ON A GENERALIZATION OF PROJECTION METHODS

MATHEMATICS

1966

SovietRxiv

Abstract

Full Text

UDC 518:519.34:517.948

MATHEMATICS

E. I. POLSKII

ON A GENERALIZATION OF PROJECTION METHODS

(Presented by Academician G. I. Petrov on 19 VI 1965)

1°. Ordinary projection methods consist in the following ([1]). Consider the equation $Lx = f$. The operator L acts from the space H_1 into the space H_2. An approximate solution

$$x_n = \sum_{i=1}^{n} c_i \varphi_i$$

is sought in the subspace $R_n \subset H_1$ from the condition that $P_n L x_n = P_n f$, where P_n is the projection operator onto the subspace $M_n \subset H_2$. As was shown in ([1]), for convergence of Lx_n to Lx for every f it is necessary and sufficient that $\tau = \lim_{n\to\infty} \tau_n > 0$; here $\tau_n = \min |P_n y_n|$, where $|y_n| = 1;\ y_n \in L_n = LR_n$.

We shall call the quantity τ_n the P_n**-projection deviation** of M_n from L_n*. The scheme presented assumes that the subspaces R_n are contained in the domain of definition of the operator L. In particular, if L is a differential operator, then the coordinate functions $\{\varphi_i\}$ must satisfy the boundary conditions of the problem. It is clear that, for a complicated boundary or complicated boundary conditions, the selection of the system $\{\varphi_i\}$ is very difficult. Such is the situation in all projection methods: Ritz, Galerkin, moments, least squares, etc.

The situation is different in Trefftz' s method ([3,4]). There the coordinate functions $\{\varphi_i\}$ are chosen to satisfy the differential expression itself. The method consists in selecting such combinations

$$x_n = \sum_{i=1}^{n} c_i \varphi_i$$

as to satisfy the boundary conditions approximately. The difficulties in applying Trefftz's method are usually caused by the complexity of the principal differential expression.

Below we consider a generalization of the projection method which does not require the coordinate functions to satisfy either the differential expression or the boundary conditions. This generalization contains, in particular, the projection methods and Trefftz' s method. Apparently, the possibility of generalizing the Ritz-Galerkin method, providing the indicated freedom in the choice of coordinate functions, was first pointed out in ([5]). A justification of such a generalization for second-order elliptic differential equations was given in ([6]). In the present paper a general scheme for the application of such methods is considered.

2°. Let D be a linear set dense in the Hilbert space H_1. In D there are given, generally speaking, unbounded operators $\mathcal{L}$ and Λ with values in the Hilbert spaces H and G, respectively. We denote the orthogonal sum of H and G by $H_2 = H \oplus G$. In this case, if $f_i \in H$, $\varphi_i \in G$, then $F_i = \{f_i, \varphi_i\} \in H_2$ $(i = 1, 2)$ and $[F_1, F_2] = (f_1, f_2) + \langle\varphi_1, \varphi_2\rangle$. We consider the problem of finding an element $x \in D \subset H_1$ which is carried by the operators $\mathcal{L}$ and Λ into the prescribed elements $f \in H$, $\varphi \in G$, respectively. Denoting the operator $L = \{\mathcal{L}, \Lambda\}$, we obtain the equation $Lx = F$, where $x \in D \subset H_1$, $F \in H_2$. Assuming that this equation is uniquely solvable–

---

* If P_n is the operator of orthogonal projection, then the quantity τ_n is connected with the gap θ_n between the subspaces M_n and L_n ([2]) by the relation $\tau_n^2 + \theta_n^2 = 1$.

in this way, for any F, we apply the projection method ([1]) to its solution.

Recall that a sequence of subspaces M_n of a Hilbert space H, together with the sequence of operators P_n projecting onto M_n, is called **projection-complete** ([1]) if $|f - P_n f| \to 0$ for every $f \in H$ as $n \to \infty$.

The approximate solution x_n is sought, as in the case of ordinary projection methods, in a certain finite-dimensional subspace $R_n \subset D \subset H_1$. The operator L maps R_n into the n-dimensional space $L_n \subset H_2 = H \oplus G$. Next, in H_2 one chooses a subspace M_n and an operator P_n projecting onto M_n; moreover, let Π_n be the projector in H_2 onto L_n. Then x_n is sought from the condition $P_n L x_n = P_n F$. As in (1):

The necessary and sufficient condition for unique solvability of the approximate equations for any F is the condition $\tau_n > 0$. In this case, in the norm of H_2 the estimate is valid:

$$|Lx - Lx_n| \le (1 + \|P_n\|/\tau_n)|F - \Pi_n F|. \tag{1}$$

If the sequence $\{L_n, \Pi_n\}$ is projection-complete in H_2, then convergence of Lx_n to Lx for every F takes place if and only if $\tau = \lim_{n\to\infty} \tau_n > 0$.

Remark. If all F are considered from some subspace, then the necessary and sufficient condition for convergence is the projection completeness of the system $\{L_n, \Pi_n\}$ in this subspace and such a choice of the system $\{M_n, P_n\}$ that $\tau > 0$.

What has been said makes it possible to transfer to the method under consideration all results obtained in various works for projection methods, for example

([1],[7],[8]).

3°. Let us note that the subspaces $U_s = \mathcal{L}R_n \subset H$ and $V_t = \Lambda R_n \subset G$ may, generally speaking, have dimensions s and t, different from n. However, always

$$\max(s, t) \leq n \leq s + t.$$

Let $\varphi_1, \dots, \varphi_n$ be a complete system of independent elements in R_n, and let $\Psi_1, \dots, \Psi_n$ be such a system in M_n, where $\Psi_i = \{g_i, \psi_i\}$; $g_i \in H$; $\psi_i \in G$; $i = 1, \dots, n$. If, in addition, P_n is the operator of orthogonal projection in H_2, then the approximate system of equations takes the form

$$[Lx_n, \Psi_i] = (\mathcal{L}x_n - f, g_i) + \langle \Lambda x_n - \varphi, \psi_i \rangle = 0, \quad i = 1, \dots, n. \tag{2}$$

We single out in particular the case when $n = s + t$. If by $D \subset H_1$ one understands the set of functions sufficiently smooth in some domain Ω, including the boundary of the domain σ, $\mathcal{L}$ is a differential operator in the domain Ω, and Λ is a collection of boundary differential expressions, then $n = s + t$ means that s functions φ_i satisfy the conditions $\Lambda\varphi_i = 0$, while t functions satisfy the conditions $\mathcal{L}\varphi_i = 0$.

In this case the subspace $L_n \subset H_2$ may be regarded as the orthogonal sum of the subspaces $U_s \subset H$ and $V_t \subset G$. Let also $M_n = U_s' \oplus V_t'$, where $U_s' \subset H$; $V_t' \subset G$. Then, if $F = \{f, \varphi\} \in H_2$, the operator Π_n, projecting H_2 onto L_n, is naturally represented in the form

$$\Pi_n F = \{S_s f, T_t \varphi\},$$

where S_s projects H onto U_s, and T_t projects G onto V_t. It is easy to see that projection completeness of the sequence $\{L_n, \Pi_n\}$ is equivalent to the simultaneous projection completeness of $\{U_s, S_s\}$, $\{V_t, T_t\}$, and

$$|F - \Pi_n F|_{H_2} = |f - S_s f|_H + |\varphi - T_t \varphi|_G.$$

Analogously, the operator P_n is represented in the form

$$P_n F = \{S_s' f, T_t' \varphi\}.$$

Theorem. Let τ_n be the projection deviation of M_n from $\dot{L}_n$, let τ' be that of U_s' from U_s, and let τ'' be that of V_t' from V_t; then

$$\tau_n = \min(\tau', \tau'').$$

Proof. By definition,

$$\begin{aligned}
\tau_n &= \min\left(|S_s' g| + |T_t' \psi|\right), \quad && \text{where } g \in U_s,\ \psi \in V_t,\ |g| + |\psi| = 1; \\
\tau' &= \min |S_s' g|, && \text{where } g \in U_s,\ |g| = 1; \\
\tau'' &= \min |T_t' \psi|, && \text{where } \psi \in V_t,\ |\psi| = 1.
\end{aligned}$$

Let, for definiteness, $\tau' \le \tau''$. Since $|S_s' g| + |T_t' \psi|$ for $\psi = 0$ assumes the same values as $|S_s' g|$, it is evident that $\tau_n \le \tau'$.

On the other hand, let τ_n be attained on the elements $\bar{g}$ and $\bar{\psi}$, with

$$|\bar{g}| = a, \qquad |\bar{\psi}| = 1 - a.$$

Then

$$\begin{aligned}
\tau_n = |S_s' \bar{g}| + |T_t' \bar{\psi}| = a \left| S_s' \left(\frac{1}{a} \bar{g} \right) \right| + (1 - a) \left| T_t' \left(\frac{1}{1 - a} \bar{\psi} \right) \right| \ge \\
\ge a\tau' + (1 - a)\tau'' \ge \tau'.
\end{aligned}$$

By virtue of the theorem just proved, estimate (1) takes the form

$$|Lx - Lx_n| \le \left(1 + \frac{\|S_s'\| + \|T_t'\|}{\min(\tau', \tau'')} \right) \left(|f - S_{sf}| + |\varphi - T_t \varphi| \right).$$

If, in addition to the condition $n = s + t$ adopted above, for example $t = 0$, i.e. $n = s$, then necessarily $\varphi = 0$ as well, since otherwise the vector $F = \{f, \varphi\}$ will not be approximated by the spaces M_n. If, however, $s = 0$, then $f = 0$ and $n = t$.

The case considered makes it possible to reduce the question of convergence of the residual in H_2 to checking the convergence conditions in H and G separately.

4°. Let us note also the case when convergence is always guaranteed by just the projection completeness of the sequence $\{L\varphi_n\}$ in H_2. Namely, let

$$\Psi_i = \{\mathcal{L}\varphi_i, \Lambda\varphi_i\},$$

and let the operator P_n project orthogonally. Then system (2) takes the form

$$[Lx_n, L\varphi_i] = (\mathcal{L}x_n - f, \mathcal{L}\varphi_i) + \langle \Lambda x_n - \varphi, \Lambda\varphi_i \rangle = 0, \qquad i = 1, \dots, n. \quad (3)$$

Here the subspaces L_n and M_n coincide, i.e. $\tau_n \equiv 1$, and the process always converges, which gives a direct generalization of the least-squares method, whose convergence was first substantiated in ([9]).

Remark. As in the general variant of system (2), so also in the case of the least-squares method (3), it is necessary to keep in mind that the process may cease to converge if one of the equations of the system $\mathcal{L}x = f$ or $\Lambda x = \varphi$ is multiplied by a constant α. In this case the condition $\tau_n > 0$ may be violated.

To preserve the convergence of the least-squares method, system (3) should be given the form

$$(\alpha\mathcal{L}x_n - \alpha f, \alpha\mathcal{L}\varphi_i) + \langle \Lambda x_n - \varphi, \Lambda\varphi_i \rangle = 0, \qquad i = 1, \dots, n.$$

Moreover, multiplication of one of the equations by α will not affect convergence and the estimate if $n = s + t$. In all cases, multiplication by α of one of the equations, or even of the functions φ_i, may affect the conditioning or stability of the approximate system.

5°. Examples of application of the considerations set forth may be differential equations, where $\mathcal{L}$ is a differential operator in a certain domain Ω, and Λ is a differential operator on the boundary σ of the domain Ω. It was noted above that convergence always takes place, for example, in the least-squares method, i.e. when

$$\Psi_i = L\varphi_i = \{\mathcal{L}\varphi_i, \Lambda\varphi_i\}.$$

In the case of methods of Galerkin type,

$$\Psi_i = V\varphi_i = \{V\varphi_i, V\varphi_i\},$$

where V is the embedding operator from H_1 into H_2. In particular, in ([6])

$$H_1 = W_2^1$$

with scalar product

$$(u, v) = \int \left[a_{ik} \frac{\partial u}{\partial x_k} \frac{\partial v}{\partial x_i} + uv \right] d\Omega,$$

and

$$H_2 = L_2(\Omega) \oplus L_2(\sigma).$$

The approximate equations take the form

$$[Lx_n - F, V\varphi_i] = 0.$$

The second-order elliptic differential operator

$$L = \{\mathcal{L}, \Lambda\}$$

in ([6]) is chosen so that the operator

$$V^* L = I + A,$$

where I is the identity, and A is a completely continuous operator in H_1.

Other examples of applications of the theorems proved may be eigenvalue problems, when the parameter enters not only into the operator $\mathcal{L}$, but also into the boundary operator Λ. Important examples of such problems are contained in [10].

The results obtained also make it possible to prove the convergence of various methods of the Trefftz type considered in [4].

Let us also note the case when $\mathcal{L}$ is an ordinary differential operator of order n. Then the space H is Hilbert, while G and the operator Λ are finite-dimensional. In this case the Trefftz method degenerates into a trivial problem of linear algebra.

The author expresses sincere gratitude to A. D. Myshkis for a useful discussion of certain questions connected with this work.

Institute of Technical Thermophysics
Academy of Sciences of the Ukrainian SSR

Received
16 VI 1965

CITED LITERATURE

[1] N. I. Pol' skii, DAN, **143**, No. 4 (1962).
[2] M. G. Krein, M. A. Krasnosel' skii, UMN, **2**, issue 3 (19) (1947).
[3] S. G. Mikhlin, *Variational Methods in Mathematical Physics*, 1957.
[4] M. Sh. Birman, Vestn. Leningrad Univ., No. 13, ser. math., mech., astr., issue 3 (1956).
[5] Ya. I. Perel' man, Prikl. matem. i mekh., **5**, issue 2 (1941).
[6] Yu. I. Trostenberg, DAN, **85**, No. 3 (1952).
[7] V. A. Medvedev, DAN, **156**, No. 2 (1964).
[8] G. M. Vainikko, Uch. zap. Tartu Univ., issue 150 (IV) (1964).
[9] S. G. Mikhlin, Uch. zap. Leningrad Univ., No. 111 (1949).
[10] S. G. Krein, DAN, **159**, No. 2 (1964).

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