Control
and Cybernetics
VOL. 16 (1987) No. 2

An exact penalty function algorithm using the operation of space dilation for constrained optimization with convex functions

by

PHAM TRONG QUAT

Faculty of Mathematics and Mechanics,
University of Hanoi

In this paper an implementable algorithm using the operation of space dilation for solving constrained minimization problems involving (not necessarily smooth) convex functions is investigated. The algorithm minimizes an exact penalty function via the subgradient method with space dilation for unconstrained minimization. A scheme for automatic limitation of penalty growth is given. Global convergence of the algorithm is established.

1. Introduction

In this paper we shall be concerned with an algorithm for solving the following constrained optimization problem

$$\text{minimize } f(x), \quad \text{subject to} \quad F(x) \leqslant 0, \tag{1}$$

where f and F are convex (possibly nonsmooth) real-valued functions defined on R^n. It is assumed throughout this paper that

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

and the Slater constraint qualification holds, i.e. there exists $\tilde{x} \in R^n$ satisfying $F(\tilde{x}) < 0$, so that the set of feasible points $S = \{x \in R^n \colon F(x) \leqslant 0\}$ has a nonempty interior. The algorithm only requires the computation of $f(x)$ and $F(x)$, and two arbitrary subgradients $g_f(x) \in \partial f(x)$ and $g_F(x) \in \partial F(x)$ at each $x \in R^n$.

The algorithm extends and modifies the subgradient method with space dilation in [4] for unconstrained minimization of an exact penalty function for problem (1).

Polak, Mayne and Wardi [5] proposed an exact penalty method for nonconvex nonsmooth problems. They developed general schemes for the extension of differentiable optimization algorithms to nondifferentiable problems. Kiwiel [2] presented a readily implementable algorithm for solving constrained minimization problems involving convex function. This algorithm minimizes an exact penalty function via the aggregate subgradient method for unconstrained minimization.

Our algorithm is based on combining, modifying and extending the ideas contained in Mifflin [3], Kiwiel [2], Shor [6], Wolfe [7] and [4]. We use the operator of space dilation for finding a direction at each iteration and our choice of step sizes is based on Mifflin's and Wolfe's ideas. A scheme for automatic limitation of penalty growth is given. By exploiting convexity, our method attains convergence properties that are stronger than those in [5].

In Section 2 we present the algorithm, while its convergence is discussed in Section 3. Section 4 contains two small illustrative numerical examples.

We use the following notation. We denote by $\langle \cdot, \cdot \rangle$ and $|\cdot|$ respectively, the usual scalar product and norm in finite-dimensional, real Euclidean space R^n. For any set $B \subset R^n$, "conv B" is the convex hull of B. If $H: R^n \to R^1$ is a convex function and $\eta > 0$, the Goldstein η-subdifferential is defined by

$$\partial H(x, \eta) = \text{conv}\,\{g \in \partial H(y) : |y - x| \leqslant \eta, \quad y \in R^n\}$$

where

$$\partial H(x) = \{g \in R^n : H(z) \geqslant H(x) + \langle g, z - x \rangle, \quad \forall z \in R^n\}$$

is the ordinary subdifferential. Note that H is continuous and the mapping $\partial H(\cdot, \cdot)$ is locally bounded and upper semicontinuous (see, e.g. [3], [5]).

In this paper we use operators of space dilation of the following type (see [6]). Let $\xi \in R^n$, $|\xi| = 1$, $\alpha > 0$. Then a linear operator $R_\alpha(\xi)$ such that

$$R_\alpha(\xi)\, x = x + (\alpha - 1)\, \xi \xi^T x$$

is referred to as the space-dilation operator acting in the direction ξ with the coefficient α.

2. Algorithm

Define the exact penalty function for problem (1)

$$P(x, c) = f(x) + c \cdot F(x)_+, \quad \text{for all} \quad x \in R^n,$$

where $c>0$ is a penalty coefficient and $F(x)_+ = \max\{F(x), 0\}$. Consider the unconstrained optimization problem

$$\min_{x\in R^n} P(x, c), \tag{3}$$

Clearly, if a point x_c solves problem (3) and $x_c \in S$, then x_c solves problem (1). This is the case if c is large enough (see [1]). We shall, therefore, use the method in [4] for solving problem (3) and choose a suitable value of c in the course of calculations.

The algorithm uses positive parameters δ, m_1, m_2, β_1, $\beta_2 < 1$ satisfying

$$m_2 < m_1 < 0.5, \tag{4}$$

$$m_1/(1-m_1) \leqslant \beta_1 < 1 \tag{5}$$

and two sequences of positive numbers $\{\delta_k\}$ and $\{\alpha_k\}$ satisfying

$$\begin{aligned} &\delta_k \to 0, \quad \text{as} \quad k \to +\infty, \\ &\alpha_k \to 0, \quad \text{as} \quad k \to +\infty \end{aligned} \tag{6}$$

and a positive number L satisfying

$$L > 2f(\bar{x}), \tag{7}$$

where $\bar{x}$ is some feasible point of problem (1).

Algorithm

Initially we have a starting point $x^0 \in R^n$, some $g_f^0 \in \partial f(x^0)$, $g_F(x^0) \in \partial F(x^0)$ and

$$g_F^0 = \begin{cases} 0 \in R^n, & \text{if} \quad F(x^0) \leqslant 0, \\ g_F(x^0), & \text{if} \quad F(x^0) > 0. \end{cases}$$

Set $c_1 = 1$, $v_1 = 1$. Suppose a point x^k and c_{k+1} and v_{k+1} are known. To find the next point x^{k+1} the algorithm realizes the following iterative process:

Step 0. Set $x^{k,0} = x^k$, $y^{k,0} = x^k$ and $d^{k,0} = g_f^k + c_{k+1} g_F^k$, where $g_f^k \in \partial f(y^{k,0})$ and

$$g_F^k = \begin{cases} 0 \in R^n, & \text{if} \quad F(x^k) \leqslant 0, \\ g_F(x^k), & \text{if} \quad F(x^k) > 0. \end{cases}$$

Set $i = 0$ and

$$\varepsilon_k = \max\left\{\sqrt{|P(x^{k-1}, c_{k+1}) - P(x^k, c_{k+1})|};\ \delta_k\right\}. \tag{8}$$

Step 1. If $|d^{k,i}| \leqslant \varepsilon_k$, set $x^{k+1} = x^{k,i}$, $l(k+1) = i$ and go to Step 7; otherwise, go to Step 2.

STEP 2. (Line search: see below for details). Find $t_L^{k,i}$ and $t_R^{k,i}$, $0 \leqslant t_L^{k,i} \leqslant t_R^{k,i}$ such that

$$P(x^{k,i} - t_L^{k,i} d^{k,i}; c_{k+1}) \leqslant P(x^{k,i}; c_{k+1}) - m_2 t_L^{k,i} |d^{k,i}|^2, \tag{9}$$

$$P(x^{k,i} - t_R^{k,i} d^{k,i}; c_{k+1}) \geqslant P(x^{k,i}; c_{k+1}) - m_1 t_R^{k,i} |d^{k,i}|^2, \tag{10}$$

$$|t_R^{k,i} - t_L^{k,i}| \leqslant \frac{\alpha_k}{d^{k,i}}. \tag{11}$$

Set $x^{k,i+1} = x^{k,i} - t_L^{k,i} d^{k,i}$ and $y^{k,i+1} = x^{k,i} - t_R^{k,i} d^{k,i}$, and go to Step 3.

STEP 3. If $|x^k - x^{k,i+1}| > \delta$ or $P(x^k, c_{k+1}) - P(x^{k,i+1}, c_{k+1}) > \delta$ set $x^{k+1} = x^{k,i+1}$, $l(k+1) = i+1$ and go to Step 7; otherwise, go to Step 4.

STEP 4. Choose $g_f^{k,i+1} \in \partial f(y^{k,i+1})$ and $g_F(y^{k,i+1}) \in \partial F(y^{k,i+1})$. Set $g^{k,i+1} = g_f^{k,i+1} + c_{k+1} g_F^{k,i+1}$, where

$$g_F^{k,i+1} = \begin{cases} 0 \in R^n, & \text{if} \quad F(y^{k,i+1}) \leqslant 0, \\ g_F(y^{k,i+1}), & \text{if} \quad F(y^{k,i+1}) > 0. \end{cases}$$

If $\langle g^{k,i+1}, d^{k,i} - g^{k,i+1} \rangle \geqslant 0$ go to Step 5; otherwise, go to Step 6.

STEP 5. Set

$$\xi^{k,i+1} = \frac{g^{k,i+1} - d^{k,i}}{|g^{k,i+1} - d^{k,i}|},$$

$$d^{k,i+1} = R_{\beta_1}(\xi^{k,i+1}) d^{k,i},$$

increase i by 1 and go to Step 1.

STEP 6. Set

$$\xi^{k,i+1} = \frac{d^{k,i} - g^{k,i+1}}{|d^{k,i} - g^{k,i+1}|},$$

$$\gamma_k = |d^{k,i}|^2 \left[1 + \frac{(\beta_1^2 - 1)(1 - 2m_1)\varepsilon_k^2}{|d^{k,i} - g^{k,i+1}|^2} \right],$$

and $q^0 = g^{k,i+1}$, $j = 0$.

i) Set $q^{j+1} = R_{\beta_2}(\xi^{k,i+1}) q^j$.

ii) If $|q^{j+1}| \leqslant \gamma_k$, set $d^{k,i+1} = q^{j+1}$, increase i by 1 and go to Step 1; otherwise, increase j by 1 and go to i).

STEP 7. If

$$|P(x^{k-1}, c_{k+1}) - P(x^k, c_{k+1})| \leqslant v_{k+1}, \tag{12}$$

$$\eta_k = \max\{|x^k - x^{k,i}| : i = 1, 2, \ldots, l(k)\} \leqslant v_{k+1}^2, \tag{13}$$

$$P(x^k, c_{k+1}) < L. \tag{14}$$

set $v_{k+2} = v_{k+1}/2$; otherwise, set $v_{k+2} = v_{k+1}$. If

$$v_{k+2} < v_{k+1}, \tag{15}$$

and

$$F(x^{k+1}) > v_{k+1}, \tag{16}$$

set $c_{k+2} = 2c_{k+1}$; otherwise, set $c_{k+2} = c_{k+1}$. Line search. Define

$$L = \{t \geqslant 0 : f(x^{k,i} - td^{k,i}) \leqslant f(x^{k,i}) - m_2\, t\, |d^{k,i}|^2\},$$

$$R = \{t \geqslant 0 : f(x^{k,i} - td^{k,i}) \geqslant f(x^{k,i}) - m_1\, t\, |d^{k,i}|^2\}.$$

Choose $t > 0$. Set $t_L = 0$; $t_R = +\infty$.

(a) If $t_R - t_L \leqslant \alpha_k / |d^{k,i}|$, go to (e); otherwise, go to (b).

(b) If $t \in L \setminus R$ go to (c).
If $t \in L \cap R$ go to (f).
If $t \in R \setminus L$ go to (d).

(c) If $t_R = +\infty$, replace t_L by t and t by $2t$; otherwise, replace t_L by t^* and t by $(t_R - t_L)/2$, and go to (a).

(d) Replace t_R by t and t by $(t_R - t_L)/2$, and go to (a).

(e) Stop.

(f) Set $t_L = t_R = t$ and stop.

Using the proof of Theorem 4.1 in [3] it is easy to see that the above process is finite.

The results in [4] show that the process of finding x^{k+1} is finite for any k and our algorithm generates an infinite sequence of points $\{x^k\}$. In the next section we shall prove that any accumulation point of the sequence $\{x^k\}$ solves problem (1).

3. Convergence of the algorithm

LEMMA 1. *Suppose that there exist numbers k_c and $c' > 0$ such that $c_k = c'$ for all $k \geqslant k_c$. Then*

i) *$\{x^k\}$ minimizes $P(\cdot, c')$,* i.e. *$P(x^k, c') \downarrow \min \{P(x, c') : x \in R^n\}$.*

ii) *$\min_{x \in R^n} P(x, c') = \min_{x \in S} f(x)$.*

iii) *There exists an accumulation point x' of the sequence $\{x^k\}$ and x' solves problems* (1) *and* (3) *with $c = c'$.*

Proof. From the description of the algorithm and the proof of the theorem in [4] it follows that the sequence $\{P(x^k, c')\}_{k \geqslant k_c}$ is nonincreasing and

$$\lim_{k \to +\infty} P(x^k, c') = \min_{x \in R^n} P(x^k, c'). \tag{17}$$

From Assumption (2), (17), and the proof of the theorem in [4] it is easy to see that there exists an infinite subset $K \subset N$ and $x' \in R^n$ such that

$$x^k \to x', \quad \text{as} \quad k \to +\infty, \quad k \in K$$
$$\eta_{k-1} = \max \{|x^{k-1} - x^{k-1,i}|: i = 1, 2, ..., l(k-1)\} \to 0, \tag{18}$$

as $k \to +\infty$, $k \in K$. From Assumption (7) and (17) it follows that there exists k' such that

$$P(x^{k-1}, c') < L \quad \text{for all} \quad k \geqslant k', \quad k \in K. \tag{19}$$

Combining (12)–(16) and (17)–(19), it is easily seen that there exists an infinite subset $K' \subset K$ such that

$$F(x^k) \leqslant v_k \quad \text{and} \quad v_k \to 0, \quad \text{as} \quad k \to +\infty, \quad k \in K'. \tag{20}$$

From (18) and (20) we obtain $F(x') \leqslant 0$, i.e. $x' \in S$. Then we have

$$\min_{x \in R^n} P(x, c') = \min_{x \in S} f(x).$$

This completes the proof. ■

LEMMA 2. *If the set* $K_c = \{k: c_{k+1} > c_k\}$ *is infinite, then the subsequence* $\{x^k\}_{k \in K_c}$ *has no accumulation point.*

Proof. For purposes of a proof by contradiction, suppose that there exists an infinite subset $K \subset K_c$ and $\bar{x} \in R^n$ such that

$$\lim_{\substack{k \to +\infty \\ k \in K}} x^k = \bar{x}. \tag{21}$$

We know that the function $F(x)_+ = \max \{F(x), 0\}$ is convex, therefore the function $F(\cdot)_+$ is locally Lipschitz (see [6]). It follows that there exists $\varrho > 0$, $M \in (0, \infty)$ such that

$$|F(x')_+ - F(x'')_+| \leqslant M |x' - x''|, \tag{22}$$

for all $x', x'' \in \cup_{2\varrho}(\bar{x})$, where $\cup_{2\varrho}(\bar{x}) = \{x \in R^n: |x - \bar{x}| \leqslant 2\varrho\}$. From (12)–(16) it follows that

$$v_k \to 0, \quad \text{as} \quad k \to +\infty, \quad k \in K, \tag{23a}$$

therefore there exists $\bar{k}$ such that

$$M v_k \leqslant 1/4, \quad v_k^2 \leqslant \varrho/2, \tag{23b}$$

for all $k \geqslant \bar{k}$, $k \in K$ and

$$x^k \in \cup_\varrho(\bar{x}), \tag{24}$$

for all $k \geqslant \bar{k}$. Now let us consider $x \in R^n$ with $|x-x^k| < 2v_k^2$. From (24) and (23b) we have

$$|x-\bar{x}| \leqslant |x-x^k| + |x^k-\bar{x}| \leqslant 2\varrho .$$

Then we obtain

$$|F(x)_+ - F(x^k)_+| \leqslant M\,|x^k-x| \leqslant 2Mv_k^2 \leqslant v_k/2, \tag{25}$$

for all $k \geqslant \bar{k}$, $k \in K$. On the other hand

$$F(x^k)_+ = F(x^k) > v_k, \tag{26}$$

for all $k \in K_c$. From (26) and (25) it follows that

$$F(x)_+ \geqslant v_k/2, \tag{27}$$

$$\text{for all } x \in R^n,\ x \in \mathrm{U}_{2v_k^2}(x^k), \quad k \geqslant \bar{k}, \quad k \in K.$$

Combining the definition of the function $F(\cdot)_+$, condition (13) and inequality (27), it is easily seen that

$$F(x^{k-1,i})_+ = F(x^{k-1,i}) \geqslant v_k/2, \quad i = 1, \ldots, l(k-1), \tag{28}$$

for all $k \geqslant \bar{k}$, $k \in K$. On the basis of the lemmas in [4] and from the description of the algorithm it is easy to see that

$$d^{k-1,l(k-1)} = d_1^{k-1,l(k-1)} + c_k\, d_2^{k-1,l(k-1)},$$

where $d_1^{k-1,l(k-1)} \in \partial f(x^k, 2v_k^2)$; $d_2^{k-1,l(k-1)} \in \partial F(x^k, 2v_k^2)$. Let us now prove that

$$|d_2^{k-1,l(k-1)}| \geqslant \bar{\varepsilon} > 0, \tag{29}$$

for k sufficiently large and $k \in K$. Assume, for contradiction purposes, that there exists an infinite subset $K' \subset K$ such that

$$\lim_{\substack{k \to +\infty \\ k \in K'}} |d_2^{k-1,l(k-1)}| = 0 .$$

From upper semicontinuity of the mapping: $(x, \eta) \to \partial F(x, \eta)$ it follows that $0 \in \partial F(\bar{x}, 0) = \partial F(\bar{x})$. On the other hand

$$F(\bar{x}) = \lim_{\substack{k \to +\infty \\ k \in K}} F(x^k) \geqslant \lim_{k \to +\infty} v_k = 0 .$$

Then we have derived a contradiction with the Slater condition. Therefore we obtain inequality (29). Combining (21), Assumption (2) and the local boundedness of the η-subdifferentials, it is easily seen that there exists $b > 0$ such that

$$|d_1^{k-1,l(k-1)}| \leqslant b, \quad \text{for all} \quad k \in K. \tag{30}$$

From the assumption of our lemma and (29), (30) it follows that

$$|d^{k-1,l(k-1)}| \to +\infty, \quad \text{as} \quad k \to +\infty, \quad k \in K. \tag{31}$$

From (6), (8), (12), (23a) and assumption of our lemma we obtain

$$\lim_{\substack{k \to +\infty \\ k \in K}} \varepsilon_{k-1} = 0. \tag{32}$$

From (32) and the description of the algorithm it follows that

$$\lim_{\substack{k \to +\infty \\ k \in K}} |d^{k-1,l(k-1)}| \leqslant \lim_{\substack{k \to +\infty \\ k \in K}} \varepsilon_{k-1} = 0,$$

so we have derived a contradiction with (31). This completes the proof. ■

Combining Lemmas 1 and 2 we shall prove the following theorem.

THEOREM 3. *Assume that conditions* (2), (4)–(7) *and the Slater condition are satisfied and let* $\{x^k\}_{k=0}^{\infty}$ *be the sequence generated by the algorithm. Then*

i) *There exists* k' *such that* $c_k = c_{k'}$ *for all* $k \geqslant k'$.

ii) *The sequence* $\{x^k\}$ *has an accumulation point, and every such accumulation point of* $\{x^k\}_{k=0}^{\infty}$ *solves problem* (1).

Proof. For purposes of a proof of assertion i) by contradiction, suppose that there exists an infinite subset $K_c = \{k\colon c_{k+1} > c_k\}$. From condition (14) we have $f(x^{k-1}) + c_k F(x^{k-1})_+ < L$, for any $k \in K_c$, where $F(x^{k-1})_+ \geqslant 0$ and $c_k > 0$. This implies

$$f(x^{k-1}) < L, \quad \text{for all} \quad k \in K_c. \tag{33}$$

Combining (33) and (2) it is easily seen that the set $\{x^{k-1}\colon k \in K_c\}$ is bounded. Therefore, from condition (13) it follows that set $\{x^k\colon k \in K_c\}$ is bounded, which means that there exists an accumulation point of $\{x^k\}_{k \in K_c}$. Then we have derived a contradiction with Lemma 2. Therefore, there exists k' such that $c_k = c_{k'}$, for all $k \geqslant k'$. From Lemmas 1 and 2 we obtain assertion ii) of Theorem 3. The theorem is proved. ■

4. Example

In this section we report on computational testing of the algorithm on two small problems. We used the following parameter values: $m_1 = 0.2$, $m_2 = 0.1$, $\beta_1 = \beta_2 = 0.3$, $\delta_k = 1/\sqrt[4]{k}$, $\alpha_k = 1/k^2$, $c_0 = 1$.

The first example is given by

$$f(x) = \max\{f_{i,100}(x)\colon i = 2, 3\}; \qquad F(x) = f_{51,100}(x),$$

$x \in R^2$, with the solution $\hat{x} = (0, 0)$; $f(\hat{x}) = 0$, where

$$f_{i,j} = x_1 \cos [2\pi (i-1)/j] + x_2 \sin [2\pi (i-1)/j].$$

Starting from $x^0 = (1, -1)$ with $P(x^0, c_0) = 0.935$; we obtained $k = 31$, $\sum_{k=1}^{31} l(k) = 235$, $P(x^{31}, c_{31}) = 7 \cdot 10^{-6}$.

The second example is

$$f(x) = x_1^2 + x_2^2 + 2x_3^2 + x_4^2 - 5x_1 - 5x_2 - 21x_3 + 7x_4$$

$$F(x) = \max \{f_i(x): i = 1, 2, 3\}, \quad x \in R^4$$

$$f_1(x) = 2x_1^2 + x_2^2 + x_3^2 + 2x_1 - x_2 - x_4 - 5,$$

$$f_2(x) = x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_1 - x_2 + x_3 - x_4 - 8,$$

$$f_3(x) = x_1^2 + 2x_2^2 + x_3^2 + 2x_4^2 - x_1 - x_4 - 10,$$

with the solution $\hat{x} = (0, 1, 2, -1)$; $f(\hat{x}) = -44$. Starting from $x^0 = (0, 0, 0, 0)$ with $P(x^0, c_0) = 0$, we obtained $k = 25$, $\sum_{k=1}^{25} l(k) = 182$; $P(x^{25}, c_{25}) = -43.997$.

References

[1] Bertsekas D. Necessary and sufficient conditions for a penalty method to be exact. *Math. Program.* **9** (1975), 87–89.

[2] Kiwiel K. C. An exact penalty function algorithm for nonsmooth convex minimization problems. *IMA Journal of Numerical Analysis*, **5** (1985), 111–119.

[3] Mifflin R. An algorithm for constrained optimization with semismooth functions. *Math. of Operations Research*, **2** (1977), 191–207.

[4] Pham T. Q. A subgradient method with space dilation for minimizing convex functions. *Control and Cybernetics*, **13** (1984), 4, 399–408.

[5] Polak E., Mayne D. Q., Wardi Y. On the extension of constrained optimization algorithms from differentiable to nondifferentiable problems. *SIAM J. Control Optim.* **21** (1983), 179–203.

[6] Shor N. Z. Methods for Minimizing Nondifferentiable Functions and Their Applications. Kiev, Naukova Dumka, 1979.

[7] Wolfe P. A Method of conjugate subgradients for minimizing nondifferentiable convex functions. Math. Program. Study 3, M. L. Balinsky and P. Wolfe eds. North-Holland, Amsterdam, 1975, pp. 145–173.

Received, January 1986

Metoda funkcji kary wykorzystująca operację rozciągania przestrzeni dla wypukłych zadań minimalizacji z ograniczeniami

W artykule tym wprowadziliśmy nowy algorytm rozwiązywania zadań minimalizacji z ograniczeniami nieliniowymi. Rozszerza on na przypadek niegładki metodę funkcji kary.

Osiągnięto to poprzez wprowadzenie reguły zwiększenia współczynnika kary. Reguła ta gwarantuje skończoność współczynnika kary w algorytmie.

Метод штрафа для выпуклых задач минимизации с ограничениями использующий операцию растягивания пространства

В работе представлен новый алгоритм решения задач минимизации с нелинейными ограничениями. Вводя закон конечного увеличения коэффициента штрафа, метод штрафных функций расширяется на случай негладких задач.