ON THE CHOICE OF A PARAMETER IN A MINIMIZATION PROBLEM

M. D. Maergoiz

1969

SovietRxiv

Abstract

Full Text

UDC 518:512.25

ON THE CHOICE OF A PARAMETER IN A MINIMIZATION PROBLEM

M. D. Maergoiz

(Presented by Academician L. V. Kantorovich on January 6, 1969)

When applying the method of steepest descent, first proposed in ([1]), to the minimization of nonquadratic functionals, it is necessary at each step to determine a parameter by solving a nonlinear equation, i.e., by carrying out an infinite computational procedure. In this connection, in ([2], [3]) special methods were proposed for choosing the parameter that ensure convergence. In the present note a very simple device is proposed for choosing the parameter (for example, on the basis of bisection) for minimizing a strongly convex function in a finite-dimensional space; the convergence of the iterative processes thereby obtained is proved, and a rapidly convergent "hybrid" method is constructed.

Let

$$X \in E^n; \qquad f(X) \in C^2; \qquad -\infty < f(X^*) = \min_X f(X) < +\infty; \tag{1}$$

the initial approximation X^0 is arbitrary.

Introduce the set $\mathfrak{M}_X = \{X : f(X) \leq f(X^0)\}$. Suppose that

$$m\|u\|^2 < (H(X)u, u) \leq \|H(x)\| \cdot \|u\|^2 < M\|u\|^2; \tag{2}$$

for all $X \in \mathfrak{M}_X,\ u \in E^n$, where

$$H(X) = \left(\frac{\partial^2 f(X)}{\partial x_i \partial x_j}\right)_{i,j=1}^{n}.$$

We describe a step of any of the iterative processes:

$$\begin{cases} t_0^k = t > 0; \quad X^{-t_j^k} = X^k - t_j^k l^k; \\ \text{if } f(X^{-t_j^k}) \geq f(X^k), \text{ then } t_{j+1}^k = t_j^k / p; \end{cases} \tag{3}$$

otherwise:

$$t^k = t_j^k; \qquad \tau^k = t^k/p; \qquad p > 1; \qquad X^{k+1} = X^k - \tau^k l^k, \tag{4}$$

where l^k satisfies the relation

$$\gamma_1 \|\nabla f(X^k)\|^2 < (\nabla f(X^k), l^k) \le \|\nabla f(X^k)\| \cdot \|l^k\| < \gamma_2 \|\nabla f(X^k)\|^2; \tag{5}$$

$$0 < \gamma_1; \qquad k = 0, 1, 2, \ldots$$

For $p = 2$ we obtain the bisection device.

Theorem 1. *If conditions (1), (2) are fulfilled, then any of the iterative processes (3), (4), (5) converges, i.e.*

$$\lim_{k\to\infty} \|X^k - X^*\| = 0. \tag{6}$$

Proof. For any X^k ($\|\nabla f(X^k)\| > 0$), consider

$$\varphi_k(t) = (\nabla f(X^k - tl^k), l^k); \qquad \varphi_k(0) = (\nabla f(X^k), l^k) > \gamma_1 \|\nabla f(X^k)\|^2 > 0. \tag{7}$$

According to (1), (2),

$$\varphi_k'(t) = -(H(X^k - tl^k) l^k, l^k) < 0, \tag{8}$$

the function $\varphi_k(t)$ is monotonically decreasing and has one positive root a^k. From (3), (4),

$$f(X^k - t^k l^k) - f(X^k) = -\int_0^{t^k} (\nabla f(X^k - tl^k), l^k)\, dt = -\int_0^{t^k} \varphi_k(t)\, dt < 0, \tag{9}$$

$$\int_0^{t^k} \varphi_k(t)\, dt > 0.$$

Taking into account (7), (8),

$$\int_0^{\tau^k} \varphi_k(t)\, dt = \int_0^{t^k/p} \varphi_k(t)\, dt > 0;$$

$$f(X^{k+1})-f(X^k) = f(X^k-\tau^k l^k)-f(X^k) = -\int_0^{\tau^k} \varphi_k(t) < 0; \qquad f(X^{k+1}) < f(X^k). \tag{10}$$

Thus, the sequence $\{f(X^k)\}$ is monotonically decreasing and bounded below $(f(X^k) \geq f(X^*))$.

Suppose that the sequence $\{f(X^k)\}$ does not converge to the minimum value of the function, i.e.,

$$\lim_{k\to\infty} f(X^k) = \bar{f} > \min_X f(X) = f(X^*). \tag{11}$$

In this case there is a $\Delta > 0$ such that for all k it will be true that

$$\|\nabla f(X^k)\| > \Delta. \tag{12}$$

In order to obtain a contradiction to assumption (11), we first estimate the parameter τ^k.

Taking into account (3), (4), $f(X^k - t^k l^k) < f(X^k)$; $f(X^k - pt^k l^k) \geq f(X^k)$. By continuity there is a number

$$\lambda^k = (1 + \theta^k(p-1))t^k; \qquad 0 < \theta^k \leq 1, \tag{13}$$

such that

$$f(X^k - \lambda^k l^k) = f(X^k); \qquad X^k - \lambda^k l^k \in \mathfrak{M}_X;$$

$$f(X^k - \lambda^k l^k) - f(X^k) = -\lambda^k(\nabla f(X^k), l^k) + \frac{(\lambda^k)^2}{2}(H(y^k)l^k, l^k) = 0; \tag{14}$$

$$\lambda^k = \frac{2(\nabla f(X^k), l^k)}{(H(y^k)l^k, l^k)}; \qquad Y^k = X^k - \omega^k\lambda^k l^k \in \mathfrak{M}_X; \qquad 0 < \omega^k < 1.$$

In connection with (2), (3), (4), (5), (13), (14),

$$\frac{2\gamma_1}{\gamma_2^2 M} < \lambda^k < \frac{2\gamma_2}{\gamma_1^2 m}; \qquad \min\left(t, \frac{2\gamma_1}{p\gamma_2^2 M}\right) < t^k < \frac{2\gamma_2}{\gamma_1^2 m};$$

$$\min\left(\frac{t}{p}, \frac{2\gamma_1}{p^2\gamma_2^2 M}\right) < \tau^k < \frac{2\gamma_2}{p\gamma_1^2 m}. \tag{15}$$

Let us estimate from below

$$\int_0^{\tau^k} (\nabla f(X^k - tl^k), l^k)\,dt = \int_0^{\tau^k} \varphi_k(t)\,dt.$$

On the basis of (1), (2), (4), (5), (13), (14),

$$-M\gamma_2^2\|\nabla f(X^k)\|^2 < \dot\varphi_k(t) = -(H(X^k - tl^k)l^k, l^k) < -m\gamma_1^2\|\nabla f(X^k)\|^2;$$

$$0 \le t \le \lambda^k; \qquad X^k - tl^k \in \mathfrak{M}_X. \tag{16}$$

$$-M\gamma_2^2\|\nabla f(X^k)\|^2 t + \varphi_k(0) \le \varphi_k(t) \le -m\gamma_1^2\|\nabla f(X^k)\|^2 t + \varphi_k(0); \qquad 0 \le t \le \tau^k;$$

$$-M\gamma_2^2\|\nabla f(X^k)\|^2 (t - \tau^k) + \varphi_k(\tau^k) \le \varphi_k(t) \le$$

$$\le -m\gamma_1^2\|\nabla f(X^k)\|^2 (t - \tau^k) + \varphi_k(\tau^k); \qquad \tau^k \le t \le \lambda^k. \tag{17}$$

Suppose first that $\tau^k \leqslant a^k$. According to (7), (8), (14), (16), (12)

$$\int_0^{\tau^k} \varphi_k(t)\,dt > \int_0^{\min(t/p,\, 2\gamma_1/p^2\gamma_2^2 M)} \left(-M\gamma_2^2\|\nabla f(X^k)\|^2 t + \varphi_k(0)\right)\,dt >$$

$$> \int_0^{\min(t/p,\, 2\gamma_1/p^2\gamma_2^2 M)} \left(-M\gamma_2^2\|\nabla f(X^k)\|^2 t + \min(t\gamma_2^2 M/2\gamma_1,\, \frac{1}{p})\varphi_k(0)\right)\,dt >$$

$$> \min\left(\frac{2\gamma_1^2}{p^4\gamma_2^2 M},\, \frac{1}{2}\frac{t^2\gamma_2^2 M}{p^2}\right)\Delta^2 = S_1. \tag{18}$$

Now suppose that $\tau^k \geqslant a^k$. According to (7), (8), (15), 17) $\varphi_k(\tau^k) \leqslant 0$;

$$\int_0^{\tau^k} \varphi_k(t)\,dt = \int_0^{\lambda^k} \varphi_k(t)\,dt - \int_{\tau^k}^{\lambda^k} \varphi_k(t)\,dt = f(X^k) - f(X^k - \lambda^k l^k) -$$

$$-\int_{\tau^k}^{\lambda^k} \varphi_k(t)\,dt = -\int_{\tau^k}^{\lambda^k} \varphi_k(t)\,dt \geqslant \int_{\tau^k}^{\lambda^k} \left(m\gamma_1^2\|\nabla f(X^k)\|^2(t-\tau^k) - \varphi_k(\tau^k)\right)\,dt \geqslant$$

$$\geqslant \int_{\tau^k}^{\lambda^k} m\gamma_1^2\|\nabla f(X^k)\|^2(t-\tau^k)\,dt = \frac{m\gamma_1^2\|\nabla f(X^k)\|^2}{2}(\lambda^k-\tau^k)^2 >$$

$$> \frac{m\gamma_1^2}{2}(t^k-\tau^k)^2\Delta^2 = \frac{m\gamma_1^2(p-1)^2}{2}\left(\min\left(\frac{t}{p}, \frac{2\gamma_1}{p^3\gamma_2^2 M}\right)\right)^2\Delta^2 = S_2. \quad (19)$$

Finally,

$$\int_0^{\tau^k} \varphi_k(t)\,dt > \min(s_1, s_2) = s > 0;$$

$$f(X^{k+1}) = f(X^k) - \int_0^{\tau^k} \varphi_k(t)\,dt < f(X^k) - s < f(X^0) - (k+1)s; \quad (20)$$

$$\lim_{k\to\infty} f(X^k) = -\infty < f(X^*) = \min_X f(X),$$

which contradicts (11).

Thus, for any $\varepsilon > 0$ there exists such a $K(\varepsilon)$ that for all $k \geqslant K(\varepsilon)$ the inequality $\|\nabla f(X^k)\| < \varepsilon$ will hold. Hence

$$\nabla f(X^k) = \nabla f(X^k) - \nabla f(X^*) = H(Z^k)(X^k - X^*);$$

$$Z^k = X^* + \theta^k(X^k - X^0) \in \mathfrak{M}_X, \quad 0 < \theta^k < 1; \quad (21)$$

$$X^k - X^* = (H(Z^k))^{-1}\nabla f(X^k); \qquad \|X^k - X^*\| < \varepsilon/m,$$

which proves the theorem.

Choose $l^k = (H(X^k))^{-1}\nabla f(X^k)$.

$$\frac{\|\nabla f(X^k)\|^2}{M} < (\nabla f(X^k), l^k) = (\nabla f(X^k), (H(X^k))^{-1}, \nabla f(X^k)) \leqslant \quad (22)$$

$$\leqslant \|(H(X^k))^{-1}\| \cdot \|\nabla f(X^k)\| < \frac{\|\nabla f(X^k)\|^2}{m};$$

$$\min\left(t, \frac{2m^2}{pM^2}\right) < t^k < \frac{2M^2}{m^2}; \qquad \frac{2m^2}{pM^2} < 2; \qquad \frac{2M^2}{m^2} > 2; \tag{23}$$

$$\min\left(t/p, \frac{2m^2}{p^2M^2}\right) < \tau^k < \frac{2M^2}{pm^2}.$$

Theorem 2. If $0 < t < 2$, then there exists a neighborhood

$$W_\delta(X^\circ) = \{X : \|X - X^\circ\| \leq \delta\} \subset \mathfrak{M}_X$$

such that for

$$X^k \in W_\delta(X^\circ); \quad X^k_{(t)} = X^k - t(H(X^k))^{-1}\nabla f(X^k) \tag{24}$$

the inequality $f(X^k_{(t)}) < f(X^k)$ holds.

Proof. From (24) we have

$$X^k_{(t)} - X^k = -t(H(X^k))^{-1}(\nabla f(X^k) - \nabla f(X^\circ)) =$$

$$= -t(H(X^k))^{-1}H(\bar{X}^k)(X^k - X^\circ);$$

$$\bar{X}^k = X^\circ + \bar{\theta}^k(X^k - X^\circ) \in W_\delta(X^\circ) \subset \mathfrak{M}_X; \tag{25}$$

$$0 < \bar{\theta}^k < 1; \quad \|X^k_{(t)} - X^k\| \leq \frac{2M}{m}\,\delta = \delta'.$$

Let us choose a number

$$0 < \varepsilon < \frac{(2-t)m^2}{Mt}. \tag{26}$$

In view of (1), for ε there exists such a $\delta' > 0$, and hence also δ, that whenever the inequality

$$\|X - X^k\| \leq \frac{2M}{m}\,\delta = \delta'$$

is satisfied, we have

$$\|H(X) - H(\bar{X}^k)\| < \varepsilon;$$

$$f(X^k_{(t)}) = f(X^k) - t(\nabla f(X^k), (H(X^k))^{-1}\nabla f(X^k)) + \frac{t^2}{2}(H(\bar{X}^k_{(t)})\times$$

$$\times(H(X^k))^{-1}\nabla f(X^k), (H(X^k))^{-1}\nabla f(X^k)); \quad \bar{X}^k_{(t)} = X^k + \bar{\theta}^k_{(t)}(X^k - X^k_{(t)});$$

$$0 < \bar{\theta}^k_{(t)} < 1; \quad \|\bar{X}^k_{(t)} - X^k\| \leq \frac{2M}{m}\,\delta = \delta';$$

$$f(X^k_{(t)}) = f(X^k) - \frac{t}{2}(\nabla f(X^k), (H(X^k))^{-1}\nabla f(X^k))(2-t)+$$
$$+\frac{t^2}{2}((H(\bar{X}^k_{(t)}) - H(X^k))(H(X^k))^{-1}\nabla f(X^k), (H(X^k))^{-1}\nabla f(X^k)) <$$
$$< f(X^k) - \frac{t}{2}\frac{\|\nabla f(X^k)\|^2}{M}(2-t) + \frac{t^2}{2}\varepsilon\frac{\|\nabla f(X^k)\|^2}{m^2} =$$
$$= f(X^k) - \frac{t^2\|\nabla f(X^k)\|^2}{2m^2}\left(\frac{(2-t)m^2}{Mt} - \varepsilon\right) < f(X^k),$$

which completes the proof.

In (3) choose $t = p,\ 1 < p < 2$. In this case the iterative process

$$X^{k+1} = X^k - \tau^k(H(X^k))^{-1}\nabla f(X^k) \tag{27}$$

converges on the basis of (22) and Theorem 1, and according to Theorem 2 there exists such a K that for all $k \geq K$ the equality

$$\tau^k = 1 \tag{28}$$

holds.

Theorem 3. The iterative process (27) converges with superlinear rate.

Proof. From (27), (28), for $k \geq K$ we have

$$X^{k+1} - X^\circ = X^k - X^\circ - (H(X^k))^{-1}(\nabla f(X^k) - \nabla f(X^\circ)) =$$
$$= X^k - X^\circ - (H(X^k))^{-1}H(\bar{X}^k)(X^k - X^\circ) = (H(X^k))^{-1}(H(X^k) - \tag{29}$$
$$-H(\bar{X}^k))(X^k - X^\circ); \quad \|X^{k+1} - X^\circ\| \leq \frac{\varepsilon}{m}\|X^k - X^\circ\|,$$

i.e. the assertion of the theorem.

If $f(X) \in C^3$, then the iterative process (27) with the choice of the parameter according to (3), (4) $(1 < t = p < 2)$ converges, as does Newton' s method for solving the nonlinear system $\nabla f(X) = 0$, with quadratic rate (see ([4])).

Institute of Cybernetics
Academy of Sciences of the Ukrainian SSR

Received
30 XII 1968

CITED LITERATURE

1. L. V. Kantorovich, DAN, **48**, No. 7, 483 (1945).

2. I. M. Glazman, DAN, **154**, No. 5, 1011 (1964).

3. A. A. Goldstein, SIAM, J. on Control, A, **3**, No. 1, 147 (1965).

4. L. V. Kantorovich, Vestn. Leningrad Univ., **2**, No. 7, 68 (1957).

Note: Figure translations are in progress. See original paper for figures.