex programming, steepest descent method, inexact line searches.

1991 AMS Classification numbers: 90C25, 90C30.

---

*Research of this author was partially supported by CNPq grant nº 301280/86

§1 Introduction

The steepest descent method (also called Cauchy's method or gradient method) is one of the oldest and simplest procedures for minimization of a real function defined on $\mathbf{R}^n$. It is also the departure point for many other more sophisticated optimization procedures. Despite its simplicity and notoriety (practically no optimization book fails to discuss it) its convergence theory is not fully satisfactory from a thoretical point of view (from a practical point of view it suffers from other disadvantages, but here we are not concerned with this issue). More precisely, standard convergence results (e.g., [**7**], [**8**]), which are based upon Zangwill's global convergence theorem ([**10**]), demand that the initial point belong to a bounded level set of the objective function f (and henceforth that f have at least one bounded level set) and fail to prove full convergence of the sequence generated by the method to a stationary point of f, establishing only that all its cluster points are stationary. Even when f is convex (in which case the stationary points are the global minimizers of f) the assumption of bounded level sets is required, and the result is just what has been called *weak convergence* to the set of minimizers of f (a sequence $\{x^k\}$ is said to be *weakly convergent* to a set S if $\{x^k\}$ is bounded, $x^{k+1} - x^k$ converges to zero and every cluster point of $\{x^k\}$ belongs to S).

It is true that from a computational point of view weak convergence is almost indistinguishable from full convergence, but failure to prove full convergence is theoretically unsatisfactory. On the other hand, the condition of bounded level sets is quite restrictive both theoretically and practically.

Basically, the steepest descent method generates a sequence $\{x^k\}$ where x^{k+1} is taken as $x^k - \lambda_k \nabla f(x^k)$ for some $\lambda_k > 0$. It is customary to distinguish the cases of exact and inexact line searches. An exact line search consists of taking λ_k as a minimizer of f on the halfline $\{x^k - \lambda \nabla f(x^k) / \lambda > 0\}$. When inexact line searches are performed, λ_k is a given predetermined value or is obtained through some finite procedure. Of course, every computational implementation of the algorithm falls in the second cathegory.

The convergence properties of the steepest descent method with inexact line searches have been studied under several strategies for the choice of the stepsize λ_k. Here we will be concerned with two methods for choosing λ_k. The first one requires that the gradient of f satisfy a Lipschitz condition with a known constant L. In this case, Polyak [**8**] has proved that for fixed $\lambda_k = \lambda \in (0, \frac{2}{L})$, the sequence obtained converges weakly to the set of stationary points of f, under a level set boundedness assumption.

The second case of stepsize selection is based on a backtracking procedure studied by Dennis-Schnabel [**3**], which considers the case when L is not known beforehand and proposes a backtracking strategy where succesive values of λ are tried until one is found so that it satisfies two inequalities. This backtracking strategy generates a sequence which, under the assumptions of level set boundedness and Lipschitz condition on $\nabla f(\cdot)$, is proven to be weakly convergent to the set of stationary points of f.

The purpose of this paper is to establish, for f convex and Lipschitzian gradient, full convergence of $\{x^k\}$ to a minimizer of f, with the first strategy, i.e., for fixed λ_k, without any hypothesis on its level sets. For the second case, we present a specific backtracking strategy for which the same results hold without requiring bounded level sets or a Lipschitz condition on the gradient of f.

Another interesting point is that this backtracking strategy finds λ_k using only one inequality instead of the two inequalities required in [**3**].

Our work is related to [**5**], where similar results are established for the steepest descent method with exact line searches, upon addition of a regularization term, reminiscent of the proximal point method, to the objective function of the one dimensional minimization.

The convergence results are based upon the notion of quasi-Fejér convergence, introduced for the first time in [**4**] for sequences of random variables and applied in [**6**] to optimization problems.

In the case of inexact line searches no regularization term is required and full convergence of the sequence is proved, by establishing quasi-Fejér convergence to the set of minimizers of f. In the case of the backtracking procedure, we made a slight generalization of Dennis-Schnabel's method. One of the inequalities in these authors' strategy compares the decrease of the function with a linear function of λ_k. In this work we consider a more general family of functions for checking the decrease of the function.

§2 The Algorithms

Take $f: \mathbf{R}^n \to \mathbf{R}$ convex and continuously differentiable. For the convergence analysis of algorithm A we will assume also that the gradient of f satisfies a Lipschitz condition with constant L, i.e. there exists $L > 0$ such that

$$\|\nabla f(x) - \nabla f(y)\| \leq L \|x - y\| \tag{1}$$

for all $x, y \in \mathbf{R}^n$. We assume from now on that the set T^* of minimizers of f is nonempty. f^* will denote the minimum value of f on $\mathbf{R}^n$.

The general form of the algorithms under consideration is:

$$x^0 \in \mathbf{R}^n \tag{2}$$

$$x^{k+1} = x^k - \lambda_k \nabla f(x^k) \tag{3}$$

where the stepsize $\lambda_k > 0$ is chosen according to one of the following criteria.

Algorithm A (known L). Take δ_1, δ_2 positive numbers such that

$$\frac{L}{2}\delta_1 + \delta_2 < 1 \tag{4}$$

and pick up λ_k satisfying

$$\delta_1 \leq \lambda_k \leq \frac{2}{L}(1 - \delta_2) \tag{5}$$

Algorithm B. Let $\psi: \mathbf{R}_+ \to \mathbf{R}$ (where $\mathbf{R}_+$ stands for the nonnegative real line) such that:

B1) ψ is convex and continuously differentiable, B2) $\psi(0) = 0$ and $\psi'(0) < 1$,

B3) $\liminf_{u \to 0^+} \frac{\psi(u)}{u^2} > 0$.

Notice that B3 implies $\psi'(0) \geq 0$. So using B2 we get $0 \leq \psi'(0) < 1$, and hence, ψ is nondecreasing by B1.

Fix positive numbers δ_1, δ_2. Define t_j (for $j = 0, 1, 2 \ldots$) as

Initialization:

$$\delta_1 < t_0 < \delta_2 \tag{6}$$

Iterative step: Given t_j,

$$\text{If} \qquad f(x^k - t_j \nabla f(x^k)) \leq f(x^k) - \psi(t_j) \left\| \nabla f(x^k) \right\|^2 \tag{7}$$

then $\lambda_k = t_j$ and the iteration stops. Otherwise

$$t_{j+1} = \frac{t_j}{2}.$$

In order for Algorithm B to be well defined it must be established that inequality (7) is satisfied after some finite number of steps. This will be done in section 4.

Example 1: $\psi(u) = \alpha u$, $\alpha \in (0, 1)$. In this case we recover the algorithm of backtracking studied by Dennis-Schnabel [**3**] and Polyak [**8**]. See Figure 1.

Example 2: $\psi(u) = \alpha u^\beta$, $\alpha > 0$, $\beta \in (1, 2]$. See Figure 2.

Figure 1

Figure 2

§3 Preliminary Results

Definition 1: A sequence $\{y^k\}$ is *quasi Fejér convergent* to a set $U \subseteq \mathbf{R}^n$ if for every

$u \in U$ there exists a sequence $\{\varepsilon_k\} \subseteq \mathbf{R}$ such that $\varepsilon_k \geq 0$, $\sum_0^\infty \varepsilon_k < \infty$ and $\|y^{k+1} - u\|^2 \leq \|y^k - u\|^2 + \varepsilon_k$ for all k.

Theorem 1. *If $\{y^k\}$ is quasi Fejér convergent to a nonempty set $U \subseteq \mathbf{R}^n$, then $\{y^k\}$ is bounded. If furthermore a cluster point y of $\{y^k\}$ belongs to U then $\lim_{k\to\infty} y^k = y$.*

Proof: Take $u \in U$. Applying iteratively Definition 1 we get

$$\|y^k - u\|^2 \leq \|y^0 - u\|^2 + \sum_{j=0}^{k-1} \varepsilon_j \leq \|y^0 - u\|^2 + \sum_{j=0}^{\infty} \varepsilon_j.$$

It follows that $\{y^k\}$ is bounded.

Let now $y \in U$ be a cluster point of $\{y^k\}$ and take $\delta > 0$. Let $\{y^{\ell_k}\}$ be a subsequence of $\{y^k\}$ convergent to y. Using Definition 1 there exist k_0 such that $\sum_{k_0}^\infty \varepsilon_j < \delta/2$, and there exist $k_1 \geq k_0$ such that $\|y^{\ell_k} - y\|^2 < \frac{\delta}{2}$ for any $k \geq k_1$. Then for any $k \geq k_1$ we have:

$$\|y^k - y\|^2 \leq \|y^{\ell_k} - y\|^2 + \sum_{j=\ell_k}^{k-1} \varepsilon_j \leq \|y^{\ell_k} - y\|^2 + \sum_{j=\ell_k}^{\infty} \varepsilon_j < \delta/2 + \delta/2 = \delta$$

It follows then that $\lim_{k\to\infty} y^k = y$. ∎

The next theorem is a version, without differentiability in the first variable, of the well-known Implicit Function Theorem, whose proof can be found, e.g. in [**9**, volume II, page 163].

Theorem 2. *Let $F: \mathbf{R}^n \times \mathbf{R} \to \mathbf{R}$ such that*

i) There exists $(x_0, u_0) \in \mathbf{R}^n \times \mathbf{R}$ such that $F(x_0, u_0) = 0$,

ii) F is continuous in a neighborhood of (x_0, u_0),

iii) F is differentiable with respect to the variable u in (x_0, u_0) and $\frac{\partial F}{\partial u}(x_0, u_0) \neq 0$.

Then there exists a neighborhood $V(x_0)$ of x_0 and at least one function $u: V(x_0) \to \mathbf{R}$ such that $u(x_0) = u_0$ and

$$F(x, u(x)) = 0 \text{ for any } x \in V(x_0). \tag{8}$$

If furthermore,

iv) $\frac{\partial F}{\partial u}(\ \cdot\ ,\ \cdot\ )$ *is continuous at* (x_0, u_0),

then the function u *is the only one that satisfies (8) and is continuous at* x_0.

Let $G = \{x \in \mathbf{R}^n / \nabla f(x) \neq 0\}$. By continuous differentiability of f, G is open.

Proposition 1. *Take* ψ *satisfying B1, B2 and B3. Then*

i) For all $x \in G$ *there exists a unique* $u(x) > 0$ *such that*

$$f(x - u(x)\nabla f(x)) = f(x) - \psi(u(x)) \|\nabla f(x)\|^2 \tag{9}$$

and

$$f(x - u\nabla f(x)) \leq f(x) - \psi(u) \|\nabla f(x)\|^2 \ \text{ iff } 0 \leq u \leq u(x) \tag{10}$$

ii) $u\colon G \to \mathbf{R}_+$ *is continuous in* G.

Proof: i) For any fixed $x \in G$, $u \in \mathbf{R}_+$, define

$$F(x,u) = f(x - u\nabla f(x)) - f(x) + \psi(u) \|\nabla f(x)\|^2 \tag{11}$$

By B1 and B2 $F(x,\ \cdot\ )$ is convex and continuously differentiable, also

$$F(x,0) = 0 \tag{12}$$

$$\frac{\partial F}{\partial u}(x,0) = \|\nabla f(x)\|^2 (\psi'(0) - 1) < 0 \tag{13}$$

and

$$F(x,u) \geq f^* - f(x) + \psi(u) \|\nabla f(x)\|^2 \tag{14}$$

¿From (12) and (13) $F(x,\ \cdot\ )$ is negative in some interval to the right of zero, and from (14), B1 and B2 $\lim\limits_{u\to\infty} F(x,u) = +\infty$. It follows that there exists $u(x) > 0$ such that $F(x, u(x)) = 0$ and (9) holds. Uniqueness of $u(x)$ is implied by convexity of $F(x,\ \cdot\ )$, and the fact that a convex function of real variable can take a given value different from its minimum at most at two different points, while $F(x,0) = F(x,u(x)) = 0$ and zero is not the minimum value of $F(x,\ \cdot\ )$ by (12) and (13). (i) is established.

ii) Let $u_0 := u(x_0)$ given by (i), for a given $x_0 \in G$. Then we have that $F(x_0, u_0) = 0$, $F(\ \cdot\ ,\ \cdot\ )$ is continuous in a neighborhood of (x_0, u_0) and also

$$\frac{\partial F}{\partial u}(x_0, u_0) = -\nabla f(x_0 - u_0 \nabla f(x_0))^t \nabla f(x_0) + \psi'(u_0) \left\| \nabla f(x_0) \right\|^2. \tag{15}$$

As $F(x_0,\ \cdot\ )$ is strictly increasing at u_0, we have that $\frac{\partial F}{\partial u}(x_0 u_0) > 0$. Using (15) we see that $\frac{\partial F}{\partial u}(\ \cdot\ ,\ \cdot\ )$ is continuous at (x_0, u_0), and all the hypotheses of Theorem 2 hold. Therefore u is continuous at x_0. ∎

Let $T = \{z \in \mathbf{R}^n / f(z) \leq \liminf_{k \to \infty} f(x^k)\}$.

Proposition 2. *For any $z \in T$ $\left\| x^{k+1} - z \right\|^2 \leq \left\| x^k - z \right\|^2 + \left\| x^{k+1} - x^k \right\|^2$, where $\{x^k\}$ is generated by (2)–(3) with any $\lambda_k > 0$.*

Proof: Take $z \in T$. Then

$$\begin{aligned} \left\| x^{k+1} - z \right\|^2 - \left\| x^k - z \right\|^2 - \left\| x^{k+1} - x^k \right\|^2 &= -2(z - x^k)^t (x^{k+1} - x^k) = 2\lambda_k (z - x^k)^t \nabla f(x^k) \\ &\leq 2\lambda_k (f(z) - f(x^k)) \leq 0. \end{aligned}$$

using (3) in the second equality, the gradient inequality in the first inequality, and definition of T in the second one. ∎

§4 Analysis of the backtracking procedure

Proposition 3. *The backtracking procedure of Algorithm B defined by (6)–(7) stops after a finite number of iterations with*

$$\min \left\{ \delta_1, \frac{u(x^k)}{2} \right\} \leq \lambda_k \leq \min \left\{ \delta_2, u(x^k) \right\} \tag{16}$$

Proof: We consider two cases for the value of t_0:

1) $t_0 \in (0, u(x^k))$

2) $t_0 \geq u(x^k)$

1) By Proposition 1(i), we get $\lambda_k = t_0$ from (6) and (7), and iteration stops at $j = 0$. (16) is established because $t_0 < \delta_2$ and $t_0 < u(x^k)$ implying $\lambda_k = t_0 < \min\{\delta_2, u(x^k)\}$, and $t_0 > \delta_1$, so $t_0 = \lambda_k \geq \min\left\{\delta_1, \frac{u(x^k)}{2}\right\}$.

2) There exist a unique $s \in \mathbf{N}$, $s \geq 1$ such that

$$2^{s-1}u(x^k) < t_0 \leq 2^s u(x^k). \tag{17}$$

Then

$$\frac{u(x^k)}{2} < \frac{t_0}{2^s} \leq u(x^k). \tag{18}$$

By (6) (7) we have that $t_j = \frac{t_0}{2^j}$, so (18) establishes that

$$\frac{u(x^k)}{2} < t_s \leq u(x^k). \tag{19}$$

We claim that $\lambda_k = t_s$. From (17) and (18) we have that $t_{s-1} > u(x^k)$ and $t_s \leq u(x^k)$ so that, using Proposition 1, (7) is satisfied by $\lambda_k = t_s$ but not by $\lambda_k = t_{s-1}$. (16) follows from (19) and the fact that $t_s \leq t_0 < \delta_2$. ∎

§5 Convergence Analysis

Proposition 4. *For Algorithms A and B it holds that:*

i) there exists $\gamma > 0$ *such that*

$$f(x^{k+1}) \leq f(x^k) - \gamma \left\|x^{k+1} - x^k\right\|^2 \qquad \textit{for all } k, \tag{20}$$

ii) $\{f(x^k)\}$ *is decreasing and convergent,* *iii)* $\sum_{k=0}^{\infty} \left\| x^{k+1} - x^k \right\|^2 < \infty.$

Proof: For Algorithm A, using the Newton-Leibniz formula:

$$\begin{aligned}
f(x^{k+1}) &= f(x^k) - \lambda_k \left\|\nabla f(x^k)\right\|^2 - \lambda_k \int_0^1 (\nabla f(x^k - u\lambda_k \nabla f(x^k)) - \nabla f(x^k))^t \nabla f(x^k)\ du \\
&\leq f(x^k) - \lambda_k \left\|\nabla f(x^k)\right\|^2 + L\lambda_k^2 \left\|\nabla f(x^k)\right\|^2 \int_0^1 u\ du \\
&= f(x^k) - \lambda_k(1 - \frac{L\lambda_k}{2}) \left\|\nabla f(x^k)\right\|^2 = f(x^k) - \frac{1}{\lambda_k}(1 - \frac{L\lambda_k}{2}) \left\|x^{k+1} - x^k\right\|^2.
\end{aligned}$$

Using $\delta_1 \leq \lambda_k \leq \frac{2}{L}(1-\delta_2)$ we get $\lambda_k^{-1}(1 - \frac{L\lambda_k}{2}) \geq \frac{\delta_2 L}{2(1-\delta_2)}$. So we establish (20) for $\gamma = \frac{\delta_2 L}{2(1-\delta_2)}$.

For Algorithm B, we have $f(x^{k+1}) \leq f(x^k) - \psi(\lambda_k) \left\|\nabla f(x^k)\right\|^2$. Then

$$\frac{\psi(\lambda_k)}{\lambda_k^2} \left\|x^{k+1} - x^k\right\|^2 = \frac{\psi(\lambda_k)}{\lambda_k^2} \lambda_k^2 \left\|\nabla f(x^k)\right\|^2 \leq f(x^k) - f(x^{k+1}). \tag{21}$$

Take $0 < \xi < \liminf_{u \to 0^+} \frac{\psi(u)}{u^2}$, using B3. By definition of ξ, there exists $\theta > 0$ such that if $\lambda \in (0, \theta)$ then

$$\frac{\psi(\lambda)}{\lambda^2} > \xi. \tag{22}$$

For each k, we have two possibilities:

a) $\lambda_k \in (0, \theta)$, so $\psi(\lambda_k)\lambda_k^2 > \xi$ by (22),

b) $\lambda_k \geq \theta$. In this case, by Proposition 3, we have that $\lambda_k \leq \min\{\delta_2, u(x^k)\} \leq \delta_2$, and it follows from B1 and B2 that ψ is increasing, implying $\psi(\lambda_k) \geq \psi(\theta)$. So we have $\frac{\psi(\lambda_k)}{\lambda_k^2} \geq \frac{\psi(\theta)}{\delta_2^2}$.

Take $\gamma = \min\{\xi, \frac{\psi(\theta)}{\delta_2^2}\}$ and use (21) to establish (20) for Algorithm B.

ii) Follows from (i), using $\gamma > 0$.

iii) By (i), there exist $\gamma > 0$ such that

$$\sum_{k=0}^{r} \left\|x^{k+1} - s^k\right\|^2 \leq \frac{1}{\gamma}(f(x^0) - f(x^r)) \leq \frac{1}{\delta}(f(x_0) - f^*).$$

Letting $r \to \infty$, we get $\sum_{k=0}^{\infty} \left\| x^{k+1} - x^k \right\|^2 < \infty$. ∎

Proposition 5. *The sequence $\{x^k\}$ generated by (2) (3) is convergent to a point $x^* \in T$.*

Proof: By Propositions 2 and 4(iii) we have that $\{x^k\}$ is quasi-Fejér convergent to T, with $\varepsilon_k = \left\| x^{k+1} - x^k \right\|^2$. Now we only need to prove that there is a cluster point of $\{x^k\}$ in T, and apply the second statement of Theorem 1. By the first statement of the same theorem, $\{x^k\}$ is bounded and so it has cluster points. Using Proposition 4(ii), any cluster point is in T. ∎

Theorem 3. *The sequence $\{x^k\}$ generated by (2)-(3) using Algorithms A or B, converges to a minimizer of f.*

Proof: By Proposition 5, $\lim_{k\to\infty} x^k = x^* \in T$, so it is enough to prove that $x^* \in T^*$, the set of minimizers of f.

For Algorithm A, we have $\left\| x^{k+1} - x^k \right\|^2 = \lambda_k^2 \left\| \nabla f(x^k) \right\|^2 \geq \delta_1^2 \left\| \nabla f(x^k) \right\|^2$ by (5). Then $\nabla f(x^*) = 0$ by Proposition 4(iii) and continuity of $\nabla f(\,\cdot\,)$, so x^* is a minimizer of f by convexity.

For Algorithm B, suppose $x^* \notin T^*$. Then, by convexity of f, $x^* \in G$ and $\|\nabla f(x^*)\| > 0$. By Proposition 1, $u(x^*) > 0$ and $u(x^k)$ converges to $u(x^*)$. So there exists k_0 such that for all $k \geq k_0$

$$u(x^k) \geq \frac{u(x^*)}{2} \quad \text{and} \quad \left\| \nabla f(x^k) \right\|^2 \geq \frac{1}{2} \left\| \nabla f(x^*) \right\|^2 \tag{23}$$

Let $\sigma = \left(\min \left\{ \delta_1, \frac{u(x^*)}{2} \right\} \right)^2 \frac{\|\nabla f(x^*)\|^2}{2}$. Then, for any $k \geq k_0$

$$\begin{aligned} \left\| x^{k+1} - x^k \right\|^2 = \lambda_k^2 \left\| \nabla f(x^k) \right\|^2 &\geq \left(\min \left\{ \delta_1, \frac{u(x^k)}{2} \right\} \right)^2 \left\| \nabla f(x^k) \right\|^2 \\ &\geq \left(\min \left\{ \delta_1, \frac{u(x^*)}{4} \right\} \right)^2 \frac{\|\nabla f(x^*)\|^2}{2} = \sigma > 0 \end{aligned} \tag{24}$$

using (3) in the first equality, Proposition 3 in the first inequality and (23) in the second one. Since (24) contradicts Proposition 4(iii) we have proved that $x^* \in T^*$. ∎

§6 Final Remarks

It can be easily checked that all our results hold under a weaker assumption than convexity, namely pseudoconvexity. We remind that f is pseudoconvex if and only if $(y - x)^t \nabla f(x) \geq 0$ implies $f(y) \geq f(x)$.

Finally, we emphasize that we are well aware of the computational shortcomings of the steepest descent method ("hemstitching" phenomena, etc), ant therefore we do not make any claim on its performance when compared to other procedures for unconstrained minimization, like quasi-Newton methods (see [**2**]), or the conjugate gradients algorithms (see [**1**]). Our main purpose is to show that the notion of quasi-Fejér convergence makes it possible to upgrade the convergence results for Polyak's and Dennis-Schnabel's procedures from weak to full convergence. Since the steepest descent method is not only one of the more basic minimization schemes but also the departure point for some more sophisticated algorithms (e.g. projected gradients, see [**2**]) we think that a full understanding of its convergence properties is relevant.

REFERENCES

[1] J. Abaffy, F. Sloboda, Extended conjugate gradient algorithms for extended quadratic functions, *Numerische Mathematik*, 1983, 42 p. 97–105.

[2] M. Avriel, *Nonlinear Programming, Analysis and Methods*, Prentice Hall, New Jersey, 1976.

[3] J.E. Dennis, R.B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Prentice Hall, New Jersey, 1983.

[4] Y.M. Ermol'ev, On the method of generalized stochastic gradients and quasi-Fejér sequences. *Cybernetics*, 1969, 5, p. 208–220.

[5] A.N. Iusem, B.F. Svaiter, A proximal regularization of the steepest descent method (to be published in *RAIRO, Recherche Opérationelle*).

[6] A.N. Iusem, B.F. Svaiter, M. Teboulle, Entropy-like proximal methods in convex programming (to be published in *Mathematics of Operations Research*).

[7] M. Minoux, *Mathematical Programming, Theory and Algorithms*. John Wiley, New York, 1986.

[8] B. Polyak, *Introduction to Optimization*. Optimization Software, New York, 1987.

[9] J. Rey Pastor, T. Pi Calleja, C.A. Trejo, *Análisis Matemático*. Editorial Kapelusz, Buenos Aires, 1957.

[10] W.I. Zangwill, *Nonlinear Programming: a Unified Approach*. Prentice Hall, New Jersey, 1969.