NOT FOR QUOTATION
WITHOUT PERMISSION
OF THE AUTHOR

ON MINIMIZING THE SUM OF A CONVEX FUNCTION AND A CONCAVE FUNCTION

L.N. POLYAKOVA

June 1984
CP-84-28

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
A-2361 Laxenburg, Austria

PREFACE

In this paper, the author presents an algorithm for minimizing the sum of a convex function and a concave function. The functions involved are not necessarily smooth and the resulting function is quasidifferentiable. The main property of such functions is the non-uniqueness of directions of steepest descent (and ascent), and therefore special precautions must be taken to guarantee that the algorithm converges to a stationary point.

This paper is a contribution to research on nondifferentiable optimization currently underway within the System and Decision Sciences Program.

ANDRZEJ WIERZBICKI
Chairman
System and Decision Sciences

ON MINIMIZING THE SUM OF A CONVEX FUNCTION AND A CONCAVE FUNCTION

L.N. POLYAKOVA

Department of Applied Mathematics, Leningrad State University, Universitetskaya nab. 7/9, Leningrad 199164, USSR

Received 27 December 1983
Revised 24 March 1984

We consider here the problem of minimizing a particular subclass of quasidifferentiable functions: those which may be represented as the sum of a convex function and a concave function. It is shown that in an n-dimensional space this problem is equivalent to the problem of minimizing a concave function on a convex set. A successive approximations method is suggested; this makes use of some of the principles of ε-steepest-descent-type approaches.

Key words: Quasidifferentiable Functions, Convex Functions, Concave Functions, ε-Steepest-Descent Methods.

1. Introduction

The problem of minimizing nonconvex nondifferentiable functions poses a considerable challenge to specialists in mathematical programming. Most of the difficulties arise from the fact that there may be several directions of steepest descent. To solve this problem requires both a new technique and a new approach. In this paper we discuss a special subclass of nondifferentiable functions: those which can be represented in the form

$$f(x) = f_1(x) + f_2(x) ,$$

where f_1 is a finite function which is convex on E_n and f_2 is a finite function which is concave on E_n . Then f is continuous and quasidifferentiable on E_n , with a quasidifferential at $x \in E_n$ which may be taken to be the pair of sets

$$Df(x) = [\underline{\partial} f(x) \; , \; \bar{\partial} f(x)] \; ,$$

where

$$\underline{\partial} f(x) = \partial f_1(x) = \{v \in E_n | f_1(z) - f_1(x) \geq (v,z-x) \quad \forall z \in E_n\} \; ,$$

$$\bar{\partial} f(x) = \partial f_2(x) = \{w \in E_n | f_2(z) - f_2(x) \leq (w,z-x) \quad \forall z \in E_n\} \; .$$

In other words, $\underline{\partial} f(x)$ is the subdifferential of the convex function f_1 at $x \in E_n$ (as defined in convex analysis) and $\bar{\partial} f(x)$ is the superdifferential of the concave function f_2 at $x \in E_n$.

Consider the problem of calculating

$$\inf_{x \in E_n} f(x) \; . \tag{1}$$

Quasidifferential calculus shows that for $x^* \in E_n$ to be a minimum point of f on E_n it is necessary that

$$-\bar{\partial} f(x^*) \subset \underline{\partial} f(x^*) \; . \tag{2}$$

We shall now show that the problem of minimizing f on the space E_n can be reduced to that of minimizing a concave function on a convex set.

Let Ω denote the *epigraph* of the convex function f_1 , i.e.,

$$\Omega = \text{epi } f = \{z = [x,\mu] \in E_n \times E_1 | h(z) \equiv f_1(x) - \mu \leq 0\} ,$$

and define the following function on $E_n \times E_1$:

$$\psi(z) = f_2(x) + \mu , \quad z = [x,\mu] \in E_n \times E_1 .$$

Set Ω is closed and convex and function ψ is quasidifferentiable at any point $z \in E_n \times E_1$. Take as its quasidifferential at $z = [x,\mu]$ the pair of sets $D\psi(z) = [\{0\} , \partial f_2(x) \times \{1\}]$, where $0 \in E_{n+1}$.

Let us now consider the problem of finding

$$\inf_{z\in\Omega} \psi(z) . \qquad (3)$$

It is well-known (see, e.g., [3]) that if a concave function achieves its infimal value on a convex set, this value is achieved on the boundary of the set.

Theorem 1. *For a point* x^* *to be a solution of problem (1), it is both necessary and sufficient that point* $[x^*,\mu^*]$ *be a solution to problem (3), where* $\mu^* = f(x^*)$.

Proof

Necessity. Let x^* be a solution of problem (1). Then

$$\mu + f_2(x) \geq f_1(x) + f_2(x) \geq f_1(x^*) + f_2(x^*) \quad \forall \mu \geq f_1(x), \quad \forall x \in E_n . \qquad (4)$$

But (4) implies that

$$\psi(z) \geq f_1(x^*) + f_2(x^*) = f_2(x^*) + \mu^* ,$$

where $\mu^* = f_1(x^*)$. Thus there exists a $z^* = [x^*,\mu^*] \in \Omega$ such that

$$\psi(z) \geq \psi(z^*) \quad \forall z \in \Omega . \tag{5}$$

This proves that the condition is necessary.

Sufficiency. That the condition is also sufficient can be proved in an analogous way by arguing backwards from inequality (5).

2. A numerical algorithm

Set $\varepsilon \geq 0$. A point $x_0 \in E_n$ is called an ε-*inf-stationary* point of the function f on E_n if

$$-\bar{\partial} f(x_0) \subset \underline{\partial}_\varepsilon f(x_0) , \tag{6}$$

where

$$\underline{\partial}_\varepsilon f(x_0) = \partial_\varepsilon f_1(x_0) = \{v \in E_n | f_1(z) - f_1(x_0) \geq$$

$$\geq (v_1,z-x_0) - \varepsilon \ \forall x \in E_n\} ,$$

i.e., $\underline{\partial}_\varepsilon f(x_0)$ is the ε-subdifferential of the convex function f_1 at x_0 . Fix $g \in E_n$ and set

$$\frac{\partial_\varepsilon f(x_0)}{\partial g} = \max_{v \in \underline{\partial}_\varepsilon f(x_0)} (v,g) + \min_{w \in \partial f(x_0)} (w,g) \ . \tag{7}$$

<u>Theorem 2</u>. *For a point* x_0 *to be an* ε*-inf-stationary point of the function* f *on* E_n *, it is both necessary and sufficient that*

$$\min_{\|g\|=1} \frac{\partial_\varepsilon f(x_0)}{\partial g} \geq 0 \ . \tag{8}$$

<u>Proof</u>

Necessity. Let x_0 be an ε-inf-stationary point of f on E_n . Then from (6) it follows that

$$0 \in w + \underline{\partial}_\varepsilon f(x_0) \qquad \forall w \in \overline{\partial} f(x_0) \ .$$

Hence

$$\min_{\|g\|=1} \max_{z \in w + \underline{\partial}_\varepsilon f(x_0)} (z,g) \geq 0 \qquad \forall w \in \overline{\partial} f(x_0) \ ,$$

and thus for every $g \in E_n$, $\|g\|=1$, we have

$$\min_{w \in \overline{\partial} f(x_0)} \max_{v \in \underline{\partial}_\varepsilon f(x_0)} (z,g) \geq 0 \ .$$

However, this means that

$$\min_{\|g\|=1} \frac{\partial_\varepsilon f(x_0)}{\partial g} \geq 0 \tag{9}$$

proving that the condition is necessary. That it is also sufficient can be demonstrated in an analogous way, arguing backwards from the inequality (9) .

Note that since the mapping

$$\underline{\partial}_{\varepsilon} f : E_n \times [0, +\infty) \longrightarrow 2^{E_n}$$

is Hausdorff-continuous if $\varepsilon > 0$ (see, e.g., [1]), then the following theorem holds.

Theorem 3. *If* $\varepsilon > 0$ *then the function* $\max\limits_{v \in \underline{\partial}_{\varepsilon} f(x)} (v,g)$ *is continuous in* x *on* E_n *for any fixed* $g \in E_n$.

Assume that x_0 is not an ε-inf-stationary point. Then we can describe the vector

$$g_{\varepsilon}(x_0) = \arg \min_{\|g\|=1} \frac{\partial_{\varepsilon} f(x_0)}{\partial g}$$

as a direction of ε-steepest-descent of function f at point x_0 .

It is not difficult to show that the direction

$$g_{\varepsilon} = -\left(\frac{v_{0\varepsilon} + w_0}{\|v_{0\varepsilon} + w_0\|}\right),$$

where $v_{0\varepsilon} \in \underline{\partial}_{\varepsilon} f(x_0)$, $w_0 \in \bar{\partial} f(x_0)$ and

$$-\max_{w \in \bar{\partial} f(x_0)} \min_{v \in \underline{\partial}_{\varepsilon} f(x_0)} \|v+w\| = -\|v_{0\varepsilon} + w_0\| = a_{\varepsilon}(x_0) ,$$

is a direction of ε-steepest-descent of function f at point x_0 .

Now let us consider the following method of successive approximations.

Fix $\varepsilon > 0$ and choose an arbitrary initial approximation $x_0 \in E_n$. Suppose that the Lebesque set

$$D(x_0) = \{x_0 \in E_n | f(x) \le f(x_0)\}$$

is bounded. Assume that a point $x_k \in E_n$ has already been found. If $-\bar{\partial} f(x_k) \subset \underline{\partial}_\varepsilon f(x_k)$, then x_k is an ε-inf-stationary point of f on E_n ; if not, take

$$x_{k+1} = x_k + \alpha_k g_{\varepsilon k} , \quad \alpha_k = \arg \min_{\alpha \ge 0} f(x_k + \alpha g_{\varepsilon k}) ,$$

where $g_{\varepsilon k} = g_\varepsilon(x_k)$ is an ε-steepest-descent direction of f at x_k .

Theorem 4. *The following relation holds:*

$$\lim_{k \to \infty} a_\varepsilon(x_k) = 0 .$$

Proof. We shall prove the theorem by contradiction. Assume that a subsequence $\{x_{k_s}\}$ of sequence $\{x_k\}$ and a number $a>0$ exist such that

$$a_\varepsilon(x_{k_s}) \le -a \quad \forall s .$$

(The required subsequence must exist since $D(x_0)$ is compact.) Without loss of generality, we can assume that $x_{k_s} \longrightarrow x^*$ (clearly, $x^* \in D(x_0)$). Then

$$f(x_{k_s} + \alpha g_{\varepsilon k_s}) = f(x_{k_s}) + \int_0^{\alpha} \left(\frac{\partial f_1(x_{k_s} + \tau g_{\varepsilon k_s})}{\partial g_{\varepsilon k_s}} \right) d\tau +$$

$$+ \alpha \left(\frac{\partial f_2(x_{k_s})}{\partial g_{\varepsilon k_s}} \right) + o(\alpha, g_{\varepsilon k_s}) \ ,$$

where

$$\frac{o(\alpha, g_{\varepsilon k_s})}{\alpha} \xrightarrow[\alpha \longrightarrow +0]{} 0 \ .$$

The term $o(\alpha, g_{\varepsilon k_s})$ appears in the above equation due to the concavity of f_2 . The fact that function f_2 is concave implies that

$$o(\alpha, g_{\varepsilon k_s}) \leq 0 \qquad \forall\, \alpha > 0 \ , \quad \forall\, g_{\varepsilon k_s} \in E_n \ ,$$

and therefore

$$f(x_{k_s} + \alpha g_{\varepsilon k_s}) \leq f(x_{k_s}) + \int_0^{\alpha} \max_{v \in \partial f_1(x_{k_s} + \tau g_{\varepsilon k_s})} (v, g_{\varepsilon k_s}) d\tau +$$

$$+ \alpha \min_{w \in \partial f_2(x_{k_s})} (w, g_{\varepsilon k_s}) \ .$$

Since $\partial_{\varepsilon} f_1(x) \supset \partial f_1(x)$ for every $x \in E_n$, we have

$$\max_{v \in \partial_{\varepsilon} f_1(x_{k_s} + \tau g_{\varepsilon k_s})} (v, g_{\varepsilon k_s}) \geq \max_{v \in \partial f_1(x_{k_s} + \tau g_{\varepsilon k_s})} (v, g_{\varepsilon k_s}) \ ,$$

and thus

$$f(x_{k_s}+\alpha g_{\varepsilon k_s}) \le f(x_{k_s}) + \int_0^{\alpha} \max_{v\in\partial_\varepsilon f_1(x_{k_s}+\tau g_{\varepsilon k_s})} (v,g_{\varepsilon k_s})d\tau +$$

$$+ \alpha \min_{w\in\partial f_2(x_{k_s})} (w,g_{\varepsilon k_s}) .$$

Since the mapping $\partial_\varepsilon f_1$ is Hausdorff-continuous at the point x^*, there exists a $\delta > 0$ such that

$$\partial_\varepsilon f_1(x) \subset \partial_\varepsilon f_1(y) + \frac{a}{2} S_1(0) \quad \forall x,y \in S_\delta(x^*) ,$$

where $S_r(z) = \{x \in E_n | \|x-z\| \le r\}$. Also, there exists a number $K > 0$ such that

$$x_{k_s} \in S_{\delta/2}(x^*) \quad \forall k_s > K ,$$

and hence

$$f(x_{k_s}+\alpha g_{\varepsilon k_s}) \le f(x_{k_s}) + \alpha(a_\varepsilon(x_{k_s}) + \frac{a}{2})$$

$$\forall \alpha \in (0,\frac{\delta}{2}] , \quad \forall k_s > K .$$

Therefore

$$f(x_{k_s+1}) = \min_{\alpha\ge 0} f(x_{k_s}+\alpha g_{\varepsilon k_s}) \le f(x_{k_s}+\frac{\delta}{2} g_{\varepsilon k_s}) \le$$

$$\le f(x_{k_s}) - \frac{\delta a}{4} . \qquad (10)$$

Inequality (10) contradicts the fact that sequence $\{f(x_k)\}$ is bounded, thus proving the theorem.

References

[1] E.A. Nurminski, "On the continuity of ε-subgradient mappings", *Cybernetics* 5(1977) 148-149.

[2] L.N. Polyakova, "Necessary condtitions for an extremum of a quasidifferentiable function", *Vestnik Leningradskogo Universiteta* 13(1980) 57-62 (translated in *Vestnik Leningrad Univ. Math.* 13(1981) 241-247).

[3] R.T. Rockafellar, *Convex Analysis* (Princeton University Press, Princeton, New Jersey, 1970).