Journal of Mathematical Sciences, Vol. 73, No. 5, 1995

RELAXATION METHODS WITH STEP REGULATION FOR SOLVING CONSTRAINED OPTIMIZATION PROBLEMS

Z. R. Gabidullina

UDC 519.853

We study the problem

$$\min_{x\in D} f(x), \tag{1}$$

where $f(x)$ is a continuously differentiable pseudoconvex function satisfying condition A (introduced in [2]) on a convex closed set $D \subset \mathbb{R}_n$.

We use the following notation: $g(x)$ is the gradient of the function $f(x)$ at the point x, x_0 is the point of initial approximation, $f^* = \min_{x\in D} f(x)$, $X^* = \{x^* \in D : f(x^*) = f^*\}$, $L = \{0, 1, \ldots\}$, and p_k^* is the projection of the recursion point x_k on the set X^*.

Definition 1 [2]. A continuous function f satisfies *Condition A* on the set D if there exists $\varkappa > 0$ such that

$$f(\alpha x + (1-\alpha)y) \geq \alpha f(x) + (1-\alpha)f(y) - \alpha(1-\alpha)\varkappa\|x-y\|^2 \quad \forall x, y \in D, \ \forall \alpha \in [0,1].$$

It was shown in [2] that for a continuously differentiable convex function satisfying Condition A on a convex set D the following estimate holds:

$$f(x) - f(y) \geq \langle g(x), x - y\rangle - \varkappa\|x-y\|^2 \quad \forall x, y \in D. \tag{2}$$

Definition 2. The direction $s \neq 0$ is an *ε-normalized direction of descent* at the point x if s can be chosen so as to satisfy the condition

$$\langle g(x), s\rangle + \varepsilon\|s\|^* \leq 0.$$

Taking account of the estimate (2), one can show that an ε-normalized direction of descent s at the point x for the function $f(x)$ has the following properties:

1) $f(x) - f(x+\lambda s) \geq -\lambda(1-\lambda)\langle g(x), s\rangle > 0$ for all $\lambda \in (0,1)$,

2) $f(x) - f(x+\lambda s) \geq \lambda(1-\lambda)\varkappa\|s\|^2$ for all $\lambda \in (0,1)$,

3) if $f(x)$ is a convex function, then there exists $\beta \in [0,1]$ such that the inequality $f(x) - f(x+s) = -\langle g(x), s\rangle - \beta\varkappa\|s\|^2 \geq 0$ holds.

In what follows we propose an relaxation method of solving the problem (1). Depending on the method of choosing the size of the recursion step, one can obtain various versions of the algorithm. The step can be chosen, for example, as follows [5]. Choose $\lambda \in (0,1)$ and determine the first index $i = 1, 2, \ldots,$ at which the following inequality holds:

$$f(x_k) - f(x_k + \lambda^i s_k) \geq -\lambda^i(1-\lambda)\langle g(x_k), s_k\rangle. \tag{3}$$

In the present paper we propose to choose the step by dividing the quantity λ until the first time the following weaker condition holds:

$$f(x_k) - f(x_k + \lambda^i s_k) \geq \lambda^i(1-\lambda)\varepsilon\|s_k\|^2. \tag{4}$$

Moreover, in the algorithms being studied the step λ_k is controlled by the ε-normalization of the direction of descent s_k.

Translated from *Issledovaniya po Prikladnoi Matematike*, No. 18, 1992, pp. 30–38.

 1072-3374/95/7305-0538 $12.50

It is assumed that the point $x_0 \in D$ and the numbers $\lambda \in (0,1)$, $\varepsilon > 0$, and σ $(0 < \sigma \leq 1)$ are known. In what follows the recursive sequence is generated according to one of the following algorithms.

Algorithm 1 (a method with adaptation of the parameter ε_k). Suppose the constant $\varepsilon_0 > 0$ has been chosen.

0. At the point $x_k \in D$, $k = 0, 1, \ldots$, compute a point y_k such that

$$\langle g(x_k), y_k - x_k \rangle + \varepsilon_k \|y_k - x_k\|^2 \leq \sigma_k \min_{x \in D} \left(\langle g(x_k), x - x_k \rangle + \varepsilon_k \|x - x_k\|^2 \right).$$

$0 < \sigma \leq \sigma_k \leq 1$. If $\langle g(x_k), y_k - x_k \rangle + \varepsilon_k \|y_k - x_k\|^2 = 0$, then x_k is a solution of the problem (1); otherwise set $s_k = y_k - x_k$ and go to Step 1.

1. Let i_k be the first index $i = 1, 2, \ldots$, at which inequality (3) holds (or inequality (4)); then set $\lambda_k = \lambda^{i_k}$. Compute the point $x_{k+1} = x_k + \lambda_k s_k$, $k = k + 1$, and go to Step 2.

2. If $i_k = 1$, then set $\varepsilon_{k+1} = \varepsilon_k$. If $i_k > 1$, then set $\varepsilon_{k+1} = \varepsilon_k / \lambda^{i_k - 1}$ and go to Step 0.

Algorithm 2 differs from Algorithm 1 in step 2: $\varepsilon_k = \varepsilon_0$ for all $k \in L$. We call Algorithm 2 *nonrelaxation*, since the parameter ε_k in the algorithm remains constant.

Remark 1. It is obvious that the direction $s_k = y_k - x_k$ is a $\varkappa$-normalized direction for all k. Indeed,

$$\langle g(x_k), y_k - x_k \rangle + \varkappa \|y_k - x_k\|^2 \leq \sigma_k \min_{x \in D} \langle g(x_k), x - x_k \rangle + \varkappa \|x - x_k\|^2 \leq \sigma \left(\langle g(x_k), x - x_k \rangle + \varkappa \|x - x_k\|^2 \right) \quad \forall x \in D.$$

from which, for $x = x_k$, we find that $\langle g(x_k), y_k - x_k \rangle + \varkappa \|y_k - x_k\|^2 \leq 0$.

Optimality Theorem 1. *Suppose*

1) *$f(x)$ is a continuously differentiable pseudoconvex function on the set D;*

2) *there exists $\tilde{f}(x)$, a pseudoconvex Lipschitz function, that extends $f(x)$ to the whole space $\mathbb{R}_n$.*

A necessary and sufficient condition for $x_k \in D$ to be a minimum of $f(x)$ on D is the inequality

$$\langle g(x_k), x - x_k \rangle \geq 0 \quad \forall x \in D. \tag{5}$$

Proof. The necessity is proved in analogy with Theorem 3 of [1, pp. 171–172].

Sufficiency. At the point x_k inequality (5) holds for all $x \in D$. Assume that x_k is not a minimum, i.e., there exists $\tilde{x} \in D$ such that $f(\tilde{x}) < f(x_k)$. Then the following inequality holds (cf. [4, pp. 47–48; 6]:

$$0 < f(x_k) - f(\tilde{x}) \leq M \left\langle \frac{g(x_k)}{\|g(x_k)\|}, x_k - \tilde{x} \right\rangle,$$

where M is the Lipschitz constant for the function $\tilde{f}(x)$; hence $\langle g(x_k), \tilde{x} - x_k \rangle < 0$, contrary to the hypothesis of the theorem. ■

Optimality Theorem 2. *Suppose the hypotheses of Optimality Theorem 1 hold. Then a necessary and sufficient condition for the point $x_k \in D$ to be a minimum of the function $f(x)$ on D is the following:*

$$\exists c > 0: \ \langle g(x_k), x - x_k \rangle + \varepsilon \|x - x_k\|^2 \geq 0 \quad \forall x \in D. \tag{6}$$

Proof. *Necessity.* Let x_k be a minimum of the function $f(x)$ on the set D. It is required to show that (6) holds. By Optimality Theorem 1 we have

$$\langle g(x_k), x - x_k \rangle + \varepsilon \|x - x_k\|^2 \geq \langle g(x_k), x - x_k \rangle \geq 0 \quad \forall x \in D.$$

Sufficiency. Assume that x_k is not a minimum point and $\tilde{x} \in D$ is such that $f(\tilde{x}) < f(x_k)$. Then according to Optimality Theorem 1 we have $\langle g(x_k), \tilde{x} - x_k \rangle < 0$. Let t be a number satisfying $0 < t < |\langle g(x_k), \tilde{x} - x_k \rangle|$. It follows from inequality (6) that $t/(\varepsilon \|\tilde{x} - x_k\|^2) < 1$. It is then easy to show that $s = t(\tilde{x} - x_k)/(\varepsilon\|\tilde{x} - x_k\|^2)$ is an ε-normalized vector, and that

$$\langle g(x_k), s \rangle + \varepsilon \|s\|^2 = \frac{t}{\varepsilon \|\tilde{x} - x_k\|^2} \big(\langle g(x_k), \tilde{x} - x_k \rangle + t \big) < 0, \quad \|s\| < \|\tilde{x} - x_k\|,$$

i.e., there exists a point $z_k = x_k + s$, $z_k \in [\tilde{x}, x_k]$ satisfying the inequality $\langle g(x_k), z_k - x_k \rangle + \varepsilon \|z_k - x_k\|^2 < 0$, and this contradicts (6). ■

Optimality Theorem 3. *Suppose the hypotheses of Optimality Theorem 1 holds. A necessary and sufficient condition for the point $x_k \in D$ to be a minimum of $f(x)$ on the set D is that*

$$\exists \varepsilon > 0: \ \ \langle g(x_k), y_k - x_k \rangle + \varepsilon \|y_k - x_k\|^2 = 0. \tag{7}$$

Proof. *Necessity.* By the choice of the direction of descent we have the inequality

$$\begin{aligned} \langle g(x_k), y_k - x_k \rangle + \varepsilon \|y_k - x_k\|^2 &\le \sigma_k \min_{x \in D} \big(\langle g(x_k), x - x_k \rangle + \varepsilon_k \|x - x_k\|^2 \big) \\ &\le \sigma \big(\langle g(x_k), x - x_k \rangle + \varepsilon_k \|x - x_k\|^2 \big) \ \ \forall x \in D. \end{aligned}$$

Hence for $x = x_k$ we obtain $\langle g(x_k), y_k - x_k \rangle + \varepsilon \|y_k - x_k\|^2 \le 0$. Then Eq. (7) holds, since the inequality $\langle g(x_k), y_k - x_k \rangle + \varepsilon \|y_k - x_k\|^2 < 0$ would contradict the fact that x_k is a minimum point.

Sufficiency. By the choice of the direction of descent we have $\langle g(x_k), x - x_k \rangle + \varepsilon_k \|x - x_k\|^2 \ge 0$ for all $x \in D$. Therefore by Optimality Theorem 2 we conclude that x_k is a minimum. ■

Remark 2. If s_k is a $\varkappa$-normalized direction of descent, then $i_k = 1$ for all $k \in L$, i.e., $\lambda_k = \lambda$ for $\lambda \in (0,1)$. It is obvious that if s_k is an ε-normalized direction of descent for $k \in L$ with $\varepsilon \ge \varkappa$, then s_k is also a $\varkappa$-normalized direction.

We shall now study the case when s_k is an ε-normalized direction of descent, but is not $\varkappa$-normalized.

Lemma 1. *Suppose*

1) $0 < \varepsilon < \varkappa$, $\lambda \in (0,1)$,

2) *$\{x_k\}$, $k \in L^*$, is a subsequence of $\{x_k\}$, $k \in L$, such that $\{s_k\}$, $k \in L^*$, are ε-normalized directions of descent, but not $\varkappa$-normalized directions, and that λ_k is chosen according to condition* (4).

Then the estimate $\lambda_k > \lambda^2 \varepsilon \varkappa^{-1}$ holds for all $k \in L^*$.

Proof. If $i_k = 1$ for all $k \in L^*$, then $\lambda_k = \lambda$, and the assertion of the lemma holds. Let $\widehat{L} \subset L^*$ be a sequence such that $i_k \ne 1$ for all $k \in \widehat{L}$. Since the step size was divided ($i_k \ne 1$), the quantity $\lambda^{i_k - 1}$ does not satisfy inequality (4). Then $\lambda^{i_k-1} > \lambda \varepsilon \varkappa^{-1}$. If we assume the contrary, i.e., $\lambda^{i_k - 1} \le \lambda \varepsilon \varkappa^{-1}$, we obtain

$$\begin{aligned} f(x_k) - f(x_k + \lambda^{i_k-1} s_k) &\ge \lambda^{i_k-1} (-\langle g(x_k), s_k \rangle - \lambda^{i_k-1} \varkappa \|s_k\|^2) \\ &\ge \lambda^{i_k-1} (-\langle g(x_k), s_k \rangle - \lambda \varepsilon \|s_k\|^2) \ge \lambda^{i_k-1} (1-\lambda) \varepsilon \|s_k\|^2. \end{aligned}$$

and this contradicts the fact that i_k is the first index for which inequality (4) holds. Hence $\lambda_k = \lambda^{i_k} > \lambda^2 \varepsilon \varkappa^{-1}$ for all $k \in L^*$. ■

It can be shown that if the step is chosen according to condition (3), then it is also bounded below.

Lemma 1 now implies the estimate $\varkappa > \varepsilon / \lambda^{i_k - 2}$. This estimate is used in Algorithm 1 for adaptation of the parameter ε_k.

Lemma 2. *If*

1) *$f(x)$ is a convex function that is defined on a convex closed set $D \subset \mathbb{R}_n$,*

2) X^* *is a nonempty bounded set,*

then for any relaxation sequence $\{x_k\}$, $k \in L$, $x_k \in D$, *the inequalities* $\|x_k - p_k^*\| \le \eta < \infty$ *will hold.*

Proof. Let $C = f(x_0) - f^*$, where $f^* = f(p_k^*)$. Then $f(x_k) - f^* \le C$ for all $k = 0, 1, 2, \ldots$, since $f(x_{k+1}) \le f(x_k)$. We assume that the assertion of the lemma is false, i.e., there exists a sequence $\{x_k\}$ for which $\|x_k - p_k^*\| \to \infty$ as $k \to \infty$. Then for an arbitrary $\varepsilon > 0$ there exist indices k such that $\rho(x_k, X^*) = \|x_k - p_k^*\| \ge \varepsilon$. For these k we consider $y_k = \varepsilon\|x_k - p_k^*\|^{-1}x_k - (1 - \varepsilon\|x_k - p_k^*\|^{-1})p_k^*$. It is obvious that $y_k \in [p_k^*, x_k]$ and $\rho(y_k, X^*) = \|y_k - p_k^*\| = \varepsilon$, i.e., $y_k \in G_\varepsilon = \{x \in D : \rho(x, X^*) = \varepsilon\}$. The set G_ε is bounded and closed, and consequently $\min\limits_{z \in G_\varepsilon}(f(z) - f^*) = \xi > 0$. Therefore

$$0 < \xi \le f(y_k) - f^*. \tag{8}$$

It follows from the convexity of the function $f(x)$ that

$$f(y_k) \le \varepsilon\|x_k - p_k^*\|^{-1} f(x_k) - (1 - \varepsilon\|x_k - p_k^*\|^{-1}) f(p_k^*) = f^* + \varepsilon\|x_k - p_k^*\|^{-1}(f(x_k) - f^*) \le f^* + \varepsilon\|x_k - p_k^*\|^{-1}.$$

Since we are assuming that $\|x_k - p_k^*\| \to +\infty$ as $k \to +\infty$, we find that $\lim\limits_{k \to \infty} f(y_k) = f^*$, and this contradicts (8). ∎

If the sequence $\{x_k\}$ constructed using Algorithm 2 is finite, then a solution of the problem (1) will have been obtained by construction. But if the sequence $\{x_k\}$ is infinite, the following theorem establishes conditions for it to converge to a solution of the problem (1).

Convergence Theorem. *If*

1) $f(x)$ *is a pseudoconvex function satisfying condition* A *on a convex compact set* D *with constant* $\varkappa > 0$,

2) *there exists* $\tilde{f}(x)$, *a pseudoconvex Lipschitzian function, that extends the function* $f(x)$ *to the whole space* $\mathbb{R}_n$,

3) *there exist* $\theta, \gamma > 0$ *such that* $\theta \le \|g(x)\| \le \gamma < \infty$ *for all* $x \in D$,

4) *the step size* λ_k *is chosen according to condition* (3),

then the sequence $\{x_k\}$, $k \in L$, *converges as a functional, and*

$$f(x_k) - f^* = O(1/k).$$

Proof. It is obvious that $f(p_k^*) = f^*$. By the pseudoconvexity of $f(x)$ we have the following inequalities [3, pp. 47–48; 5]:

$$0 < f(x_k) - f(p_k^*) \le M\left\langle \frac{g(x_k)}{\|g(x_k)\|}, x_k - p_k^* \right\rangle \le M\theta^{-1}\langle g(x_k), x_k - p_k^*\rangle \quad \forall\, k \in L,$$

where M is the Lipschitz constant of the function $\tilde{f}(x)$. Therefore $\langle g(x_k), p_k^* - x_k\rangle < 0$ for all $k \in L$. Let $L_1 \subset L$ be the set of subscripts k such that the inequalities $\langle g(x_k), p_k^* - x_k\rangle + \frac{\varepsilon}{\lambda}\|p_k^* - x_k\|^2 \le 0$ hold and $L_2 \subset L$ the set of subscripts k for which the opposite inequality holds. It is easy to show that the vector

$$q_k = \begin{cases} p_k^* - x_k & \forall\, k \in L_1, \\ \frac{\lambda t_k (p_k^* - x_k)}{\varepsilon\|x_k - p_k^*\|}, & \forall\, k \in L_2,\ 0 < t_k = |\langle g(x_k), p_k^* - x_k\rangle|, \end{cases}$$

is (ε/λ)-normalized. Then the point $z_k = x_k + q_k$, $k \in L$, is such that

1) $z_k \in [x_k, p_k^*]$, since $\|z_k - x_k\| \le \|x_k - p_k^*\|$,

2) $\langle g(x_k), z_k - x_k\rangle + \varepsilon\lambda^{-1}\|z_k - x_k\|^2 \le 0$.

I. Let us consider the case $0 < \varepsilon \le \varkappa$. We distinguish $L_3 \subset L$, the set of subscripts such that s_k, is a $\varkappa$-normalized direction of descent for $k \in L_3$. We then have

$$f(x_k) - f(x_{k+1}) \ge -\lambda(1-\lambda)\langle g(x_k), s_k\rangle \ge -\lambda(1-\lambda)(\langle g(x_k), s_k\rangle + \varepsilon\|s_k\|^2) \ \forall k \in L_3. \tag{9}$$

Let $L_4 = L \setminus L_3$. Then, taking account of the assertion of Lemma 1, we have

$$\begin{aligned} f(x_k) - f(x_{k+1}) &\ge -\lambda_k(1-\lambda)\langle g(x_k), s_k\rangle > -\lambda^2\varepsilon\varkappa^{-1}(1-\lambda)\langle g(x_k), s_k\rangle \\ &\ge -\lambda^2\varepsilon\varkappa^{-1}(1-\lambda)(\langle g(x_k), s_k\rangle + \varepsilon\|s_k\|^2) \ \forall k \in L_4. \end{aligned} \tag{10}$$

From estimates (9) and (10) we obtain

$$f(x_k) - f(x_{k+1}) \ge -C_1(\langle g(x_k), s_k\rangle + \varepsilon\|s_k\|^2) \ \forall k \in L,$$

where $C_1 = \lambda^2\varepsilon\varkappa^{-1}(1-\lambda)$. Extending this estimate by taking account of the choice of the direction of descent s_k, we obtain

$$\begin{aligned} f(x_k) - f(x_{k+1}) &\ge -C_1(\langle g(x_k), y_k - x_k\rangle + \varepsilon\|y_k - x_k\|^2) \ge -C_1\sigma_k \min_{x\in D}(\langle g(x_k), x - x_k\rangle + \varepsilon\|x - x_k\|^2) \\ &\ge -C_1\sigma(\langle g(x_k), x - x_k\rangle + \varepsilon\|x - x_k\|^2) \quad \forall x \in D, \ \forall k \in L. \end{aligned} \tag{11}$$

II. In the case $\varepsilon > \varkappa > 0$ we have, in analogy with estimate (11),

$$\begin{aligned} f(x_k) - f(x_{k+1}) &\ge -\lambda(1-\lambda)(\langle g(x_k), s_k\rangle + \varepsilon\|s_k\|^2) \\ &\ge -\lambda(1-\lambda)\sigma(\langle g(x_k), x - x_k\rangle + \varepsilon\|x - x_k\|^2) \ \forall x \in D, \forall k \in L. \end{aligned} \tag{12}$$

Thus it follows from estimates (11) and (12) that there exists a constant $C > 0$ such that the following inequalities hold:

$$f(x_k) - f(x_{k+1}) \ge -C(\langle g(x_k), x - x_k\rangle + \varepsilon\|x - x_k\|^2) \ \forall x \in D, \forall k \in L.$$

For $x = z_k$ we obtain in particular from this inequality

$$f(x_k) - f(x_{k+1}) \ge -C(\langle g(x_k), z_k - x_k\rangle + \varepsilon\|z_k - x_k\|^2) \ge -C(1-\lambda)\langle g(x_k), q_k\rangle \ \forall k \in L.$$

Then if $\eta = \sup_{x,y\in D} \|x - y\|$, we have the inequalities

$$f(x_k) - f(x_{k+1}) \ge C(1-\lambda)\langle g(x_k), x_k - p_k^*\rangle \ge C(1-\lambda)(\eta\gamma)^{-1}\langle g(x_k), x_k - p_k^*\rangle^2 \ \forall k \in L_1. \tag{13}$$

We have the following inequalities:

$$f(x_k) - f(x_{k+1}) \ge C(1-\lambda)\frac{\lambda t_k}{\|p_k^* - x_k\|^2}\langle g(x_k), x_k - p_k^*\rangle \ge \frac{C\lambda(1-\lambda)}{\varepsilon\eta^2}\langle g(x_k), x_k - p_k^*\rangle^2 \ \forall k \in L_2. \tag{14}$$

It follows from estimates (13) and (14) that

$$f(x_k) - f(x_{k+1}) \ge C_2\langle g(x_k), x_k - p_k^*\rangle^2 \ \forall k \in L,$$

where $C_2 = C(1-\lambda)\eta^{-1}\min\{\gamma^{-1}, (\varepsilon\eta)^{-1}\} > 0$. Let $C_3 = \theta^2 M^{-2} C_2$. Then $f(x_k) - f(x_{k+1}) \ge C_3(f(x_k) - f(p_k^*))^2$. Therefore according to Lemma 4 of [1, p. 102] the sequence $\{x_k\}$, $k \in L$, converges in the functional sense: $f(x_k) - f^* \le C_3^{-1}k^{-1}$. ■

The convergence of the method when the step size is chosen from the condition (4) is proved similarly. Here the following estimate is used in the proof: $f(x_k) - f(x_{k+1}) \ge -\lambda_k(1 + \lambda_k\varkappa(\varepsilon(1-\lambda))^{-1})^{-1}\langle g(x_k), s_k\rangle$.

We remark that if the function $f(x)$ to be minimized is convex, we need not require the set D to be bounded in order to obtain an estimate of the rate of convergence; it suffices that the set X^* be bounded. In this situation Condition 2 in the convergence theorem is superfluous.

It follows from the condition $\varkappa < +\infty$ that after a finite number of increases of ε_k in Algorithm 1 this quantity exceeds $\varkappa$ and ceases to change. From that point on the relaxation algorithm begins to work with a fixed constant $\varepsilon \geq \varkappa$; hence the convergence of Algorithm 1 follows from the convergence theorem. We remark that from that point on $\lambda_k = \lambda$, $k = 0, 1, \ldots$. Here in order to compute the recursion step only two computations of the objective function are carried out.

Literature Cited

1. F. P. Vasil'ev. *Numerical Methods of Solving Extremal Problems* [in Russian], Nauka, Moscow (1980).
2. Z. R. Gabidullina, "On the convergence of the constrained gradient method of a class of nonconvex functions," *Issled. po Priklad. Mat.*, No. 14, 15–25 (1987).
3. Ya. I. Zabotin, A. I. Korablev, "Pseudoconvex functionals and their extremal properties," *Izv. Vuzov. Mat.*, No. 4, 27–31 (1974).
4. I. V. Konnov, "Application of the method of conjugate subgradients to minimizing quasiconvex functionals," *Issled. po Priklad. Mat.*, No. 12, 46–58 (1984).
5. B. N. Pshenichnyi and Yu. M. Danilin, *Numerical Methods in Extremal Problems* [in Russian], Nauka, Moscow (1975).
6. F. Plastria, "Lower subdifferentiable functions and their minimization by cutting planes," *J. Opt. Th. Appl.*, **46**, No. 1, 37–53 (1985).