GRADIENT METHODS FOR THE MINIMISATION OF FUNCTIONALS*

B.T. POLYAK

(Moscow)

(Received 2 July 1962)

Let $f(t)$ be a functional defined in the (real) Hilbert space H. The problem consists in finding its minimum value $f^* = \inf f(x)$ and some minimum point x^* (if such exists).

We shall assume everywhere below that the function is continuously differentiable, i.e. for any x, y

$$f(x+y) = f(x) + (h(x), y) + O(\| y \|), \tag{1}$$

where $h(x) \in H$ is the gradient of the functional $f(x)$ at the point x and is assumed to depend continuously on x.

Gradient minimisation methods consist in constructing the minimising sequence $x^0, x^1, \ldots, x^n, \ldots$ according to the formula

$$x^{n+1} = x^n - \alpha_n h(x^n). \tag{2}$$

The size of the step $\alpha_n \geqslant 0$ can be chosen in various ways. Thus, in the simple iteration method $\alpha_n = \text{const}$. In the method of fastest descent α_n is chosen to minimise the functional at each step, i.e. α_n gives the minimum $\psi(\alpha) = f(x^n - \alpha h(x^n))$.

When $f(x)$ is a quadratic functional, the minimisation problem is equivalent to the solution of the linear equation $h(x) = 0$. This case has received detailed study, and a survey of this work can be found in [1]. Although gradient methods are widely used in the practical solution

* *Zh. vych. mat.* 3, No. 4, 643-653, 1963.

of extremal problems for non-quadratic functionals also (mainly in finite-dimensional spaces, as in [2], for instance) little study has been made of the questions of convergence and rate of convergence in this case. There are several results for a finite-dimensional space in [3], [4], for a Hilbert space in [5] and for a Banach space in [6].

In most of the research (in [5], [6] and [7] in particular) the problem of the extremum is considered as an auxiliary problem associated with the solution of the equation $h(x) = 0$. The conditions for the convergence are therefore expressed in the form of restrictions on $h(x)$. We shall use another approach: we take the extremum problem to be the initial problem, the equation $h(x) = 0$ being simply a necessary condition for an extremum. For this reason we lay certain restrictions on the functional itself (as a rule these are simple restrictions, such as boundedness below) and not only on its gradient. Since the majority of applied problems are formulated with the use of variational principles, testing these conditions does not usually cause any difficulty.

Let us first give a theorem which is useful for the proof of the convergence of many methods of descent (not necessarily gradient methods).

Theorem 1. Let Z be a topological space, $f(z)$ a continuous functional on it, P an operator mapping Z into Z, where $f(Pz) \leqslant f(z)$ and $f(Pz)$ is a semi-continuous function of z from above. Then if the sequence z^0, $z^1 = Pz^0, \ldots, z^{n+1} = Pz^n, \ldots$ has a limit point z^*, $f(Pz^*) = f(z^*)$.

Proof. Obviously $f(z^*) \leqslant f(z^n)$, $\lim\limits_{n\to\infty} f(z^n) = f(z^*)$. Let $f(z^*) - f(Pz^*) = \varepsilon > 0$. It follows from the fact that $f(Pz)$ is semi-continuous above that there is a neighbourhood U of the point z^* such that $f(Pz) - f(Pz^*) < \varepsilon$ for any $z \in U$. Let us take $z^n \in U$. Then $f(Pz^n) - f(Pz^*) = f(z^{n+1}) - f(Pz^*) < \varepsilon$, i.e. $f(z^{n+1}) < f(z^*)$, which is impossible.

Let us give an example of the use of this general theorem.

Theorem 2. If the set $S = \{x : f(x) \leqslant c\}$ is compact for some c and $x^0 \in S$, then in the method of fastest descent $h(x^n) \underset{n\to\infty}{\to} 0$.

Proof. It is obvious that $f(x^{n+1}) < f(x^n)$, if $h(x^n) \neq 0$, and $f(x^{n+1}) = f(x^n)$, if $h(x^n) = 0$. Let us show further that $f(x^{n+1}) = f(Px^n)$ is a function of x^n, semi-continuous above.

Let Q be the sphere $\| x - x^{n+1} \| < \rho$ such that $|f(x) - f(x^{n+1})| < \varepsilon$ for all $x \in Q$. Due to the continuity of $h(x)$ we can find a sphere Q'

(namely $\| x - x^n \| < \rho'$) such that $\alpha_n \| h(x) - h(x^n) \| < \rho/2$. Here α_n is given by the equation $x^{n+1} = x^n - \alpha_n h(x^n)$ and we can take $\rho' \leqslant \rho/2$. Let us show that $\bar{x} = x - \alpha_n h(x) \in Q$ for $x \in Q'$. For $\| \bar{x} - x^{n+1} \| = \| x - \alpha_n h(x) - x^n + \alpha_n h(x^n) \| \leqslant \| x - x^n \| + \alpha_n \| h(x) - h(x^n) \| < \rho$. Thus $| f(\bar{x}) - f(x^{n+1}) | < \varepsilon$. But by the definition of the method of fastest descent $f(Px) \leqslant f(\bar{x})$, i.e. $f(Px) - f(x^{n+1}) < \varepsilon$ when $x \in Q'$, and this means semi-continuity above. Thus the conditions of Theorem 1 are satisfied. Since all the $x^n \in S$, we can find a limit point x^* and for this point $f(Px^*) = f(x^*)$, i.e. $h(x^*) = 0$. The theorem follows from this. For let us take for each limit point x^* and arbitrary $\varepsilon > 0$ the neighbourhood $U(x^*)$ such that $\| h(x) \| < \varepsilon$ for $x \in U(x^*)$. Let

$$U = \bigcup_{x^*} U(x^*).$$

Then only a finite number of points $x^n \overline{\in} U$. For if there were an infinite number of them, there would have to be a limit point $x^* \overline{\in} U$, for them, and this is impossible. Thus for any $\varepsilon > 0$ we can find N such that $x^n \in U$ for $n \geqslant N$, i.e. $\| h(x^n) \| < \varepsilon$.

Corollary. If, in addition to the conditions of Theorem 2, the functional $f(x)$ is convex (i.e. $\left(\frac{x+y}{2}\right) \leqslant \frac{1}{2} f(x) + \frac{1}{2} f(y) \Big)$, then $f(x^n) \to f^*$.

For in this case $h(x^*) = 0$ implies that $f(x^*) = f^*$.

Theorem 2 is proved for the finite-dimensional case in [4].

Theorem 3. If $f(x)$ satisfies the conditions

$$\| h(x + y) - h(x) \| \leqslant R \| y \|, \tag{A}$$

$$f^* > -\infty, \tag{B}$$

then for the sequence (2) $\lim_{n \to \infty} h(x^n) = 0$ for any x^0 if

$$\varepsilon_1 \leqslant \alpha_n \leqslant \frac{2}{R} - \varepsilon_2 \text{ for } \varepsilon_1, \varepsilon_2 > 0. \tag{3}$$

Proof. It follows from condition (1) (see [6]) that:

$$f(x + y) = f(x) + \int_0^1 (h(x + ty), y)\, dt,$$

$$f(x^{n+1}) - f(x^n) = -\alpha_n \int_0^1 (h(z), h(x^n))\,dt = -\alpha_n \|h(x^n)\|^2 + \\ + \alpha_n \int_0^1 (h(x^n) - h(z), h(x^n))\,dt \leqslant -\alpha_n \|h(x^n)\|^2 + \\ + \alpha_n^2 \|h(x^n)\|^2 R \int_0^1 t\,dt = \alpha_n \left(\frac{\alpha_n R}{2} - 1\right) \|h(x^n)\|^2 \leqslant 0,$$

$$z = x^n - t\alpha_n h(x^n).$$

Thus the sequence $f(x^n)$ is monotonically non-increasing and bounded below, i.e. has a limit. Therefore $\lim\limits_{n\to\infty} [f(x^n) - f(x^{n+1})] = 0$, but from the last inequality

$$\|h(x^n)\|^2 \leqslant \frac{f(x^n) - f(x^{n+1})}{\alpha_n\left(1 - \frac{R\alpha_n}{2}\right)} \leqslant \frac{f(x^n) - f(x^{n+1})}{\varepsilon_1 \frac{R\varepsilon_2}{2}}, \tag{4}$$

i.e. $\lim\limits_{n\to\infty} h(x^n) = 0$.

Of course, it does not follow from this theorem that $\lim\limits_{n\to\infty} f(x^n) = f^*$ (for example, when there is a local minimum) or that the sequence x^n converges (for example, the functional $e^{-(x,x)}$). For the convergence of the method we must lay additional restrictions on the functional.

Theorem 4. Let $f(x)$ satisfy the conditions (A), (B) and

$$\|h(x)\|^2 \geqslant 2r\,[f(x) - f^*], \qquad r > 0. \tag{C}$$

Then if α_n satisfies (3) for any x^0 we have the convergence $x^n \to x^*$, $f(x^n) \to f^*$, with

$$f(x^n) - f^* \leqslant q^n\,[f(x^0) - f^*], \qquad \|x^n - x^*\|^2 \leqslant c_0\,q^n, \qquad 0 \leqslant q < 1. \tag{5}$$

Note. The conditions of the theorem lay quite weak restrictions on the functional. Thus, the functional will not necessarily be convex or have a unique minimum point.

Proof. Put $\varphi(x) = f(x) - f^*$. It follows from (4) that

$$\varphi(x^n) - \varphi(x^{n+1}) = f(x^n) - f(x^{n+1}) \geqslant \alpha_n\left(1 - \frac{R\alpha_n}{2}\right)\|h(x^n)\|^2 \geqslant \\ \geqslant \alpha_n(2 - R\alpha_n)\,r\varphi(x^n),$$

$$\varphi(x^{n+1}) \leqslant [1 - \alpha_n(2 - R\alpha_n) r] \varphi(x^n) = p_n \varphi(x^n).$$

Further, from (4)

$$\|h(x^n)\|^2 \leqslant \frac{\varphi(x^n)}{\alpha_n(1 - R\alpha_n/2)},$$

and it follows, after a comparison with (C), that

$$r \leqslant \frac{1}{\alpha_n(2 - R\alpha_n)},$$

i.e. $p_n \geqslant 0$. On the other hand

$$p_n \leqslant 1 - Rr\varepsilon_1\varepsilon_2 = q < 1.$$

Thus, $\varphi(x^{n+1}) \leqslant q\varphi(x^n), \varphi(x^n) \leqslant q^n \varphi(x^0)$, and the second part of the theorem is proved.

Further

$$\|x^{n+1} - x^n\|^2 = \alpha_n^2 \|h(x^n)\|^2 \leqslant \frac{\alpha_n \varphi(x^n)}{(1 - R\alpha_n/2)} \leqslant \frac{(2/R - \varepsilon_2)\varphi(x^0)}{R\varepsilon_2/2} q^n = c_1 q^n.$$

Hence

$$\|x^n - x^m\| \leqslant \sum_{k=m}^{n-1} \|x^{k+1} - x^k\| \leqslant \sqrt{c_1} \sum_{k=m}^{n-1} q^{k/2} \leqslant \frac{\sqrt{c_1} q^{m/2}}{1 - \sqrt{q}}$$

for all $n > m$, i.e. the sequence x^n converges in itself and thus converges to some x^* where

$$\|x^* - x^n\|^2 \leqslant \frac{c_1 q^n}{(1 - \sqrt{q})^2} = c_0 q^n.$$

In Theorem 4 we required that the conditions be satisfied in the whole space H. In fact, it is sufficient that they should be satisfied in some neighbourhood of x^0. Let us modify Theorem 4 in this way for the method of simple iteration.

Theorem 5. Suppose that in the region $S = \{x: \|x - x^0\| \leqslant \rho\}$ conditions (A), (B), (C) are satisfied, where $\gamma = \sqrt{2R\varphi(x^0)}/r\rho < 1$. Then in S there exists the minimum point x^* and $\|x^0 - x^*\| \leqslant \gamma\rho$. If α

satisfies the condition

$$0 < \alpha < \bar{\alpha}, \qquad \bar{\alpha} = 2\frac{1-\gamma}{R - r\gamma^2},$$

for the method of simple iteration we have the convergence

$$\varphi(x^n) \leqslant \varphi(x^0)\, q^n, \qquad \| x^n - x^* \| \leqslant C(\alpha)\, q^{n/2},$$

$$q = 1 - \alpha(2 - R\alpha)\, r, \qquad 0 \leqslant q < 1, \qquad C(\alpha) = \frac{\alpha \sqrt{2R\varphi(x^0)}}{1 - \sqrt{q}}.$$

The magnitude of q is minimal and equal to $1 - r/R$ for $\alpha = 1/R$.

Proof. Since $\bar{\alpha} \leqslant 2/R$, as above we find that $\varphi(x^n) \leqslant \varphi(x^0)\, q^n$ $q = 1 - \alpha(2 - R\alpha)\, r$, $0 \leqslant q < 1$. Putting $\alpha_n = 1/R$ in (4) we obtain $\| h(x)\|^2 \leqslant 2R\varphi(x)$, and so $\| x^{n+1} - x^n\| = \alpha\| h(x^n)\| \leqslant \alpha \sqrt{2R\varphi(x^0)}\, q^{n/2}$. Hence

$$\| x^n - x^m \| \leqslant \sum_{k=m}^{n-1} \| x^{k+1} - x^k \| \leqslant C(\alpha)\, q^{m/2}, \qquad n > m.$$

It remains to show that all $x^n \in S$. The use of conditions (A), (B), (C) in the proof is then justified. For $\| x^0 - x^n \| \leqslant C(\alpha)$. Let us show that $C(\alpha) \leqslant \rho$. It is not difficult to verify that on $[0, 2/R]$ $C(\alpha)$ is an increasing function of α. But $C(\bar{\alpha}) = \rho$, and so $C(\alpha) \leqslant \rho$ for $0 < \alpha < \bar{\alpha}$, i.e. $x^n \in S$. Finally, for any $0 < \alpha < \alpha$ there exists x^* such that $\| x^0 - x^* \| \leqslant C(\alpha)$, and since $\{x^*\}$ is closed, there exists x^* such that $\| x^0 - x^* \| \leqslant C(+0) = \gamma\rho$.

Let us now give the conditions guaranteeing the uniqueness of the minimum:

$$(h(x + y) - h(x),\, y) \geqslant m \| y \|^2, \qquad m > 0. \tag{D}$$

The functionals satisfying conditions (A) and (D) (and also weaker conditions of the same kind) are considered in detail in [6]. Their most important feature is that when these conditions are satisfied the functional $f(x)$ has a unique minimum x^*. The proof is based on the use of the principle of contracted mappings for the equation $x = x - (1/R)h(x)$, equivalent to the equation $h(x) = 0$. In the case when the existence of a point x^* for which $h(x^*) = 0$ is known, condition (D) can be replaced by the weaker condition

$$(h(x^* + y),\, y) \geqslant m \| y \|^2, \qquad m > 0. \tag{D'}$$

Lemma. Conditions (B) and (C) follow from conditions (A) and (D) (or (D′)).

Proof. The lemma was proved for (B) above. From (D) or (D′) we have

$$\| h(x) \| \| x - x^* \| \geqslant (h(x), x - x^*) \geqslant m \| x - x^* \|^2, \quad \| h(x) \| \geqslant m \| x - x^* \|.$$

On the other hand

$$f(x) - f^* = \int_0^1 (h(y), x - x^*)\, dt \leqslant \| x - x^* \| \int_0^1 \| h(y) \|\, dt \leqslant \frac{R}{2} \| x - x^* \|^2,$$
$$y = x^* + t(x - x^*),$$

i.e.

$$\| h(x) \|^2 \geqslant m^2 \| x - x^* \|^2 \geqslant \frac{2m^2}{R} [f(x) - f^*],$$

i.e. (C) is satisfied and we may take m^2/R for r.

Corollary. Theorems 4 and 5 are valid for functionals satisfying conditions (A) and (D) (or (D′)).

Theorems 4 and 5 for this case follow from the more general theorems in [6].

By strengthening conditions (A) and (D) to some extent, we can simplify the testing of them. Let the functional $f(x)$ be twice differentiable, i.e. suppose that for any x, y

$$f(x + y) = f(x) + (h(x), y) + \tfrac{1}{2}(A(x) y, y) + o(\| y \|^2), \tag{6}$$

where $A(x)$ is a bounded self-conjugate linear operator. Suppose, further, for all x

$$M(x) \leqslant M, \qquad M(x) = \sup_{\| y \| = 1} (A(x) y, y); \tag{A′}$$
$$m(x) \geqslant m > 0, \qquad m(x) = \inf_{\| y \| = 1} (A(x) y, y). \tag{D″}$$

Then, obviously, (A) and (D) are satisfied, and we can take $R = M$ and m the same as in (D″).

Corollary. Theorems 4 and 5 are valid for functionals which satisfy conditions (A′) and (D″).

The statement analogous to Theorem 4 for this case is proved in [5].

We note that on passing to conditions (A) and (D) or (A$'$) and (D$'$) we have not attempted to obtain the most accurate estimates for the rate of convergence (the quantity q). Let us now make them more accurate for the most interesting case, the simple iteration method.

Theorem 6. Suppose that the functional satisfies conditions (A$'$) and (D$'$). Then in the simple iteration method, for $0 < \alpha < 2/M$

$$\| x^n - x^* \| \leqslant q^n \| x^0 - x^* \|, \qquad q = \max\{|1 - \alpha m|, |1 - \alpha M|\}, \tag{7}$$

the minimum value of q occurs when $\alpha = 2/(M + m)$ and is equal to $(M - m)/(M + m)$.

Proof.

$$\| x^{n+1} - x^* \| = \| x^n - x^* - \alpha h(x^n)\| = \left\| x^n - x^* - \alpha \int_0^1 A(y)(x^n - x^*)\, dt \right\| =$$

$$= \int_0^1 (E - \alpha A(y))(x^n - x^*)\, dt \,\| \leqslant \| x^n - x^* \| \int_0^1 \| E - \alpha A(y)\| \, dt \leqslant q \| x^n - x^* \|,$$

$$E - \alpha A(y)\| = \sup_{\|x\| = 1} |((E - \alpha A(y))x, x)| \leqslant \sup_{m \leqslant \mu \leqslant M} |1 - \alpha\mu| =$$

$$= \max\{|1 - \alpha m|, \ |1 - \alpha M|\} = q.$$

$$y = x^* + t(x^n - x^*).$$

It is obvious that q attains a minimum when $1 - \alpha m = -(1 - \alpha M)$, i.e. when $\alpha = 2/(M + m)$.

Now let us discuss the method of fastest descent in detail. In this case Theorem 4 is inapplicable, since the condition $\alpha_n < 2/R$ cannot be satisfied.

Theorem 7. Let conditions (A), (B), (C) be satisfied in the region $S = \{x : f(x) \leqslant c\}$. Then for any $x^0 \in S$ in the method of fastest descent $f(x^n) \to f^*$ and

$$\varphi(x^n) \leqslant \varphi(x^0)(1 - r/R)^n.$$

Further, let conditions (A$'$) and (D$'$) be satisfied in

$$T = \left\{ x : \| x - x^0 \| \leqslant \sqrt{\frac{8\varphi(x^0)}{m}} \right\}.$$

Then $x^n \to x^*$ at the rate of a geometric progression:

$$\|x^n - x^*\| \leqslant \sqrt{\frac{2\varphi(x^0)}{m}}\left(\frac{M-m}{M+m}\right)^n.$$

Proof. Put $\bar{x}^{n+1} = x^n - (1/R)\,h\,(x^n)$. Then, as in the proof of Theorem 4, $\varphi\,(\bar{x}^{n+1}) \leqslant (1 - r/R)\,\varphi\,(x^n)$. But, by the definition of the method of fastest descent, $\varphi\,(x^{n+1}) \leqslant \varphi\,(\bar{x}^{n+1})$, i.e. $\varphi\,(x^{n+1}) \leqslant (1 - r/R)\ \varphi\ (x^n)$, which proves the first part of the theorem.

In exactly the same way as in the case when conditions (A′) and (D″) are satisfied we obtain

$$\varphi\,(x^n) \leqslant \left(\frac{M-m}{M+m}\right)^{2n}\varphi\,(x^0).$$

Further, it follows from (D″) that

$$\varphi\,(x) = \tfrac{1}{2}(A\,(\xi)\,(x - x^*),\,(x - x^*)) \geqslant \tfrac{1}{2}\,m\,\|x - x^*\|^2,$$
$$\xi = x + \theta\,(x^* - x), \qquad 0 \leqslant \theta \leqslant 1.$$

Therefore

$$\|x^n - x^*\|^2 \leqslant \frac{2}{m}\,\varphi\,(x^n) \leqslant \frac{2\varphi\,(x^0)}{m}\left(\frac{M-m}{M+m}\right)^{2m}.$$

Finally

$$\|x^0 - x^n\| \leqslant \|x^0 - x^*\| + \|x^n - x^*\| \leqslant$$
$$\leqslant \sqrt{\frac{2\varphi\,(x^0)}{m}}\left(1 + \left(\frac{M-m}{M+m}\right)^n\right) \leqslant 2\sqrt{\frac{2\varphi\,(x^0)}{m}},$$

and so $x^n \in T$ and conditions (A′) and (D″) are applicable.

We could examine the problem of the minimisation of a functional on the linear manifold $G = F + z$, where F is a sub-space of H, $z \in H$, instead of on the whole space. All the theorems we have proved would remain valid, with obvious changes in the conditions. Thus, in (1) we would have to take $x \in G$, $y \in F$ and instead of $h(x)$, $\bar{h}(x)$, the projection of the gradient on F, and so on.

Now let us consider some examples.

1. The quadratic functional

$$f\,(x) = \tfrac{1}{2}(Ax,\,x) - (b,\,x), \tag{8}$$

where A is a non-negative definite operator, and b is orthogonal to the sub-space E_1 corresponding to a zero eigenvalue. We note that, in particular, the problem of minimising $\|Bx - d\|^2$, where B is some bounded linear operator from one Hilbert space into another reduces to such a functional. Let us assume in addition that if zero belongs to the spectrum of A it is an isolated point of the spectrum.

We shall show that in this case conditions (A), (B) and (C) are satisfied. Let $H = E_1 \oplus E_2$. Then the operator A^{-1} is defined in E_2 and is positive definite there. Since $g = Ax - b \in E_2$, we have $f(x) = \frac{1}{2}(A^{-1}g, g) - \frac{1}{2}(A^{-1}b, b) \geqslant -\frac{1}{2}(A^{-1}b, b)$. Obviously, $f^* = -\frac{1}{2}(A^{-1}b, b)$, and $x^* = A^{-1}b + x_1$; $x_1 \in E_1$ is arbitrary. Further $h(x) = Ax - b$, $\|h(x + y) - h(x)\| = \|Ay\| \leqslant \|A\| \, \|y\|$, i.e. in conditions (A) we can take $R = \|A\|$. Finally

$$f - f^* = \frac{1}{2}(A^{-1}h(x), h(x)) \leqslant \frac{1}{2\lambda_{\min}} \|h(x)\|^2,$$

i.e. in conditions (C) we can take $r = \lambda_{\min}$ where $\lambda_{\min}$ is the smallest positive eigenvalue of A.

Thus Theorems 4 and 5 are valid for the functional (8). The point x^* to which the sequence x^n converges is obviously $A^{-1}b + x_1^0$, where $x^0 = x_1^0 + x_2^0$, $x_1^0 \in E_1$, $x_2^0 \in E_2$.

2. Let us consider the simplest variational problem, the minimisation of

$$f(x) = \int_a^b F(t, x(t), x'(t))\, dt, \qquad x(a) = x_1, \qquad x(b) = x_2. \tag{9}$$

We shall solve the problem in the Hilbert space $W_2^{(1)}$ (see [8]). Its elements are absolutely continuous functions with a square-integrable derivative, and the scalar product is given by:

$$(x, y) = x(a)\, y(a) + x(b)\, y(b) + \int_a^b x'(t)\, y'(t)\, dt.$$

We shall assume below that the function $F(t, x, x')$ possesses all the necessary derivatives. Let us find the gradient of the functional $f(x)$, assuming in (1) that $x(t)$ has fixed values at the end-points, and $y(t)$ has zero values there. From the usual formula for the first variation

$$(h(x), y) = \int_a^b (F_x y + F_{x'} y')\, dt = \int_a^b \Big[F_{x'}(t, x, x') - \int_a^t F_x(\tau, x, x')\, d\tau\Big]\, y'(t)\, dt,$$

i.e.

$$\frac{d}{dt} h(x) = F_{x'}(t, x, x') - \int_a^t F_x(\tau, x, x')\, d\tau;$$

$$h(x) = \int_c^t \Big[F_{x'}(s, x, x') - \int_a^s F_x(\tau, x, x')\, d\tau\Big]\, ds.$$

Since we are looking for the extremum only on the linear manifold corresponding to the boundary conditions, and not in the whole space, in accordance with the remark we made above we must take the projection of the gradient on the sub-space of functions with zero boundary conditions. To do this we note that any linear function $kt + p$ is orthogonal to this sub-space:

$$(kt + p, y) = 0 + \int_a^b k y'\, dt = 0.$$

We therefore take $\overline{h}(x) = h(x) + kt + p$ and choose k and p so that $\overline{h}(x)\,|_{t=a} = \overline{h}(x)\,|_{t=b} = 0$. We obtain

$$\overline{h}(x) = \int_a^t \Big[F_{x'}(s, x, x') - \int_a^s F_x(\tau, x, x')\, d\tau\Big]\, ds -$$

$$- \frac{t-a}{b-a} \int_a^b \Big[F_{x'}(s, x, x') - \int_a^s F_x(\tau, x, x')\, d\tau\Big]\, ds.$$

We note that if we differentiate the equation $\overline{h}(x) = 0$ twice with respect to t we obtain the ordinary Euler equation

$$F_x - \frac{d}{dt} F_{x'} = 0.$$

Thus, the gradient method for solving a variational problem consists in choosing the initial approximation $x^0(t)$, $x^0(a) = x_1$, $x^0(b) = x_2$ and constructing subsequent approximations from the formula $x^{n+1}(t) = x^n(t) - \alpha_n \overline{h}(x^n)$.

The theorems proved above enable us to evaluate the convergence of these methods in certain cases. In addition, we obtain new variants of Tonelli's theorem about the absolute extremum [9].

Let us just give the simplest conditions on F for which the theorems can be applied. Let the second variation of the functional exist; then

$$(A\ (x)\ y,\ y) = \int_a^b \left[\left(F_{xx} - \frac{d}{dt} F_{xx'}\right) y^2 + F_{x'x'} y'^2\right] dt.$$

Let $F \geqslant c_1$, $\left| F_{xx} - \frac{d}{dt} F_{xx'} \right| \leqslant c_2$, $| F_{x'x'} | \leqslant c_3$ for any $x \in G$ and all t. Then conditions (A) and (B) are satisfied. For it is obvious that (B) is satisfied and

$$(A\ (x)\ y,\ y) \leqslant c_2 \int_a^b y^2\, dt + c_3 \int_a^b y'^2\, dt \leqslant [c_2 (b-a) + c_3] \int_a^b y'^2\, dt = \\ = [c_2 (b-a) + c_3] \| y \|^2,$$

i.e. (A′) and thus (A) is satisfied. We can therefore use Theorem 3 in this case.

Now let

$$0 \leqslant F_{xx} - \frac{d}{dt} F_{xx'} \leqslant c_4, \qquad 0 < c_5 \leqslant F_{x'x'} \leqslant c_6.$$

In this case conditions (A) and (D) are satisfied. (A) is tested in the same way, and from

$$(A\ (x)\ y,\ y) \geqslant \int_a^b F_{x'x'} y'^2\, dt \geqslant c_5 \| y \|^2$$

it follows that (D′) is satisfied. Thus Theorems 4 and 5 can be used here.

We note further that if $h(x) \to 0$ in $W_2^{(1)}$ then

$$F_{x'} - \int_a^t F_x(\tau,\ x,\ x')\, d\tau \to 0$$

in L^2 and if $x \to x^*$ in $W_2^{(1)}$ then $x \to x^*$ uniformly on $[a,\ b]$.

A similar method is examined in [10] from a different point of view. This method for solving variational problems can be extended to more complicated cases (multidimensional problems, for instance). Some qualitative considerations are given in [11].

To conclude, let us briefly consider the continuous analogue of the gradient method. Here, instead of constructing the sequence (2) we look for $x(t)$, the solution of the equation

$$\frac{dx}{dt} = -h(x). \tag{10}$$

We can solve extremal problems by this method on continuous action computers. Let us just consider a few analogues of the preceding theorems.

We shall put $f(t) = f(x(t))$, $h(t) = h(x(t))$; $\varphi(t) = f(t) - f^*$.

Theorem 8. If $f(x)$ satisfies conditions (A), (B), then for any $x(0)$ $\lim\limits_{t\to\infty}$, $h(t) = 0$.

Proof.

$$\frac{df}{dt} = \left(h(t), \frac{dx}{dt}\right) = -\|h(t)\|^2 \leqslant 0.$$

Therefore the function $f(t)$ is monotonic non-increasing, and so, because of (8), tends to a limit $\bar{f}$. But

$$\frac{d\|x(t) - x(0)\|}{dt} \leqslant \left\|\frac{dx}{dt}\right\| =$$

$$= \|h(t)\| \leqslant \|h(t) - h(0)\| + \|h(0)\| \leqslant R\|x(t) - x(0)\| + \|h(0)\|,$$

i.e.

$$\|x(t) - x(0)\| \leqslant \|h(0)\|(e^{Rt} - 1).$$

Let $T = (1/R)\ \ln 2$. Then

$$f(0) - f(T) = \int_0^T \|h(t)\|^2\, dt \geqslant \int_0^T [\|h(0)\| - R\|x(t) - x(0)\|]^2\, dt \geqslant$$

$$\geqslant \|h(0)\|^2 \int_0^T (2 - e^{Rt})^2\, dt = \frac{\|h(0)\|^2}{R}\left(4\ln 2 - \frac{5}{2}\right) \geqslant 0.27\,\frac{\|h(0)\|^2}{R}.$$

Thus $\|h(t)\|^2 \leqslant \frac{R}{0.27}(f(t) - \bar{f})$, and so $h(t) \to 0$ as $t \to \infty$.

Theorem 9. Let conditions (A), (B), (C), with $\gamma = \frac{\sqrt{2R\varphi(0)}}{r\rho} \leqslant 1$. be satisfied in the region $S = \{x : \|x - x(0)\| \leqslant \rho\}$ Then there exists the minimum point x^* in S and $\|x(0) - x^*\| \leqslant \gamma\rho$. In addition

$$\varphi(t) \leqslant \varphi(0)\, e^{-2rt}, \qquad \| x(t) - x^* \| \leqslant \gamma \rho e^{-rt}.$$

Proof.

$$\frac{d\varphi}{dt} = \frac{df}{dt} = -\| h(t) \|^2 \leqslant -2r\varphi(t), \qquad \text{i.e. } \varphi(t) \leqslant \varphi(0)\, e^{-2rt}.$$

Further

$$\| x(t_2) - x(t_1) \| = \left\| \int_{t_1}^{t_2} h(t)\, dt \right\| \leqslant \int_{t_1}^{t_2} \| h(t) \|\, dt \leqslant \sqrt{2R} \int_{t_1}^{t_2} \sqrt{\varphi(t)}\, dt \leqslant$$
$$\leqslant \frac{\sqrt{2R\varphi(0)}}{r} (e^{-rt_1} - e^{-rt_2}), \qquad t_1 < t_2.$$

It follows that $\lim\limits_{t\to\infty} x(t) = x^*$, and $\| x(t) - x^* \| \leqslant \dfrac{\sqrt{2R\varphi(0)}}{r} e^{-rt}$. We can use conditions (A), (B) and (C) since

$$\| x(0) - x(t) \| \leqslant \frac{\sqrt{2R\varphi(0)}}{r} (1 - e^{-rt}) \leqslant \gamma\rho \leqslant \rho,$$

i.e. $x(t) \in S$ for all t.

Translated by R. Feinstein

REFERENCES

1. Kantorovich, L.V., *Uspekhi matem. nauk*, 11, No. 6, 99-116, 1956.
2. Fel'dbaum, A.A., *Vychislitel'nye ustroistva v avtomaticheskikh sistemakh (Computing Assemblies in Automatic Systems)*. Fizmatgiz, Moscow, 1959.
3. Crocket, J.B. and Chernoff, H., *Pacif. J. Math.*, 5, No. 1, 33-50, 1955.
4. Curry, H.B., *Quart. Appl. Math.*, 2, No. 3, 268-271, 1944.
5. Guan' Chzhao-Chzhi, *Acta math. sinica*, 6, No. 4, 638-650, 1956.
6. Vainberg, M.M., *Sibirskii matem. zh.*, 2, No. 2, 201-220, 1961.
7. Vainberg, M.M., *Variatsionnye metody issledovaniya nelineinykh*

operatorov (Variational Methods for the Study of Non-Linear Operators). Gostekhizdat, Moscow, 1956.

8. Sobolev, S.L., *Nekotorye primeneniya funkteional'nogo analiza v matematicheskoi fizike (Some Applications of Functional Analysis in Mathematical Physics)*. Izd-vo Leningrad State University, 1950.

9. Akhiezer, N.I., *Lektsii po variatsionnomu ischisleniyu (Lectures on Variational Analysis)*. Gostekhizdat, Moscow, 1955.

10. Stein, M.L., *J. Res. Nat. Bur. Standards*, 50, No. 3, 277-297, 1953.

11. Tompkins, C.B., In the book *Sovremennaya matem. dlya inzh (Contemporary Mathematics for Engineers)*. Izd-vo in. lit., Moscow, 1958, 441-472.