MATHEMATICS

M. M. VAINBERG

1960

SovietRxiv

Abstract

Full Text

MATHEMATICS

M. M. VAINBERG

ON THE CONVERGENCE OF THE METHOD OF STEEPEST DESCENT FOR NONLINEAR EQUATIONS

(Presented by Academician S. L. Sobolev on 1 IX 1959)

1. The method of steepest descent has been applied in a number of works (see, for example, (1-4)) to the solution of equations in finite-dimensional spaces. For linear equations in Hilbert space this method was proposed and developed by L. V. Kantorovich (5-7). Later this method was applied to the solution of certain nonlinear equations in Hilbert space by Yu. Lumiste (10) and Guan Chzhao-chzhi (8).

In the present work we extend the method of steepest descent to more general cases: we show its convergence for nonlinear equations in Banach spaces and establish estimates for the error of the approximate solution.

2. Let $F(x)$ be a potential operator (9), defined in a Banach space E, and let $f(x)$ be its potential. Proceeding from the idea of the method of steepest descent and using certain considerations from (7), we arrive at the conclusion that a minimizing sequence $\{x_n\}$ for the functional $f(x)$ can be defined by the following iterative process:

$$x_{n+1} = x_n - \varepsilon_n AF(x_n), \tag{1}$$

where ε_n are certain positive numbers and A is a bounded linear operator from E^* into E (E^* is the space conjugate to E), normalized by the condition

$$(y, Ay) \geqslant \|y\|^2, \tag{2}$$

for every $y \in E^*$. Here and below (y, x) denotes the value of the linear functional $y \in E^*$ on the vector $x \in E$. If process (1) converges to some vector x_0, then at this vector x_0 the functional has a minimum, and therefore $\operatorname{grad} f(x) = F(x)$ vanishes, i.e. $F(x_0) = 0$. Convergence of process (1) to a solution of the equation

$$F(x) = 0 \tag{3}$$

is established by the following theorem.

Theorem 1. Let the Gateaux-differentiable potential operator $F(x)$, defined in E, satisfy the condition

$$\|h\|\gamma(\|h\|) \leqslant (F'(x)h, h) \leqslant M(\|x\|)\|h\|^2, \tag{4}$$

where $\gamma(t)$ is an increasing function such that $\gamma(0) = 0$, $\lim_{t\to+\infty} \gamma(t) = +\infty$; A is a bounded linear operator from E^* into E, satisfying condition (2).

Then equation (3) has a unique solution in E, to which the iterative process (1) converges for any initial approximation x_1, if

$$\frac{1}{2M_n\|A\|^2} \leqslant \varepsilon_n \leqslant \frac{1}{M_n\|A\|^2},$$

where $M_n = \max\{1, M(R_n)\}$, $R_n = \|x_n\| + \|A\|\|F(x_n)\| > 0$.

Let us outline the proof. The existence of a solution x_0 of equation (3) is ensured by Theorem 9.4 of [(9)], while in proving the uniqueness of the solution x_0 the following is used.

Lemma. *If the operator $\Phi(x)$ from E into the Banach space E_1 satisfies the condition $(\mathfrak{A}h, \Phi'(x)h) \geqslant \beta(\|x\|, \|h\|)$, where $\mathfrak{A}$ is some operator from E into E_1^* (linear or nonlinear) and $\beta(\tau, t)$ is a nonnegative function of nonnegative arguments, vanishing only when $t = 0$, then the equation $\Phi(x) = 0$ can have in E at most one solution.*

This lemma follows directly from the generalized Lagrange formula [(9)]. In proving the convergence to x_0 of the process (1), we first establish that $\{f(x_n)\}$ is a convergent sequence of decreasing numbers, whence, according to inequality (9.3) of [(9)], we find that $\{x_n\}$ is a bounded sequence. From the boundedness of $\{x_n\}$ follows the boundedness of $\{F(x_n)\}$, $\{R_n\}$, and hence also of $\{M_n\}$. Hence, and from the inequality $f(x_n) - f(x_{n+1}) \geqslant \frac{1}{2}\varepsilon_n\|F(x_n)\|^2$, which is derived from the conditions of the theorem, it follows that

$$\lim_{n\to\infty} F(x_n) = 0.$$

With the help of this equality the convergence of the sequence $\{x_n\}$ to the solution x_0 is proved, and the estimate $\gamma(\|x_n - x_0\|) \leqslant \|F(x_n)\|$ is established.

3. Theorem 1, like Theorem 9.4 of [(9)], admits other closely related formulations. In view of this we note that the proof of these theorems is preserved if the inequality $(F'(x)h, h) \geqslant \|h\|\gamma(\|h\|)$ is replaced by the inequality

$$(F'(x)h, h) \geqslant \|h\|\gamma(\|x\|, \|h\|) \geqslant 0,$$

where the function of nonnegative arguments $\gamma(\tau, t)$ satisfies the condition

$$\lim_{R\to+\infty} \int_0^1 \int_0^1 \gamma(\tau t R, tR)\, d\tau\, dt = +\infty.$$

In particular, Theorem 1 and Theorem 9.4 of [9] remain valid when the inequality $(F'(x)h, h) \geqslant \|h\|\gamma(\|h\|)$ is replaced by

$$(F'(x)h, h) \geqslant m(\|x\|)\|h\|^2,$$

where $m(t)$ is a positive nonincreasing function and

$$\int_0^{+\infty} m(t)\, dt = +\infty.$$

We further note that Theorem 1 is applicable to various examples from [9] (see Theorems 10.2, 10.4, 24.3, and 24.5 of [9]) and to the examples considered at our suggestion by R. I. Kachurovskii [11] and N. V. Kirpotina.

4. Under the conditions of Theorem 1, the iterative process (1) leads to a sequence $\{x_n\}$ that is minimizing not only for $f(x)$, but also for $\|F(x)\|$. This fact can be used for finding the solution of equation (3) by process (1) also when the operator $F(x)$ is not potential. Suppose, for example, that the operator A satisfies the following condition:

(χ) A is a linear bounded operator from E^* into E, satisfying inequality (2), and moreover $(A^{-1})^* A = I$, where I is the identity operator in E^*.

Then the following proposition holds:

Theorem 2. *Suppose the following conditions are fulfilled: the Gâteaux differentiable operator $F(x)$ from E into E^* satisfies, for all $x, h \in E$, the inequalities*

$$\|F'(x)\| \leqslant M, \qquad (F'(x)h, h) \geqslant m\|h\|^2, \qquad 0 < m = \text{const}, \quad M = \text{const};$$

the operator A satisfies condition (χ), and

$$\|A\|(1 - m^2/M^2) < 1.$$

Then equation (3) has in E a unique solution x_0, to which the iterative process (1) converges (for $\varepsilon_n = \varepsilon = \text{const}$) at the rate of a geometric progression, starting from any initial approximation x_1, if $\varepsilon_1 \leqslant \varepsilon \leqslant \varepsilon_2$, where $\varepsilon_1, \varepsilon_2$ are the roots of the equation

$$a^3 M^2 \varepsilon^2 - 2a^2 m\varepsilon + a - q^2 = 0, \qquad a = \|A\|; \qquad \sqrt{a(1 - m^2/M^2)} < q < 1.$$

For estimating the error, the formula holds

$$\|x_n - x_0\| \leqslant \alpha(1 - q)^{-1} q^{n-1},$$

where $\alpha = a\varepsilon\|F(x_1)\|$.

The proof uses the inequality

$$\|F(x_{n+1})\| \leqslant (a^3M^2\varepsilon^2 - 2a^2m\varepsilon + a)^{1/2}\|F(x_n)\|,$$

which is derived directly from the hypotheses of the theorem. We note that Theorem 2 is applicable to the equation $\Phi(x) = 0$, when Φ maps E into E_1, if the operator $F(x) = B\Phi(x)$ satisfies the hypotheses of Theorem 2, where B is a bounded linear operator from E_1 into E^* having a bounded inverse.

5. Here we shall assume that the norm in E is Fréchet differentiable, and we introduce

Definition 1. We shall say that the norm in E is **smooth** if the remainder of the Fréchet differential satisfies the inequality

$$|\omega(x,h)| \equiv \left|\|x+h\|^2 - \|x\|^2 - (\operatorname{grad}\|x\|^2, h)\right| \leqslant C(\|x\|)\|h\|^2, \quad x \neq 0.$$

Without loss of generality one may here assume that $C(\|x\|) \geqslant 1$.

Theorem 3. *Let the Gateaux-differentiable operator $F(x)$, acting from the space E with a smooth norm into E, satisfy the conditions*

$$\|F'(x)\| \leqslant M = \text{const},$$

$$(\operatorname{grad}\|h\|,\, F'(x)h) \geqslant m\|h\|, \qquad 0 < m = \text{const}$$

for all $x, h \in E$.

Then equation (3) has in E a unique solution x_0, to which the iterative process converges

$$x_{n+1} = x_n - \varepsilon F(x_n), \tag{5}$$

starting from any initial approximation x_1, if $\varepsilon_1 \leqslant \varepsilon \leqslant \varepsilon_2$, where $\varepsilon_1, \varepsilon_2$ are the (positive) roots of the equation

$$CM^2\varepsilon^2 - 2m\varepsilon + 1 - q^2 = 0; \qquad \sqrt{1 - \frac{m^2}{CM^2}} < q < 1, \qquad C(\|x\|) \equiv C = \text{const}.$$

For estimating the error, the formula holds

$$\|x_{n+1} - x_0\| \leqslant \frac{\varepsilon q^{n-1}}{1-q}\,\|F(x_1)\|.$$

The proof uses the inequality

$$\|F(x_{n+1})\| \leqslant (CM^2\varepsilon^2 - 2m\varepsilon + 1)^{1/2}\|F(x_n)\|,$$

which is derived directly from the hypotheses of the theorem.

In the case where the quantities C, M, and m depend on $\|x\|$, one may use the following proposition:

Theorem 4. *Let the following conditions be fulfilled: 1) equation (3), where $F(x)$ is an operator from the space E with a smooth norm into E, has a solution x_0, for which the estimate $\|x_0\| < r$ is known; 2) the Gateaux-differentiable operator $F(x)$ satisfies in the ball $D(\|x\| \leqslant R = 2r)$ the conditions: $\|F'(x)\| \leqslant M = M(R)$, $(\operatorname{grad}\|h\|, F'(x)h) \geqslant m\|h\|$, $0 < m = m(R)$.*

Then x_0 is the unique solution of equation (3) in D, to which the iterative process (5) converges, if ε is chosen in the same way as in Theorem 3, where $C = C(R)$.

For estimating the error, the formula holds:

$$\|x_n - x_0\| \leqslant q^{n-1}\|x_1 - x_0\|,$$

where the initial approximation x_1 is chosen so that $\|x_1\| \leqslant r$.

The proof uses the inequality $\|x_{n+1} - x_0\| \leq (CM^2\varepsilon^2 - 2m\varepsilon + 1)^{1/2}\|x_n - x_0\|$, which is obtained directly from the conditions of the theorem.

6. We make the following remarks.

1) If the operator $\Phi(x)$ maps E into E_1, then Theorems 3 and 4 can be applied to the operator $F(x) = B\Phi(x)$, where B is a linear bounded operator from E_1 into E having a bounded inverse, since in this case the equations $F(x) = 0$ and $\Phi(x) = 0$ are equivalent.

2) If, in addition, one requires the uniform continuity of the derivative $F'(x)$ on every bounded set, and also that the function $m(t)$ satisfy the conditions of item 3, then in the hypotheses of Theorem 4 one may dispense with the requirement that a solution of equation (3) exist, since for the existence of a solution of equation (3) it is sufficient that the following conditions hold: a) uniform continuity of $F'(x)$ on every bounded set; b) uniform continuity of $\operatorname{grad}\|x\|$ on every annulus $0 < r_1 \leq \|x\| \leq r_2$; c) $(\operatorname{grad}\|h\|, F'(x)h) \geq m(\|x\|)\|h\|$, where the function $m(t)$ satisfies the conditions of item 3.

Moscow Regional Pedagogical Institute
named after N. K. Krupskaya

Received
28 VIII 1959

CITED LITERATURE

1. H. B. Curry, Quart. Appl. Math., **2**, No. 3, 258 (1944).
2. A. D. Booth, Quart. J. Mech. and Appl. Math., **2**, 460 (1949).
3. T. C. Koopmans, H. Rubin, R. B. Leipnik, *Measuring the Equations Systems of Dinamic Economics*, N. Y., 1950.
4. J. B. Crockett, H. Chernoff, Pac. J. Math., **5**, 33 (1955).
5. L. V. Kantorovich, DAN, **48**, No. 7, 483 (1945).
6. L. V. Kantorovich, DAN, **56**, No. 3, 233 (1947).
7. L. V. Kantorovich, *Uspekhi Mat. Nauk*, **3**, issue 6 (28), 89 (1948).
8. Kwan Chao-chih, Acta Math. Sinica, **6**, No. 4, 638 (1956).
9. M. M. Vainberg, *Variational Methods for the Study of Nonlinear Operators*, 1956.
10. Yu. Lumiste, *Scientific Notes of Tartu University*, No. 37, 106 (1955).
11. R. I. Kachurovskii, *Scientific Notes of the Moscow Regional Pedagogical Institute named after N. K. Krupskaya*, **77**, 203 (1959).

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