Control and Cybernetics

VOL. **21** (1992) No. 3/4

On the convergence of a conditional ϵ–subgradient method for nondifferentiable convex optimization

by

Youssef Aly Mohamed Baghdadi

Institute of Mathematics
University of Łódź
ul. S. Banacha 22
90–238 Łódź
Poland

We present a new numerical method for minimizing any convex not necessarily differentiable function of several variables. At each iteration, the method uses one conditional ϵ–subgradient of the given function, which can be computed numerically by applying the algorithm of N.Z. Shor [6, §1.3]. Consequently, no analytical formula for computing subgradients is required.

Keywords: conditional ϵ–subgradient, nondifferentiable optimization, Lipschitzian functions.

1. Introduction

In this paper we present a new numerical method for the minimization of a convex nondifferentiable function $f: \mathbb{R}^n \to \mathbb{R}$. Unlike the majority of methods which can be found in the literature (see e.g. [2,3,4,6]), neither the computation of subgradients nor ϵ–subgradients of the given function are required in our method. Instead, at each iteration, one conditional ϵ–subgradient (as defined in (1.1) below) is used. It can be computed numerically by applying an algorithm described by N.Z. Shor in [6, §1.3]. In its original version, the algorithm of Shor enables one to find, for given $x_0 \in \mathbb{R}^n$ and $\epsilon, \delta > 0$, a point $y \in \mathbb{R}^n$ satisfying $\| y - x_0 \| < \delta$ and a vector z^* for which there exists a subgradient $z \in \partial f(y)$ satisfying $\| z^* - z \| < \epsilon$, where $\| \cdot \|$ is the Euclidean norm corresponding to the standard inner product. The computation of y and z^* requires only a finite number of arithmetic operations and a finite number of evaluations of f at some appropriately constructed points.

In the paper we shall use the following definition which can be found in [2, §I.10]:

DEFINITION 1.1 *Let f be a real–valued convex function defined on a convex subset Ω of $\mathbb{R}^n$. For fixed $x_0 \in \Omega$ and $\epsilon \geq 0$, define*

$$\partial_\epsilon^\Omega f(x_0) = \{ v \in \mathbb{R}^n \mid f(z) - f(x_0) \geq \langle v, z - x_0 \rangle - \epsilon \ \text{ for all } \ z \in \Omega \} \quad (1.1)$$

The set $\partial_\epsilon^\Omega f(x_0)$ is said to be the conditional ϵ–subdifferential of the function f at the point $x_0 \in \Omega$ with respect to the set Ω. An element $v \in \partial_\epsilon^\Omega f(x_0)$ is said to be a conditional ϵ–subgradient of the function f at the point x_0 with respect to the set Ω. In particular, if $\Omega = \mathbb{R}^n$, the set $\partial_\epsilon^{\mathbb{R}^n} f(x_0)$ is simply denoted by $\partial_\epsilon f(x_0)$ and called the ϵ–subdifferential of f at x_0.

We shall prove here (cf. Proposition 2.1 below) that the algorithm of Shor, with suitably chosen parameters ϵ and δ, can be used for exact calculation (if we neglect the roundoff error) of some conditional ϵ–subgradient of f at x_0. Consequently, this finite procedure can easily be included as a subprogram to be used at each iteration of our minimization method.

This paper deals only with a general convergence theorem for the proposed minimization method. We indicate here the way of choosing the parameters ϵ, δ at each iteration, but we do not go into the details of Shor's algorithm. A more detailed description of an implementable version of our method will be the subject of further research.

The minimization method presented here can be applied in those situations in which we are able to compute the values of f at any given point, but we have no analytical formula for computing subgradients or ϵ–subgradients.

In this paper we take as the set Ω (in the definition of the conditional ϵ-subdifferential – see (1.1)) some closed ball with centre x_0.

It should be noted that S.V. Rževskiĭ in paper [7], the title of which is similar to the one adapted here, also discusses the problem of minimizing a convex nondifferentiable function $f: \mathbb{R}^n \to \mathbb{R}$. However, he considers constrained minimization problems and takes the set Ω as the constraint set. Consequently, the approach presented in [7] (see also [8]) is entirely different from ours.

The paper is organized as follows. In §2 we present a result concerning Shor's algorithm (cf. Proposition 2.1 below). The convergence theorem is established in §3. Finally, §4 contains some remarks on the implementation of the method.

2. A result concerning Shor's algorithm

The proposition stated below shows that an element of $\partial_\mu^{B(x_0,r)} f(x_0)$, where $B(x_0, r)$ is the closed ball with centre x_0 and radius $r > 0$, can be computed by applying the algorithm given by N.Z. Shor with suitably chosen parameters ϵ and δ.

PROPOSITION 2.1 *Let $x_0 \in \mathbb{R}^n$ and $\mu, r > 0$ be given. Let L be the Lipschitz constant for f on $B(x_0, 1), \bar{\mu} = \min\{\mu, 4L\}, \delta = \bar{\mu}/4L$ and $\epsilon = \mu/2r$. Then Shor's algorithm gives a vector $z^* \in \partial_\mu^{B(x_0,r)} f(x_0)$.*

PROOF. By Lemma 6.1 in [4], we have

$$\bigcup_{u \in B(x_0, \bar{\mu}/4L)} \partial f(u) \subset \partial_{\mu/2} f(x_0)$$

Now apply Shor's algorithm with the given parameters δ and ϵ to find elements y, z, z^* satisfying $\| y - x_0 \| < \delta$, $z \in \partial f(y)$, $\| z^* - z \| < \epsilon$ (see Introduction). In particular, we have

$$z \in \partial f(y) \subset \bigcup_{u \in B(x_0, \delta)} \partial f(u) \subset \partial_{\mu/2} f(x_0)$$

This means that

$$\langle z, v - x_0 \rangle \leq f(v) - f(x_0) + \mu/2 \quad \text{for all} \quad v \in \mathbb{R}^n$$

Now, assume that $v \in B(x_0, r)$. Then

$$\begin{aligned} \langle z^*, v - x_0 \rangle &= \langle z, v - x_0 \rangle + \langle z^* - z, v - x_0 \rangle \\ &\leq f(v) - f(x_0) + \mu/2 + \|z^* - z\| \|v - x_0\| \\ &\leq f(v) - f(x_0) + \mu/2 + \epsilon r \quad = \quad f(v) - f(x_0) + \mu \end{aligned}$$

We have thus verified that $z^* \in \partial_\mu^{B(x_0,r)} f(x_0)$.

3. Convergence of the minimization method

The proof of our main result stated below is a suitable modification of the proof of the convergence of the ϵ_k-subgradient method (cf. [2, §3.4]).

THEOREM 3.1 *Let $f: \mathbb{R}^n \to \mathbb{R}$ be a convex function which has a bounded non-empty set of minimum points M^*. For any x_0, consider the sequence $\{x_k; k \geq 0\}$ generated according to the formula*

$$x_{k+1} = x_k + z_k \tag{3.1}$$

where

$$z_k = \begin{cases} 0 & \text{if } v_k = 0, \\ -\lambda_k v_k \|v_k\|^{-1} & \text{if } v_k \neq 0, \end{cases}$$

$v_k \in \partial_{\epsilon_k}^{B(x_k,\delta_k)} f(x_k)$. *We assume that the numbers λ_k, ϵ_k, and δ_k $(k = 0, 1, \ldots)$ satisfy the conditons: $\lambda_k > 0, \lim_{k\to\infty} \lambda_k = 0$, $\sum_{k=0}^{\infty} \lambda_k = \infty$, $\epsilon_k > 0$, $\lim_{k\to\infty} \epsilon_k = 0$, and the sequence $\{\epsilon_k; k \geq 0\}$ is nonincreasing, $\delta_0 > 0$ and $\delta_k = 2\sum_{i=0}^{k-1} \lambda_i$ for $k > 0$. Then*

$$\lim_{k\to\infty} \rho(x_k, M^*) = 0 \quad \text{and} \quad \lim_{k\to\infty} f(x_k) = f^* \tag{3.2}$$

where

$$f^* = \inf_{x\in\mathbb{R}^n} f(x) \quad \text{and} \quad \rho(x, M^*) = \inf_{u\in M^*} \|x - u\|.$$

To prove Theorem 3.1, we need the following lemma:

LEMMA 3.1 *Let D be an arbitrary bounded set in $\mathbb{R}^n$, and let the sequence $\{x_k\}$ be generated according to (3.1). Then there exists k_0 such that*

$$D \subset B(x_k, \delta_k) \quad \text{for all} \quad k \geq k_0. \tag{3.3}$$

PROOF. First, we prove that

$$B(x_0, \delta_k/2) \subset B(x_k, \delta_k) \quad k = 1, 2, \ldots. \tag{3.4}$$

It follows from the construction of algorithm (3.1) that

$$\|x_i - x_{i+1}\| = \lambda_i^* \quad i = 0, 1, 2, \ldots, \tag{3.5}$$

where λ_i^* is equal to either λ_i or 0 (depending on whether $v_k \neq 0$ or $v_k = 0$).

Let $z \in B(x_0, \delta_k/2)$, i.e., $\|z - x_0\| \leq \sum_{i=0}^{k-1} \lambda_i$. Using this inequality and (3.5), we obtain

$$\begin{aligned} \|z - x_k\| &\leq \|z - x_0\| + \sum_{i=0}^{k-1} \|x_i - x_{i+1}\| \\ &\leq \sum_{i=0}^{k-1} \lambda_i + \sum_{i=0}^{k-1} \lambda_i^* \\ &\leq \sum_{i=0}^{k-1} \lambda_i + \sum_{i=0}^{k-1} \lambda_i \quad = \quad \delta_k, \end{aligned}$$

thus $z \in B(x_k, \delta_k)$, and (3.4) is proved.

Since D is bounded, there exists $M > 0$ such that

$$\|u\| \leq M \quad \text{for all} \quad u \in D.$$

Using the property that $\delta_k/2 \to \infty$ as $k \to \infty$, we can find some k_0 such that

$$\delta_k/2 \geq M + \| x_0 \| \quad \text{for all} \quad k \geq k_0.$$

Then we have, for all $u \in D$ and $k \geq k_0$,

$$\|u - x_0\| \leq \|u\| + \|x_0\| \leq M + \|x_0\| \leq \delta_k/2,$$

and so, in view of (3.4),

$$D \subset B(x_0, \delta_k/2) \subset B(x_k, \delta_k) \quad \text{for all} \quad k \geq k_0.$$

PROOF OF THEOREM 3.1. From the boundedness of M^* we obtain the boundedness of the set

$$D_\epsilon = \{x \in \mathbb{R}^n \mid f(x) \leq f^* + \epsilon\}$$

for any $\epsilon > 0$ (cf. [2, Corollary 2 of Lemma I.9.1]).

First, we shall prove that

$$\underline{\lim}_{k\to\infty} \rho(x_k, M^*) = 0. \tag{3.6}$$

Assume the contrary; then there exists $a > 0$ and $k_1 \in \mathbb{N}$ such that

$$\rho(x_k, M^*) \geq 2a > 0 \quad \text{for all} \quad k \geq k_1. \tag{3.7}$$

We may assume that, for all k, $v_k \neq 0$. Indeed, suppose that there exists an infinite set $K \subset \mathbb{N}$ of k for which $v_k = 0$. Then we have, for all $k \in K$,

$$0 = v_k \in \partial_{\epsilon_k}^{B(x_k,\delta_k)} f(x_k),$$

i.e.

$$f(x) - f(x_k) \geq -\epsilon_k \quad \text{for all} \quad x \in B(x_k, \delta_k),$$

that is, for all $k \in K$,

$$f(x_k) \leq f(x) + \epsilon_k \quad \text{for all} \quad x \in B(x_k, \delta_k). \tag{3.8}$$

Since M^* is bounded, it follows from Lemma 3.1 that there exists k_0 such that

$$M^* \subset B(x_k, \delta_k) \quad \text{for all} \quad k \geq k_0. \tag{3.9}$$

Take any $x^* \in M^*$ Then we have, by (3.8) and (3.9), for all $k \in K$,

$$f(x_k) \leq f(x^*) + \epsilon_k = f^* + \epsilon_k \quad \text{for all} \quad k \geq k_0. \tag{3.10}$$

Hence, for all $k \in K$ and $k \geq k_0$,

$$x_k \in D_{\epsilon_k} \subset D_{\epsilon_0}.$$

Since D_{ϵ_0} is bounded, there exists a convergent subsequence $x_{k_s} \to \tilde{x}$ as $s \to \infty$. We have, for all s,

$$f(x_{k_s}) \leq f^* + \epsilon_{k_s},$$

and so, by the continuity of f,

$$f(\tilde{x}) \leq f^* + 0,$$

i.e., $\tilde{x} \in M^*$. This implies that $\rho(x_{k_s}, M^*) \to 0$ as $s \to \infty$, which contradicts assumption (3.7).

Since the set M^* is bounded, therefore, according to Corollary 4 of Lemma I.9.1 in [2], there exists $c > 0$ such that

$$D_c = \{x \mid f(x) \leq f^* + c\} \subset S_a(M^*) \tag{3.11}$$

where

$$S_a(M^*) = \{x \mid \rho(x, M^*) \leq a\}.$$

Since D_c is bounded, we can find, by using Lemma 3.1, some $k_2 \in \mathbb{N}$ such that

$$D_c \subset B(x_k, \delta_k) \quad \text{for all} \quad k \geq k_2 \tag{3.12}$$

where

$$\delta_k = 2\sum_{i=0}^{k-1} \lambda_i.$$

Denote

$$C_{2a}(M^*) = \{x \mid \rho(x, M^*) \geq 2a\}.$$

We shall prove that

$$\inf_{x \in C_{2a}(M^*)} f(x) - f^* = d > c > 0. \tag{3.13}$$

Suppose, to the contrary that

$$\inf_{x \in C_{2a}(M^*)} f(x) - f^* \leq c.$$

Then, for any $m \in \mathbb{N}$, there exists y_m such that

$$\rho(y_m, M^*) \geq 2a \quad \text{and} \quad f(y_m) \leq f^* + c + 1/m. \tag{3.14}$$

Observe that the sequence $\{y_m\}$ is bounded since it is included in the bounded set D_{c+1}. Thus we can choose a convergent subsequence $y_{m_s} \to \tilde{x}$ as $s \to \infty$. Passing to the limit in (3.14), we obtain

$$\rho(\tilde{x}, M^*) \geq 2a \quad \text{and} \quad f(\tilde{x}) \leq f^* + c. \tag{3.15}$$

Hence $\tilde{x} \in D_c$, and so, by (3.11), $\tilde{x} \in S_a(M^*)$, which contradicts the first part of (3.15). This completes the proof of (3.13).

It can easily be shown that there exists $r > 0$ satisfying

$$B(\bar{x}, r) = \{x \mid \|x - \bar{x}\| \leq r\} \subset D_c \quad \text{for all} \quad \bar{x} \in M^*. \tag{3.16}$$

By the definition of $\partial_{\epsilon_k}^{B(x_k,\delta_k)} f(x_k)$ we have, for every k,

$$f(x) - f(x_k) \geq \langle v_k, x - x_k \rangle - \epsilon_k \quad \text{for all} \quad x \in B(x_k, \delta_k). \tag{3.17}$$

It follows from (3.7) and (3.13) that

$$f(x_k) \geq f^* + d \quad \text{for all} \quad k \geq k_1. \tag{3.18}$$

Using the inequality $d > c$ and the fact that $\epsilon_k \rightarrow 0+$ as $k \rightarrow \infty$, we can find $k_3 \geq \max\{k_1, k_2\}$ such that

$$\epsilon_k < d - c \quad \text{for all} \quad k \geq k_3. \tag{3.19}$$

Conditions (3.12), (3.17), (3.18), (3.19) and the definition of D_c imply

$$\langle v_k, x - x_k \rangle \leq f(x) - f(x_k) + \epsilon_k < f^* + c - f^* - d + d - c = 0$$

for all $x \in D_c$ and $k \geq k_3$. Thus, for all $k \geq k_3$, we have

$$\langle z_k, x_k - x \rangle < 0 \quad \text{for all} \quad x \in D_c. \tag{3.20}$$

Let us fix an arbitrary $\bar{x} \in M^*$ and define $\tilde{x}_k = \bar{x} - r z_k \parallel z_k \parallel^{-1}$. It follows from (3.16) that $\tilde{x}_k \in D_c$ and hence, from (3.20), that

$$\langle z_k, x_k - \tilde{x}_k \rangle < 0 \quad \text{for all} \quad k \geq k_3.$$

Therefore

$$\langle z_k, x_k - \bar{x} \rangle + \langle -\lambda_k v_k \parallel v_k \parallel^{-1}, -r\lambda_k v_k \parallel v_k \parallel^{-1} \lambda_k^{-1} \rangle < 0,$$

i.e.,

$$\langle z_k, x_k - \bar{x} \rangle < -r\lambda_k.$$

Furthermore,

$$\begin{aligned} \|x_{k+1} - \bar{x}\|^2 &= \|x_k + z_k - \bar{x}\|^2 \\ &= \|x_k - \bar{x}\|^2 + 2\langle z_k, x_k - \bar{x} \rangle + z_k^2 \\ &< \|x_k - \bar{x}\|^2 - 2r\lambda_k + \lambda_k^2. \end{aligned}$$

Since $\lambda_k \rightarrow 0+$ as $k \rightarrow \infty$, we have, for sufficiently large k,

$$\|x_{k+1} - \bar{x}\|^2 < \|x_k - \bar{x}\|^2 - r\lambda_k.$$

Going on analogously, we obtain

$$\|x_{k+s} - \bar{x}\|^2 < \|x_k - \bar{x}\|^2 - r \sum_{i=0}^{s-1} \lambda_{k+i}. \tag{3.21}$$

Since $\sum_{k=0}^{\infty} \lambda_k = \infty$, we have $\parallel x_k - \bar{x} \parallel \rightarrow -\infty$ as $k \rightarrow \infty$, which is impossible by nonnegativity of $\| \cdot \|$. This contradiction proves (3.6).

Now, let us fix $\alpha > 0$. As above, we find $r_\alpha > 0$ such that the inequality

$$\parallel x_{k+1} - \bar{x} \parallel^2 < \parallel x_k - \bar{x} \parallel^2 - r_\alpha \lambda_k \quad \text{for all} \quad \bar{x} \in M^* \tag{3.22}$$

holds for all sufficiently large k satisfying $\rho(x_k, M^*) > \alpha$. From (3.22) we obtain

$$\rho^2(x_{k+1}, M^*) < \rho^2(x_k, M^*) - r_\alpha \lambda_k. \tag{3.23}$$

By (3.6), there exist numbers k such that

$$\rho(x_k, M^*) \leq \alpha. \tag{3.24}$$

Since $x_{k+1} = x_k + z_k$, $\| z_k \| = \lambda_k$, we may assume, by taking $\lambda_k \to 0+$ into account, that $\rho(x_{k+1}, M^*) < 2\alpha$ for all sufficiently large k satisfying (3.24). If, moreover, $\rho(x_{k+1}, M^*) > \alpha$, then, by (3.23),

$$\rho(x_{k+2}, M^*) < \rho(x_{k+1}, M^*) < 2\alpha.$$

Further, we argue analogously. Namely, if $\rho(x_{k+i}, M^*) \leq \alpha$, then $\rho(x_{k+i+1}, M^*) < 2\alpha$; if, instead, $\alpha < \rho(x_{k+i}, M^*) < 2\alpha$, then $\rho(x_{k+i+1}, M^*) < \rho(x_{k+i}, M^*) < 2\alpha$. Finally, for all $i \geq 1$, we have $\rho(x_{k+i}, M^*) < 2\alpha$. Since α was arbitrary, we have $\rho(x_k, M^*) \to 0$ as $k \to \infty$, and so, $f(x_k) \to f^*$ as $k \to \infty$, which finishes the proof of the theorem.

4. Remarks on the implementation of the method

1) If the function f is globally Lipschitzian, then, to compute the elements $v_k \in \partial_{\epsilon_k}^{B(x_k, \delta_k)} f(x_k)$, we apply the algorithm of Shor as in Proposition 2.1, with the global Lipschitz constant L.

2) If the function f is locally Lipschitzian, then, to compute the elements v_k, we apply the algorithm of Shor with the Lipschitz constant L_k for f on $B(x_k, 1)$. In this case, we must be able to determine the constant L_k for each point x_k.

Acknowledgment

I wish to express my thanks to my supervisor Prof. Dr. M. Studniarski for many helpful discussions and suggestions.

References

[1] Clarke F.H. Optimization and Nonsmooth Analysis. J.Wiley, New York, 1983.

[2] DEMYANOV V.F., VASILEV L.V. Nondifferentiable Optimization (in Russian). Nauka, Moscow, 1981 (English Translation: Optimization Software, Inc., New York, 1985).

[3] KIWIEL K.C. Methods of Descent for Nondifferentiable Optimization. Lecture Notes in Mathematics 1133 , Springer–Verlag, Berlin, 1985.

[4] LEMARÉCHAL C. Nondifferentiable optimization. In: Handbooks in OR & MS, Vol.1, North–Holland 1989, 529–572.

[5] ROCKAFELLAR R.T. Convex Analysis. Princeton University Press, Princeton, New Jersey,1970.

[6] SHOR N.Z. Minimization Methods for Nondifferentiable Functions and their Applications (in Russian). Naukova Dumka, Kiev, 1979 (English Translation: Springer–Verlag, Berlin, 1985).

[7] RŽEVSKIĬ S.V. Conditional ϵ–subgradient method for solving convex programming problems (in Russian). Kibernetika, No.1 (1987), 69 - 72.

[8] RŽEVSKIĬ S.V. On the structure of the conditional ϵ–subgradient method for simultaneous solution of the primal and dual convex programming problems. Doklady AN SSSR, Vol. 311, No.5 (1990), 1055–1059.

O zbieżności metody warunkowego ϵ-subgradientu w wypukłej optymalizacji nieróżniczkowalnej

Przedstawiono nową metodę numeryczną minimalizacji dowolnej funkcji wypukłej – niękoniecznie różniczkowalnej – wielu zmiennych. W każdej iteracji algorytm odwołuje się do jednego warunkowego ϵ–subgradient danej funkcji, który może być wyznaczony numerycznie przy pomocy algorytmu N.Z. Shora [6, §1.3]. Tak więc, nie jest konieczna znajomość wzoru analitycznego dla wyznaczania subgradientów.

О сходимости метода условного ϵ-субградиента в выпуклой недифференцируемой оптимизации

Представлен новый численный метод минимизации произвольной – не обязательно дифференцируемой – выпуклой функции многих переменных. В каждой итерации алгоритм обращается к одному условному ϵ-субградиенту данной функции, который может быть определен с помощью

алгоритма Н.З. Шора [6, §1.3]. Таким образом нет необходомости чтобы аналитическая формула определения субградиента была известна.