A SUBGRADIENT ALGORITHM WITH SPACE DILATION FOR SOLVING MINIMAX PROBLEMS

PHAM TRONG QUAT

Abstract. *In this paper an implementable algorithm using the operation of space dilation for solving the problem of finding the minimax of convex functions is investigated. The algorithm is based on combining and modifying the nonsmooth optimization works of Shor [8], Wolfe [9] and the author [7] [5]. The algorithm is conceptually simple and easy to be implemented. Global convergence of the algorithm is shown.*

1. INTRODUCTION

This paper presents an algorithm for solving the following minimax problem

$$\min f(x) = \min[\max f_i(x)], \qquad (1)$$
$$x \in R^n, \qquad x \in R^n, \qquad 1 \leq i \leq m,$$

where f_i $i = 1, 2, \ldots, m$ are strictly convex real-valued functions defined on R^n. Throughout this paper it is assumed that the functions f_i $(i = 1, 2, \ldots, m)$ are continuously differentiable and

$$\lim_{|x| \to +\infty} f(x) = +\infty. \qquad (2)$$

This problems is "nonsmooth" in the sense that the function f needs not to be differentiable everywhere.

Our algorithm is a modification of the subgradient method with space dilation developed by Shor in [8] for solving problem (1). Shor's algorithm is a nondescent method. Shor suggested to transform the space metric at each iteration so as to accelerate convergence of the subgradient method. He used operators of space dilation of the following type. Let $S \in R^n$, $|S| = 1$, $\alpha > 0$. Then a linear operation $R_\alpha(S)$ such that

$$R_\alpha(S)x = x + (\alpha - 1)SS^T x, \qquad x \in R^n$$

is referred to as the space-dilation operator acting in the direction S with the coefficient α. Shor's algorithm constructs an iterative process of the type

$$x_{k+1} = x_k + \alpha_k d_k,$$

where d_k is a vector determining the direction and α_k is a numerical factor whose value determines the length of the step in the direction of d_k. The method of choosing α_k in Shor's algorithm is practically impossible since usually the value of constant μ is unknown (see [7]). Therefore, we aim to construct a new algorithm which modifies the method of choosing the step length. We use the operator of space dilation for finding direction at each iteration and our choice of the step length is based on Wolfe's idea. Global convergence of the algorithm is established. The algorithm is conceptually simple and easy to be implemented. In particular, it does not require the solution of an auxiliary problem for generating search direction as in [2], [3], [4]... Hence it can be used for solving large scale problems.

In Section 2 we present the algorithm, while its convergence is discussed in Section 3. In Section 4 we present a simple numerical example.

2. ALGORITHM

The algorithm uses positive parameters β, m_1, m_2 satisfying

$$0 < m_2 < m_1 < 0.5 \quad \text{and} \quad 0 < \beta < 1. \tag{3}$$

Initially we have a starting point $x^1 \in R^n$, $g^0 = 0 \in R^n$ and $B_0 = I$, where I is an identity matrix. Suppose a point x^k, a vector g^{k-1} and a matrix B_{k-1} are known. To find the next point x^{k+1}, the vector g^k and the matrix B_k the algorithm realizes the following iterative process.

Step 1: Take $i(k) \in I(x^k) = \{i \ : \ f_i(x^k) = f(x^k)\}$ such that

$$(B_{k-1}^T f'_{i(k)}(x^k), g^{k-1}) \le m_1 |g^{k-1}|^2, \tag{4}$$

where $f'_i(x)$ denotes the gradient of f_i at x. If $f'_{i(k)}(x^k) = 0$, terminate. Otherwise, compute

$$p^{k-1} = B_{k-1}^T f'_{i(k)}(x^k). \tag{5}$$

Step 2: Set

$$S^k = (p^{k-1} - g^{k-1})/|p^{k-1} - g^{k-1}|. \tag{6}$$

Step 3: Compute

$$B_k = B_{k-1} R_\beta(S^k) \tag{7}$$

where $R_\beta(S^k)$ is the space dilation operator acting in the direction S^k with the coefficient β.

Step 4: Compute

$$g^k = B_k^T f'_{i(k)}(x^k) = R_\beta(S^k) p^{k-1}. \tag{8}$$

Step 5: Set

$$d^k = B_k g^k = B_k B_k^T f'_{i(k)}(x^k). \tag{9}$$

Step 6: Find $t^k \geq 0$ such that

$$f(x^k - t^k d^k) \leq f(x^k) - m_2 t^k |g^k|^2 \tag{10}$$

and

$$(B_k^T f'_{i(k+1)}(x^k - t^k d^k), g^k) \leq m_1 |g^k|^2, \tag{11}$$

for some index $i(k+1) \in I(x^k - t^k d^k)$.

Step 7: Set $x^{k+1} = x^k - t^k d^k$, increase k by 1 and go to Step 1.

Remark 1. To find $t^k \geq 0$ satisfying the conditions (10) and (11) we realize the following process. Initially, we determine

$$f'(x^k, -d^k) = \max_{i \in I(x^k)} (f'_i(x^k), -d^k) = (f'_r(x^k), -d^k). \tag{12}$$

Now let us consider the following two cases.

First case. We have

$$f'(x^k, -d^k) \geq -m_1 |g^k|^2.$$

Set $t^k = 0$ and $i(k+1) = r$.

Second case. We have

$$f'(x^k, -d^k) < -m_1 |g^k|^2.$$

There exists $t^k > 0$ satisfying condition (10) and

$$f(x^k - t^k d^k) \geq f(x^k) - m_1 t^k |g^k|^2. \tag{12}$$

The finite iterative process to finding t^k satisfying conditions (10) and (13) is proposed in [5], [7].

The above remark and the following result show that the algorithm is well-defined.

Lemma 1. *Assume that conditions (2) - (3) are satisfied and let t^k be a positive value satisfying the conditions (10) and (13). Then condition (11) holds.*

P r o o f. From Lemma 2.1 in [5] we have

$$(g, d^k) \leq m_1 |g^k|^2$$

for any $g \in \partial f(x^k - t^k d^k)$, where $\partial f(x)$ denotes the set of all subgradients of f at x. It is known that $f_i'(x^k - t^k d^k) \in \partial f(x^k - t^k d^k)$ for any $i \in I(x^k - t^k d^k)$. Therefore, it follows

$$(f_i'(x^k - t^k d^k), d^k) \leq m_1 |g^k|^2 \tag{14}$$

for any $i \in I(x^k - t^k d^k)$. Combining relation (9) and inequality (14) we obtain

$$(B_k^T f_i'(x^k - t^k d^k), g^k) = (f_i'(x^k - t^k d^k), d^k) \leq m_1 |g^k|^2$$

for any $i \in I(x^k - t^k d^k)$.

This completes the proof.

3. CONVERGENCE

In this section we show that if the algorithm generates an infinite sequence $\{x^k\}$ then $\{x^k\}$ converges to a solution of problem (1).

Lemma 2. *If the algorithm terminates in STEP 1, then x^k solves problem (1).*

From now on we suppose that the algorithm generates an infinite sequence $\{x^k\}$. We use the following notation. Let C be a compact convex set and $S \in R^n$, $|S| = 1$. Two values

$$\begin{aligned} d_s(C) = \min\{d \ : \ (s, x) - d \leq 0, \ \forall x \in C\} \\ - \max\{d \ : \ (s, x) - d \geq 0, \ \forall x \in C\} \end{aligned}$$

and $d(C) = \min_{|s|=1} d_s(C)$ are called a width of the set C in the direction S and a width of the set C, respectively. We denote $D(C) = \sup\{|x-y| \; : \; x \in C, y \in C\}$. For $\delta > 0$ and $\varepsilon > 0$ we define

$$p_{\delta,\varepsilon}(x) = \partial f(x,\delta) \bigcup \Big(\bigcup_{z \in \partial f(x)} B_\varepsilon(z) \Big)$$

where $\partial f(x,\delta) = \text{conv}\; U\{\partial f(y) \; : \; |y-x| < \delta\}$ is called the Goldstein δ - subdifferential and $B_\varepsilon(z) = \{y \in R^n \; : \; |y-z| \le \varepsilon\}$. For any convex set $C \subset R^n$ we define the following functions:

$$e_s(C) = \inf_{z \in C} |(S,z)|, \quad \text{where} \quad S \in R^n, |S| = 1,$$

$$K_S(C) = \begin{cases} \dfrac{d_s(C)}{e_s(C)}, & \text{if} \quad e_s(C) \neq 0 \\ +\infty, & \text{if} \quad e_s(C) = 0 \end{cases}$$

and

$$F(C) = \inf_{|S|=1} K_S(C).$$

Lemma 3. *Assume that conditions (2), (3) are satisfied and let $\{x^k\}_{k=0}^{\infty}$ be the sequence generated by the algorithm. Then*

i) $\lim_{k \to +\infty} f(x^k) = f_\infty \ge f^* = \min_{x \in R^n} f(x)$,

ii) $|x^{k+1} - x^k| \longrightarrow 0$, as $k \longrightarrow +\infty$,

iii) $|p^{k-1} - g^{k-1}| \ge |p^{k-1}|$.

P r o o f. From inequality (10) and condition (2) it is evident that the sequence $\{f(x^k)\}_{k=1}^{\infty}$ is nonincreasing and bounded from below, therefore there exists

$$\lim_{k \to \infty} f(x^k) = f_\infty \ge f^* = \min_{x \in R^n} f(x).$$

Let us now prove that $|x^{k+1} - x^k| \to 0$, as $k \to +\infty$. Assume, to the contrary, that $|x^{k-1} - x^k| \not\to 0$, as $k \to +\infty$. Then we can always choose an infinite subset of indices $K \subset N$ such that.

$$|x^{k+1} - x^k| \ge \delta > 0, \quad \text{for all} \quad k \in K, \tag{15}$$

$x^k \to x'$, $x^{k+1} \to x''$ and $g_f(x^k) \to g'$, as $k \to +\infty$ and $k \in K$, where $g_f(x^k) = f'_{i(k)}(x^k)$. Then

$$
\begin{aligned}
0 \le f(x^k) - f(x^{k+1}) &\le (g_f(x^k), x^k - x^{k+1}) \\
&= (g_f(x^k), t^k B_k B_k^T g_f(x^k)) = t^k |g^k|^2 \\
&\le \frac{f(x^k) - f(x^{k+1})}{m_2} \longrightarrow 0, \quad \text{as} \quad k \longrightarrow +\infty.
\end{aligned}
$$

This implies $(g', x' - x'') = 0$, and $g' \in \partial f(x')$, $f(x'') = f(x') = f_\infty$, which conflicts with the assumption that the function f is strictly convex. Thus $|x^{k+1} - x^k| \longrightarrow 0$, as $k \longrightarrow +\infty$.

From the description of the algorithm we have

$$
\begin{aligned}
|p^{k-1} - g^{k-1}| &= |B_{k-1}^T f'_{i(k)}(x^k) - B_{k-1}^T f'_{i(k-1)}(x^{k-1})|^2 \\
&= |B_{k-1}^T f'_{i(k)}(x^k)|^2 - 2(B_{k-1}^T f'_{i(k)}(x^k), B_{k-1}^T f'_{i(k-1)}(x^{k-1})) \\
&\quad + |g^{k-1}|^2 \ge |p^{k-1}|^2 - 2m_1 |g^{k-1}|^2 + |g^{k-1}|^2 \\
&= |p^{k-1}|^2 + (1 - 2m_1)|g^{k-1}|^2.
\end{aligned}
$$

From condition (3) it is clear that $|p^{k-1} - g^{k-1}| \ge |p^{k-1}|$, for all k. This completes the proof.

Combining Lemma 3 and Theorem 3.11 in [8] it is easy to obtain the following.

Theorem 1. *Assume that conditions (2), (3) are satisfied and let $\{x^k\}_{k=0}^{\infty}$ be the sequence generated by the algorithm. Then for any $\nu \in R$, $\sqrt[n]{\beta} < \nu < 1$, $\varepsilon > 0$, $\delta > 0$, $k \in N$, there exists $\tilde{k} > k$ such that*

$$
F(\mathrm{conv}\, p_{\delta,\varepsilon}(x^{\tilde{k}})) \ge \sqrt{\frac{\nu^2 \sqrt[n]{\alpha^2} - 1}{\alpha^2 - 1}}
$$

where $\alpha = 1/\beta$ and conv C denotes the convex hull of C.

Theorem 2. *Assume that condition (2), (3) are satisfied and $0 \notin$ aff $\partial f(x)$ for $x \in M_1 = \{x \in R^n \ : \ f(x) \le f(x^1)\ x \ne x^*\}$ where aff C is a k-plane generated by C $(k \le n-1)$ and x^* is a minimum point of f. Then the sequence generated by the algorithm $\{x^k\}_{k=0}^{\infty}$ converges to the minimum point x^* of f on R^n.*

P r o o f. From Lemma 3 we have

$$
\lim_{k \to +\infty} f(x^k) = f_\infty \ge f^*.
$$

Let us now prove that $f_\infty = f^*$. Assume the contrary that $f_\infty > f^*$. We denote

$$M = \{x \in R^n \;:\; f(x) = f_\infty\}.$$

Thus $x^* \notin M$. From the continuity of the convex function f and condition (2) it is easy to see that M is a compact set. Let us denote

$$\gamma = \inf \{\, |y| \;:\; y \in \operatorname{aff} \partial f(x);\; x \in M \,\}.$$

We shall show that $\gamma > 0$. Indeed, if $\gamma = 0$, there exist sequences $\{y^k\}_{k=0}^\infty$ and $\{z^k\}_{k=0}^\infty$ such that $y^k \in \operatorname{aff} \partial f(z^k)$, $z^k \in M$ and $|y^k| \to 0$ as $k \to +\infty$. We know that

$$\partial f(x) = \operatorname{conv} \{f_i'(x) \;:\; i \in I(x)\}.$$

Hence we have

$$y^k = \sum_{i \in I(z^k)} \lambda_i^k f_i'(z^k) \to 0; \qquad \sum_{i \in I(z^k)} \lambda_i^k = 1. \tag{16}$$

Since $I(z^k) \subset \{1, 2, \ldots, n\}$ for any k, there exists a subsequence $\{z^k\}_{k \in K}, K \subset N$ such that $I(z^k) = I \subset \{1, 2, \ldots, n\}$ for any $k \in K$.

From relation (16) and the fact that $f_i'(x)$ is continuous, M is compact, $f_i'(x) \neq 0$ at $x \in M$ and $i \in I(x)$ it is easily seen that there exists an infinite subset of indices $K_1 \subset K$ such that

a) $\lambda_i^k \to \lambda_i$, as $k \to +\infty$, $k \in K_1$ and $i \in I$,

b) $z^k \to z^0$, as $k \to +\infty$, $k \in K_1$.

Then we obtain

$$(y^k) \to 0 = \sum_{i \in I} \lambda_i f_i'(z^0), \; \sum_{i \in I} \lambda_i = 1, \tag{17}$$

as $k \to +\infty$, $k \in K_1$. Hence $z^0 \in M$. We have

$$f_i(z^k) = f(z^k) = \max_{1 \le j \le n} f_j(z^k), \quad \text{for } i \in I.$$

From the continuity of the functions f and f_i, $i \in I$, it follows

$$f_i(z^0) = \lim_{k \to +\infty} f_i(z^k) = \lim_{k \to +\infty} (z^k) = f(z^0) = \max_{1 \le j \le n} f_j(z^0).$$

This implies

$$I \subset I(z^0). \tag{18}$$

Combining (17) and (18) we get $0 \in \text{aff}\, \partial f(z^0)$ and $z^0 \in M$. This contradicts to the assumption. Thus we have $\gamma > 0$. Let us denote

$$W(\eta) = \{y \in R^n : |y - x| \leq \eta,\ x \in W \subset R^n\}.$$

From the assumption that $0 \notin aff\ \partial f(x)$, for $x \in M_1$ it is easy to see that dim aff $\partial f(x) < n$, for $x \in M_1$. Hence $d[\partial f(x)(\eta)] \leq 2\eta$, for any $x \in M \subset M_1$. Let us choose $\eta > 0$ such that

$$\frac{2\eta}{\gamma - \eta} < \sqrt{\frac{\nu^2 \sqrt[n]{\alpha^2 - 1}}{\alpha^2 - 1}}.$$

Since f is the convex function, for every $y \in M$, there exists $\delta(y) > 0$ such that

$$\partial f(z) \subset \partial f(y)\Big(\frac{\eta}{2}\Big), \quad \text{for } z \in B_{\delta(y)/2}(y).$$

Since the set M is compact and $M \subset \bigcup_{y \in M} B_{\delta(y)/4}(y)$, it follows that there exist finite points $\{y^1, y^2, \ldots, y^t\} \subset M$ such that

$$M \subset \bigcup_{i=1}^{t} B_{\delta(y^i)/4}(y^i).$$

Let us set $\delta = \min_{1 \leq i \leq t} \frac{1}{2}\delta(y^i) > 0$. Then we can always choose some finite $\overline{K}$ sufficiently large and every x^k, $k > \overline{K}$ we find y^i, $1 \leq i \leq t$ such that

$$x^k \in B_{\delta(y^i)/2}(y^i) \quad \text{and} \quad B_\delta(x^k) \subset B_{\delta(y^i)}(y^i).$$

Thus for $\varepsilon > 0$, $0 < \varepsilon \leq \eta/2$ we have

$$\text{conv } P_{\delta,\varepsilon}(x^k) \subset \partial f(y^i)(\eta), \quad \text{for } k > \overline{K}.$$

Assume that a vector $a(x) \in R^n$, $|a(x)| = 1$ is orthogonal to aff $\partial f(x)$. From the definition of the width of the set we obtain

$$d_{a(y^i)}[\text{conv } P_{\delta,\varepsilon}(x^k)] \leq d_{a(y^i)}\big(\partial f(y^i)(\eta)\big) \leq 2\eta, \quad k > \overline{K}.$$

Then

$$F\big(\text{conv } P_{\delta,\varepsilon}(x^k)\big) = \min_{|S|=1} \frac{d_s[\text{conv } P_{\delta,\varepsilon}(x^k)]}{e_s[\text{conv } P_{\delta,\varepsilon}(x^k)]}$$

$$\frac{d_{a(y^i)}[\text{conv } P_{\delta,\varepsilon}(x^k)]}{e_{a(y^i)}[\text{conv } P_{\delta,\varepsilon}(x^k)]} \leq \frac{2\eta}{\gamma - \eta} < \sqrt{\frac{\nu^2 \sqrt[n]{\alpha^2} - 1}{\alpha^2 - 1}}$$

$k > \overline{K}$. So we have derived a contradiction with theorem 1. This implies

$$\lim_{k \to +\infty} f(x^k) = \min_{x \in R^n} f(x) = f^*.$$

Moreover, since f is the strictly convex function, it is easy to see that the sequence $\{x^k\}_{k=0}^{\infty}$ converges to the minimum point x^* of f on R^n. The theorem is proved.

Remark 2. We know, that if the vectors $f_i'(x)$, $i \in I(x)$ are linearly independent, then $0 \notin \text{aff } \partial f(x)$.

4. EXAMPLE

The objective function to be minimized is

$$f(x) = \max\{f_1(x), f_2(x)\},$$

where $x \in R^2$, $f_1(x) = 4x_1^2 + (x_2 - 4)^2$; $f_2(x) = (2x_1 - 4)^2 + x_2^2$. For this problem, the optimal solution is $x^* = (1, 2)$ with $f(x^*) = 8$. Let IP denote the number of calculations of the function values and IG that of the function subgradients.

Our algorithm used the starting point $x^0 = (2, 0)$ with $f(x^0) = 32$. For $\beta = 0,3$, $m_1 = 0.25$, $m_2 = 0.1$, we obtained

K = 41, IG = 43, IP = 622

$X_1^{41} = 1.00011606083$, $X_2^{41} = 2.0023198041$

$f(x^{41}) = 8.0000164193.$

REFERENCES

1. Demjanov V. F., Vasiliev L. V., Nondifferentiable Optimization (in Russian), Moscow, 1981.
2. Kiwiel K. C., An aggregate subgradient method for nonsmooth convex minimization. Mathematical Programming 27 (1983), 320-341.
3. Lemarechal C., Nonsmooth Optimization and Descent Methods RR-78-4. International Institute for Applied Systems Analysis, Laxenburg, Australia, 1978.
4. Mifflin R., An algorithm for constrained optimization with semismooth function. Mathematics of Operations Research 2(1977), 191-207.

5. Pham Trong Quat. Methods with space dilation for solving nonsmooth convex optimization problems and their applications. Ph. D. thesis, Technical University of Warsaw (in Polish), 1985.
6. Pham Trong Quat. An exact penalty function algorithm using the operation of space dilation for constrained optimization with convex function. Control and Cybernetics, Vol 16(1987). N 2, 103-112.
7. Pham Trong Quat. A subgradient method with space dilation for minimizing convex function. Control and Cybernetics, Vol 13(1984) N 4, 399-408.
8. Shor N. Z., Methods for Minimizing Nondifferentiable Functions and Their Applications. Kiev, Naukova Dumka 1979 (in Russian).
9. Wolfe P., Method of conjugate subgradients for minimizing nondifferentiable functions. In M.L. Balinskii and P. Wolfe eds. Nondifferentiable Optimization. Math. Programming Study 3, Amsterdam, North-Holland, 1975, 145-173.

Faculty of Mathematics, Mechanics and Informatics
Hanoi State University
Nguyen Trai Str. 90
Hanoi, Vietnam