UNIVERSITATIS IAGELLONICAE ACTA MATHEMATICA, FASCICULUS XXVI

1987

On a Reasonable Choice of the Coordinate Functions in the Faedo-Galerkin Method

by Jan BOCHENEK

Introduction. In Mikhlin's monograph [7] conditions of a reasonable choice of coordinate functions for an approximate solution of the operator equation

$$Au = f, \tag{1}$$

were given with the use of the Ritz method. It is shown in [7] that on the choice of these coordinate functions, among other, the following properties depend:

1° convergence of the approximate solution and, possibly, convergence of the "residuum $Au_n - f$" to zero,

2° rate of convergence of the approximating sequence.

In the papers [1], [2], [3] this result of Mikhlin is generalized and transfered to the Bubnov-Galerkin method for the equation of the type (1), when the operator A is not self-adjoint.

The purpose of this paper is to give a certain fashion of a reasonable choice of the coordinate functions in the method of Faedo-Galerkin for the approximate solution of the equation of the parabolic type (cf. [4]).

Introductory concepts. Let X be a real separable Hilbert space with the norm $|\cdot|$ and the inner product $(\cdot, \cdot)$ and let A be a linear operator with domain $D(A) \subset X$. We assume that A is a self-adjoint positive definite operator i.e., the operator A satisfies the following condition

$$(Au, u) \geqslant p|u|^2, \quad u \in D(A), \ p > 0. \tag{2}$$

Let us denote by $L^2(a, b; X)$ the linear space of the functions $f\colon [a, b] \to X$, such that

$$\int_a^b |f(t)|^2 dt < \infty,$$

where $|f(t)|$ denotes the norm of the element $f(t) \in X$, and $-\infty \leqslant a < b \leqslant +\infty$. The space $L^2(a, b; X)$ is a Hilbert space with the norm

$$\|f\| := \{\int_a^b |f(t)|^2 dt\}^{1/2} \tag{3}$$

and with the inner product

$$\langle f, g\rangle := \int_a^b (f(t), g(t))dt, \tag{4}$$

in particular,

$$L^2(X) := L^2(-\infty, +\infty; X).$$

Let us denote by $D^k(a, b; X)$, for arbitrary integer $k \geqslant 0$, the set of functions f: $[a, b] \to X$ which are of the class C^k in (a, b) such that $f, f', \ldots, f^{(k)} \in L^2(a, b, X)$. In the set $D^k(a, b; X)$ we introduce the structure of a linear space and the norm by formula

$$\|f\|_{D^k(a,b;X)} := \{\sum_{p=0}^{k} \int_a^b |f^{(p)}(t)|^2 dt\}^{1/2}. \tag{5}$$

The closure of the space $D^k(a, b; X)$ in the norm defined by (5) is denoted by $H^k(a, b; X)$.

Let $0 < T \leqslant +\infty$ be fixed. We shall denote by $H_0^k(0, T; X)$ the closure of the subspace $D_0^k(-\infty, T, X) := \{f \in D^k(-\infty, T; X): f(t) = f'(t) = \ldots = f^{(k)}(t) = 0 \text{ for } t < 0\}$ in the norm defined by (5).

In this paper the spaces $H^0(0, T; X) = L^2(0, T; X)$, $H^1(0, T; X)$ and $H_0^1(0, T; X)$ play a substantial part. Since all these spaces are subspaces of the space $L^2(0, T; X)$, we shall assume that the norm and the inner product of these spaces are defined by the formulas

$$\|f\| := \{\int_0^T |f(t)|^2 dt\}^{1/2} \tag{6}$$

$$\langle f, g\rangle := \int_0^T (f(t), g(t))dt. \tag{7}$$

Let P be a linear operator such that

$$P: H_0^1(0, T; X) \to H^0(0, T; X), \tag{8}$$

defined by the formula

$$Pu := \frac{du}{dt} + \mathscr{A}u \quad \text{for } u \in H_0^1(0, T; X), \tag{9}$$

where $\mathscr{A}$ is a linear operator such that

$$\mathscr{A} := L^2(0, T; X) \to L^2(0, T; X)$$

defined by the formula

$$(\mathscr{A}u)(t) := Au(t) \text{ for } t \in [0, T] \text{ and } u \in L^2(0, T; X), \tag{10}$$

where A is the operator defined above. We shall prove the following

LEMMA 1. *The operator P defined by the formula* (9) *has the inverse bounded operator P^{-1} in the space $L^2(0, T; X)$.*

Proof. Let u be an element of $H_0^1(0, T; X)$. If $T < +\infty$, then u: $[0, T] \to X$ is a continuous function, therefore $u(0)$ and $u(T)$ have sense and $u(0) = 0$ (cf. [6], Ch. 1). By (2) and (4) we have

$$\langle Pu, u\rangle = \int_0^T (u'(t), u(t))dt + \int_0^T (Au(t), u(t))dt \geqslant \frac{1}{2}\int_0^T \frac{d}{dt}|u(t)|^2 dt$$

$$+p\int_0^T |u(t)|^2 dt = \frac{1}{2}|u(T)|^2 + p||u||^2 \geqslant p||u||^2.$$

If, however, $T = +\infty$, then for each $T_1 < +\infty$ we have

$$\int_0^{T_1} (u'(t) + Au(t), u(t))dt \geqslant p\int_0^{T_1} |u(t)|^2 dt.$$

From this, letting $T_1 \to +\infty$, we get

$$\langle Pu, u\rangle \geqslant p||u||^2.$$

Therefore, for each $T \leqslant +\infty$, we get

$$\langle Pu, u\rangle \geqslant p||u||^2. \tag{11}$$

Putting $v = Pu$ in (11) we have

$$||P^{-1}v|| \leqslant \frac{1}{p}||v||. \tag{12}$$

The proof is completed.

Apart from the operator A we now consider a linear self-adjoint, positive definite operator B such that $D(B) = D(A)$. We assume that the operator B posesses a discrete spectrum. We denote by $\{\lambda_n\}$ the denumerable sequence of eigenvalues of the operator B, and by $\{\varphi_n\}$ the corresponding orthonormal sequence of eigenvectors which form a complete system in the space X.

From the assumptions on the operators A and B it follows that AB^{-1}, BA^{-1}, $A^{-1}B$ and $B^{-1}A$ are bounded operators in X (cf. [7], Ch. I).

Let us denote by $\mathscr{B}$ the linear operator such that

$$\mathscr{B} := L^2(0, T; X) \to L^2(0, T; X) \tag{13}$$

defined by the formula

$$(\mathscr{B}u)(t) := Bu(t) \text{ for } t \in [0, T] \text{ and } u \in L^2(0, T; X), \tag{14}$$

where B is the operator defined above.

It is easy to prove that $\mathscr{A}\mathscr{B}^{-1}$, $\mathscr{B}\mathscr{A}^{-1}$, $\mathscr{A}^{-1}\mathscr{B}$ and $\mathscr{B}^{-1}\mathscr{A}$ are bounded operators in the space $L^2(0, T; X)$.

LEMMA 2. *The operator $P^{-1}\mathscr{B}$ is a bounded operator in the space $L^2(0, T; X)$.*

Proof. First let us observe that for the operator P there exists the adjoint operator P^*, defined by the formula

$$P^*v = -\frac{dv}{dt} + \mathscr{A}v \quad \text{for } v \in D(P^*), \tag{15}$$

where

$$D(P^*) = \{v \in H^1(0, T; X): v(T) = 0\} \quad \text{when } T < +\infty$$

and

$$D(P^*) = H^1(0, +\infty; X) \quad \text{when } T = +\infty.$$

Hence $D(P^*)$ is a dense subset of the space $H^1(0, T; X)$ and, for any $v \in D(P^*)$, we have

$$\langle P^*v, v\rangle = -\int_0^T (v'(t), v(t))dt + \int_0^T (Av(t), v(t))dt \geqslant -\tfrac{1}{2}|v(t)|^2|_0^T$$
$$+p\int_0^T |v(t)|^2 dt = \tfrac{1}{2}|v(0)|^2 + p\|v\|^2 \geqslant p\|v\|^2.$$

Therefore the operator P^* has the inverse bounded operator $(P^*)^{-1}$ and

$$\|(P^*)^{-1}\| \leqslant \frac{1}{p}. \tag{16}$$

We shall prove that $R(P^*) = D((P^*)^{-1})$ is the whole space $L^2(0, T; X)$. To this purpose it is sufficient to prove that $R(P^*)$ is dense in $D(P^*)$. Let $z \in D(P^*)$ and $y \in R(P^*)$. We shall prove that if $\langle z, y\rangle = 0$ for any $y \in R(P^*)$, then $z = 0$. Indeed, if $y \in R(P^*)$, then $y = P^*x$, where $x \in D(P^*)$, whence $0 = \langle y, z\rangle = \langle Px, z\rangle = \langle x, 0\rangle = \langle x, Pz\rangle$, i.e., $Pz = 0$. On the other hand, by (11) we have $p\|z\|^2 \leqslant \langle Pz, z\rangle = 0$ whence $z = 0$. It means that $R(P^*)$ is dense in $D(P^*)$. Since $D(P^*)$ is dense in $H^1(0, T; X)$ and $H^1(0, T; X)$ is dense in $L^2(0, T; X)$, then $R(P^*)$ is dense in $L^2(0, T; X)$. Since P^* is a closed operator, whence $R(P^*)$ is a closed and dense domain in $L^2(0, T; X)$. Therefore $R(P^*) = D((P^*)^{-1}) = L^2(0, T; X)$.

From the definition of the operator $\mathscr{B}$ it follows that $D(P^*) \subset D(\mathscr{B})$. Therefore we can define $\mathscr{B}(P^*)^{-1}$. Since the operator $\mathscr{B}(P^*)^{-1}$ is closed and defined in the whole space $L^2(0, T; X)$, it is bounded (cf. [8], p. 557). It follows that the operator $P^{-1}\mathscr{B}$ is bounded as the adjoint operator to $\mathscr{B}(P^*)^{-1}$.

The Faedo-Galerkin method. The Faedo-Galerkin method is an approximate method which is used, among other things, to solve equations of the type

$$Pu = f, \tag{17}$$

where f is an element of the space $L^2(0, T; X)$, and P is the operator defined by the formula (9).

Let $\{x_n\}$ be an orthonormal and complete system of vectors of the space X. Let us denote

$$u_n(t) := \sum_{k=1}^{n} c_k(t) x_k \quad \text{for } t \in [0, T], \tag{18}$$

where c_k $(k = 1, \ldots, n)$ are real-valued functions defined on the interval $[0, T]$, such that for every $t \in (0, T)$ the following equations hold

$$(Pu_n(t), x_k) = (f(t), x_k), \quad k = 1, \ldots, n, \tag{19}$$

where $Pu(t) := u'(t) + Au(t)$ for every $u \in H_0^1(0, T; X)$ and $t \in (0, T)$.

The system (19), as follows from the definition of the operator P, is a linear system of n ordinary differential equations of the first order with the unknown functions $c_1, \ldots, c_n$. The system (19) is called the Faedo-Galerkin system for the equation (17). The sequence $\{u_n\}$ defined by (18), where the coefficients c_k $(k = 1, \ldots, n)$ satisfy (19), is called the approximate sequence for the equation (17) in the sense of Faedo-Galerkin. It is proved that for any orthonormal and complete system $\{x_n\}$ in X, the sequences $\{u_n(t)\}$ and $\{u_n'(t)\}$ are convergent to $u(t)$ and $u'(t)$, respectively, for every $t \in (0, T)$, where u is the unique solution of the equation (17) (cf. [9]). We shall prove the following

THEOREM 1. *If in the Faedo-Galerkin method for the equation* (17) *we take the sequence* $\{\varphi_n\}$ *of the eigenvectors of the operator* B, *defined above, then for every* $t \in (0, T)$

$$|Pu_n(t) - f(t)| \to 0 \quad \text{if } n \to +\infty. \tag{20}$$

Proof. Let us denote by M_n and L_n the closed linear subspaces spaned by $\{\varphi_1, \ldots, \varphi_n\}$ and $\{\psi_1, \ldots, \psi_n\}$, respectively, and by P_n and π_n the projectors of X onto M_n and L_n, respectively, where $\psi_k = AB^{-1}\varphi_k$, $k = 1, \ldots, n$. By the definition of the operator P and by (18) the Faedo-Galerkin system (19) takes the form

$$c_j'(t) + \sum_{k=1}^{n} c_k(t)(A\varphi_k, \varphi_j) = (f(t), \varphi_j), \quad j = 1, \ldots, n. \tag{21}$$

Let us denote

$$z_n(t) := Au_n(t) \tag{22}$$

and let us observe that

$$z_n(t) = Au_n(t) = \sum_{k=1}^{n} c_k(t) A\varphi_k = \sum_{k=1}^{n} c_k(t) \lambda_k AB^{-1}\varphi_k = \sum_{k=1}^{n} a_k(t) \psi_k,$$

where $a_k(t) = \lambda_k c_k(t)$, $\psi_k = AB^{-1}\varphi_k$, $k = 1, \ldots, n$.
Now the system (21) may be written in the form

$$c_j'(t) + \sum_{k=1}^{n} a_k(t)(\psi_k, \varphi_j) = (f, (t), \varphi_j), \quad j = 1, \ldots, n, \tag{23}$$

or in another form

$$P_n u_n'(t) + P_n z_n(t) = P_n f(t), \quad t \in (0, T). \tag{24}$$

In the paper [2] it was proved that for every $v_n \in L_n$, we have the inequality

$$|v_n| \leqslant C\lambda_n^{\frac{1}{2}}|B^{-\frac{1}{2}}P_n v_n|, \quad C>0. \tag{25}$$

Since $z_n(t)-\pi_n[f(t)-u_n'(t)] \in L_n$, by (25) we get

$$|z_n(t)-\pi_n[f(t)-u_n'(t)]| \leqslant C\lambda_n^{\frac{1}{2}}|B^{-\frac{1}{2}}P_n\{z_n(t)-\pi_n[f(t)-u_n'(t)]\}|.$$

From this, by (24), we have

$$|z_n(t)-\pi_n[f(t)-u_n'(t)]| \leqslant C\lambda_n^{\frac{1}{2}}|B^{-\frac{1}{2}}P_n\{[f(t)-u_n'(t)]-\pi_n[f(t)-u_n'(t)]\}|. \tag{26}$$

Owing to the paper [2], the operators $B^{-\alpha}$ and P_n are commutative for every $\alpha>0$ and $n \in N$. From this and from inequality $|P_n| \leqslant 1$, by (26), we have

$$|z_n(t)-\pi_n[f(t)-u_n'(t)]| \leqslant C\lambda_n^{\frac{1}{2}}|B^{-\frac{1}{2}}\{[f(t)-\pi_n f(t)]-[u_n'(t)-\pi_n u_n'(t)]\}|. \tag{27}$$

Making the most of the boundedness of the operators AB^{-1} and $B^{-1}A$, the completeness of the sequence $\{\varphi_n\}$ and $\{\psi_n\}$ in the space X, the convergence of the sequence $\{u_n'(t)\}$ in the norm of the space X and reasoning analogously as in the proof of the Theorem 1 in the paper [2], we can get

$$\lim_{n\to\infty} \lambda_n^{\frac{1}{2}}|B^{-\frac{1}{2}}\{[f(t)-\pi_n f(t)]-[u_n'(t)-\pi_n u_n'(t)]\}| = 0 \tag{28}$$

From the inequality (27) by (28) follows that

$$\lim_{n\to\infty} |z_n(t)-\pi_n[f(t)-u_n'(t)]| = 0. \tag{29}$$

Let us observe that

$$\begin{aligned} |u_n'(t)+z_n(t)-f(t)| &= |f(t)-u_n'(t)-z_n(t)| \\ &= |\{[f(t)-u_n'(t)]-\pi_n[f(t)-u_n'(t)]\}-\{z_n(t)-\pi_n[f(t)-u_n'(t)]\}| \\ &\leqslant |[f(t)-u_n'(t)]-\pi_n[f(t)-u_n'(t)]|+|z_n(t)-\pi_n[f(t)-u_n'(t)]|, \end{aligned} \tag{30}$$

and

$$\begin{aligned} |[f(t)-u_n'(t)]-\pi_n[f(t)-u_n'(t)]| &\leqslant |f(t)-\pi_n f(t)|+|u_n'(t)-u'(t)| \\ &+|u'(t)-\pi_n u'(t)|+|\pi_n[u_n'(t)-u'(t)]| \leqslant |f(t)-\pi_n f(t)| \\ &+2|u_n'(t)-u'(t)|+|u'(t)-\pi_n u'(t)|, \end{aligned} \tag{31}$$

where $u'(t)$ is the limit of the sequence $\{u_n'(t)\}$ for every $t \in (0, T)$, in the norm of the space X.

From the inequalities (30) and (31) by (29) and the completeness of the sequence $\{\psi_n\}$, we get the equality

$$\lim_{n\to\infty} |u_n'(t)+z_n(t)-f(t)| = 0 \quad \text{for } t \in (0, T),$$

which is equivalent to (20) i.e., to the thesis of Theorem 1.

The rate of convergence of approximation in the sense of Faedo-Galerkin. Analogously as in the case of the equation (1), we shall prove the following

THEOREM 2. *If the operators A, B, $\mathscr{A}$, $\mathscr{B}$ and P are the operators defined in section* 1, *then*

$$\|u_n - u\| = o(\lambda_n^{-1}), \tag{32}$$

where $\{u_n\}$ is defined by (18), *u is the unique solution of the equation* (17), *but $\{\lambda_n\}$ is the sequence of eigenvalues of the operator B.*

Proof. Let us denote

$$\delta_n = Pu_n - f \tag{33}$$

and let us observe that for every $t \in (0, T)$, we have

$$\delta_n(t) = \sum_{k=1}^{\infty} (\delta_n(t), \varphi_k)\varphi_k .$$

By orthonormalization of $\{\varphi_n\}$ in the space X, we get

$$|\delta_n(t)|^2 = \sum_{k=1}^{\infty} |(\delta_n(t), \varphi_k)|^2.$$

On the other hand, by (19), for $x_k = \varphi_k$, $k = 1, \ldots, n$, we get

$$(\delta_n(t), \varphi_k) = (Pu_n(t) - f(t), \varphi_k) = 0 \quad \text{for } k = 1, \ldots, n .$$

It means that

$$|\delta_n(t)|^2 = \sum_{k=n+1}^{\infty} |(\delta_n(t), \varphi_k)|^2. \tag{34}$$

Now we shall estimate $|B^{-1}\delta_n(t)|$. We have

$$|B^{-1}\delta_n(t)|^2 = \sum_{k=1}^{\infty} |(B^{-1}\delta_n(t), \varphi_k)|^2 = \sum_{k=1}^{\infty} |(\delta_n(t), B^{-1}\varphi_k)|^2$$

$$= \sum_{k=1}^{\infty} \frac{|(\delta_n(t), \varphi_k)|^2}{\lambda_k^2} = \sum_{k=n+1}^{\infty} \frac{|(\delta_n(t), \varphi_k)|^2}{\lambda_k^2} \leqslant \frac{|\delta_n(t)|^2}{\lambda_{n+1}^2}.$$

Therefore

$$|B^{-1}\delta_n(t)| \leqslant \frac{|\delta_n(t)|}{\lambda_{n+1}}. \tag{35}$$

From the estimation (35), by the definition of the operator $\mathscr{B}$ (cf. (13) and (14)), and by (6) we have

$$\|\mathscr{B}^{-1}\delta_n\|^2 = \int_0^T |(\mathscr{B}^{-1}\delta_n)(t)|^2 dt = \int_0^T |B^{-1}\delta_n(t)|^2 dt \leqslant \frac{1}{\lambda_{n+1}^2} \int_0^T |\delta_n(t)|^2 dt = \frac{\|\delta_n\|^2}{\lambda_{n+1}^2},$$

that is

$$\|\mathscr{B}^{-1}\delta_n\| \leqslant \frac{\|\delta_n\|}{\lambda_{n+1}}.$$

On the other hand, we have

$$(36)\quad \|u_n-u\| = \|P^{-1}P(u_n-u)\| = \|P^{-1}(Pu_n-f)\| = \|P^{-1}\delta_n\|$$
$$= \|P^{-1}\mathscr{B}\mathscr{B}^{-1}\delta_n\| \leqslant \|P^{-1}\mathscr{B}\|\,\|\mathscr{B}^{-1}\delta_n\| \leqslant \frac{\|P^{-1}\mathscr{B}\|\,\|\delta_n\|}{\lambda_{n+1}}.$$

In virtue of Lebesgue's theorem, we have

$$\lim_{n\to\infty}\|\delta_n\|^2 = \lim_{n\to\infty}\int_0^T |\delta_n(t)|^2 dt = \int_0^T \lim_{n\to\infty} |\delta_n(t)|^2)dt = 0.$$

Since by Lemma 2 $\|P^{-1}\mathscr{B}\| \leqslant C$, so (36) implies (32), that is the thesis of Theorem 2.

Theorem 2 gives an estimation of the rate of convergence of the sequence $\{u_n\}$ in the norm of the space $L^2(0, T; X)$. As it follows from the definition of the norm in the space $L^2(0, T; X)$, the convergences in the spaces $L^2(0, T; X)$ and X are not equivalent.

To prove the next theorem, we shall need the following

LEMMA Fatou (see [5]). *If $\{f_n\}$ is a sequence of nonnegative an integrable functions on the set $G \subset R^n$ which converges to a function f for almost every $x \in G$, then*

$$(37)\quad \int_G f(x)\,dx \leqslant \varliminf_{n\to\infty} \int_G f_n(x)\,dx.$$

THEOREM 3. *Under the assumptions of the Theorem 2, the following estimation is true*

$$(38)\quad |u_n(t)-u(t)| = o(\lambda_n^{-1}),$$

for almost every $t \in (0, T)$.

Proof. Let us denote

$$(39)\quad f_n(t) := \lambda_n|u_n(t)-u(t)|,\ t \in [0, T],\ n \in N,$$

where λ_n, u_n and u have the same meaning as in Theorem 2. From the previous considerations it follows that for every $n \in N$

$$f_n\colon [0, T] \to R$$

is the continuous and nonnegative function in the interval $[0, T]$. From this, by (32), we have

$$(40)\quad \lim_{n\to\infty}\int_0^T f_n(t)\,dt = 0.$$

On the other hand because $u_n(t) \to u(t)$ for every $t \in [0, T]$, when $n \to \infty$, we have

$$\lim_{n\to\infty} f_n(t) = f(t) \quad \text{for every } t \in [0, T],$$

where f is any function defined in $[0, T]$. By (40) owing to Theorem 6 of [5], we get that the function f is integrable in $(0, T)$. From this, by Lemma Fatou, we have

$$(41)\quad \int_0^T f(t)\,dt \leqslant \varliminf_{n\to\infty}\int_0^T f_n(t)\,dt = 0.$$

By (41), we get

$$\lim_{n\to\infty} f_n(t) = 0 \text{ for almost every } t \in (0, T). \tag{42}$$

The equality (42), by (39), is equivalent to the estimation (38), that is the thesis of Theorem 3.

References

[1] J. Bochenek, *On the reasonable choice of the coordinate functions in the Bubnov-Galerkin method*, Comm. Math. 17 (1973), 9—16.

[2] —, *Some remarks on the convergence of Ritz and Galerkin methods*, Comm. Math. 17 (1973), 17—27.

[3] А. В. Джишкариани, *О методе Бубнова-Галеркина*, Жур. выч. Мат.-Физ. 7 (1967), Но. 2, 1398—1402.

[4] S. Faedo, *Un nuovo metodo per l'analisi esistenziale e quantitativa del problemi di propagazione*, Ann. Sc. Norm. Sup. Pisa (1949), 1—40.

[5] S. Hartman J. Mikusiński, *Teoria miary i całki Lebesgue'a*, PWN Warszawa 1957.

[6] J. L. Lions, E. Magenes, *Problemes aux limites non homogènes et application*, vol. I., Dunod, Paris 1968.

[7] С. Г. Михлин, *Численная реализация вариационных методов*, Москва 1966.

[8] В. И. Смирнов, *Курс высшей математики*, т. 5, Москва 1959.

[9] М. И. Вишик, *Задача Коши для уравнений с операторными коэффициентами, смешанная краевая задача для систем дифференциальных уравнений и приближенный метод их решения*, Матем. сб. 39 (81) вып. 1 (1956), 51—148.

Received November 2, 1981.