PACIFIC JOURNAL OF MATHEMATICS
Vol. 65, No. 1, 1976

SOME CONVERGENCE PROPERTIES OF THE BUBNOV-GALERKIN METHOD

S. R. SINGH

We generalize the Bubnov-Galerkin method to approximate the resolvent of the m-sectorial operator associated with a densely defined, closed, sectorial form in a Hilbert space. Some special cases of interest are also discussed.

1. Introduction. The Bubnov-Galerkin method [3] was originally devised to approximate the solutions of the equations of the form

$$(1) \qquad (z - A)f = g$$

where A is an operator in a Hilbert space, $\mathscr{H}$, g is a vector in $\mathscr{H}$ and z is a complex number. The method proceeds with solving the following set of equations:

$$(2) \qquad \sum_{j=1}^{n} \alpha_j(\phi_i \,|\, (z - A)\phi_j) = (\phi_i \,|\, g) \qquad i = 1, \cdots, n\,;$$

where $(.\,|\,.)$ denotes the scalar product in $\mathscr{H}$ and $\{\phi_i\} \subset \mathscr{D}(A)$ is some linearly independent (l.i.) set in $\mathscr{H}$. $\mathscr{D}(\cdot)$ denotes the domain. The questions of interest are the existence and the convergence of the solutions of equation (2). Until recently, the only cases that received a detailed treatment have been when A is compact, bounded or essentially self-adjoint [3, 6]. However, recently the following result was proven by Masson and Thewarapperuma [2]:

R.1. Let A be symmetric, bounded below by b, z be at a nonzero distance from $[b, \infty)$ and $\{\phi_i\}$ be the orthonormal set formed from $\{A^i h\}$ where h is in $\mathscr{D}(A^i)$ for each i. Then $\lim_{n\to\infty} \|\sum_{j=1}^{n} \alpha_j \phi_j - (z - A_F)^{-1} g\| = 0$, where $\|.\|$ denotes the norm in $\mathscr{H}$ and A_F is the Friedrichs extension of A.

Consider the following set of equations:

$$(3) \qquad \sum_{j=1}^{n} \alpha_j[z(\phi_i \,|\, \phi_j) - t(\phi_i, \phi_j)] = (\phi_i \,|\, g) \qquad i = 1, \cdots, n\,;$$

where t is a densely defined, closable, sectorial, sesquilinear form in $\mathscr{H}$. The sector of t will be denoted by S and since it causes no loss of generality, the vertex will be taken to be one. In the present note we determine the limit of $f_n = \sum_{j=1}^{n} \alpha_j \phi_j$ as n becomes large.

R.1. and some other generalizations of it, will follow from our main result (Theorem 1).

2. **Results.** Define a new scalar product $(.\,|\,.)_t$ on $\mathscr{D}(t)$ by $(u\,|\,v)_t = \text{Re.}\, t(u, v)$, [1, pp. 309–10] and complete $\mathscr{D}(t)$ in the new metric to a Hilbert space $\mathscr{H}_t$. Let the closure of t be $\bar{t}$. We have that $\mathscr{D}(t) \subset \mathscr{D}(\bar{t}) = \mathscr{H}_t \subset \mathscr{H}$. The norm in $\mathscr{H}_t$ will be denoted by $\|.\|_t$. Also $\mathscr{B}(X, Y)$ will denote the space of bounded operators with $\mathscr{D}(\cdot) \subset X$ and range $\mathscr{R}(\cdot) \subset Y$, and $\mathscr{B}(X) = \mathscr{B}(X, X)$.

LEMMA 1. *Let t be as in equation* (3), $\{\phi_i\} \subset \mathscr{D}(t)$ and $g \in \mathscr{H}$. *Equation* (3) *is equivalent to*

$$(4) \qquad \sum_{j=1}^{n} \alpha_j(\phi_i\,|\,[1 - T(z)]\phi_j)_t = -(\phi_i\,|\,Bg)_t \qquad i = 1, \cdots, n\,;$$

where $B \in \mathscr{B}(\mathscr{H}, \mathscr{H}_t)$, $T(z) = (zB_t - C) \in \mathscr{B}(\mathscr{H}_t)$ *and* B_t *is the restriction of* B *to* $\mathscr{D}(t)$.

Proof. Since $t_1 = (t - \text{Re.}\, t)$ is a bounded form on $\mathscr{H}_t$ [1, p. 314], there is a $C \in \mathscr{B}(\mathscr{H}_t)$ such that

$$t_1(u, v) = (u\,|\,Cv)_t;\ u, v \in \mathscr{D}(t)\,.$$

Also from Ref. [4] pp. 332–3, it follows that there is a unique $B \in \mathscr{B}(\mathscr{H}, \mathscr{H}_t)$ such that $\mathscr{D}(B) = \mathscr{H}$ and for $u \in \mathscr{H}_t$, $w \in \mathscr{H}$,

$$(5) \qquad (u\,|\,\omega) = (u\,|\,B\omega)_t\,.$$

In particular, in equation (3), $(\phi_i\,|\,g) = (\phi_i\,|\,Bg)_t$ and $(\phi_i\,|\,\phi_j) = (\phi_i\,|\,B\phi_j)_t = (\phi_i\,|\,B_t\phi_j)_t$.

The assertion now follows from direct substitution.

LEMMA 2. *In the notation of Lemma* 1, *we have that B_t, C are closable, B is closed and invertible and $B^{-1}(1 + \bar{C}) = A_t$ where A_t is the unique m-sectorial operator associated with $\bar{t}$.*

Proof. Since B_t and C are bounded and densely defined, they are closable. Since B is bounded and $\mathscr{D}(B) = \mathscr{H}$, it is closed. Invertibility of B has been proven in Reference [4] p. 333.

Now, $\mathscr{D}([B^{-1}(1 + \bar{C})]) \subset \mathscr{H}_t = \mathscr{D}(\bar{t})$ and for $u, v \in \mathscr{D}(t)$,

$$\begin{aligned}(u\,|\,B^{-1}(1 + \bar{C})v) &= (u\,|\,B^{-1}(1 + C)v) \\ &= (u\,|\,(1 + C)v)_t \qquad \text{(equation (5))} \\ &= t(u, v)\end{aligned}$$

From the closability of t, this result extends for $u, v \in \mathscr{H}_t$. The result now follows from Theorem 2.1, Chapter 6, Reference [1].

THEOREM 1. *In addition to the assumptions of Lemma* 1 *and* 2, *let* $\{\phi_i\}$ *be* l.i. *and complete in* $\mathscr{H}_t$, *and* z *be at a nonzero distance from* S. $f_n = \sum_{j=1}^{n} \alpha_j \phi_j$ *of equation* (3) *is then defined for each* n *and* $\lim_{n\to\infty} \| f_n - (z - A_t)^{-1} g \| = 0$.

Proof. From Lemma 1, equation (3) is equivalent to equation (4). Also without loss of generality, we may assume $\{\phi_i\}$ to be an orthonormal basis in $\mathscr{H}_t$. It is straightforward to check that (4) is equivalent to

$$(1 - T_n(z))f_n = -P_n B g$$

where $T_n(z) = P_n T(z) P_n$, and P_n is the ortho-projection on the n-dimentional subspace of $\mathscr{H}_t$ determined by $\{\phi_i\}$, $i = 1$ to n. It follows, for $h \in \mathscr{H}_t$, that

$$\lim_{n\to\infty} \| (T_n(z) - \bar{T}(z))h \|_t = 0 \, .$$

Also, since z is at a nonzero distance from S, dist. $(1, W(\bar{T}(z))) = d' > 0$, where $W(\cdot)$ denotes the numerical range. Further, since the spectrum of T_n, $\sigma(T_n) \subset (W(\bar{T}(z)) \cup \{0\})$, for each n, $(1 - T_n(z))^{-1} \in \mathscr{B}(\mathscr{H}_t)$ with $\| (1 - T_n(z))^{-1} \|_t \leqq 1/d$ where $d = \min.(1, d')$. Also $(1 - \bar{T}(z))^{-1} \in \mathscr{B}(\mathscr{H}_t)$.

Hence for $h \in \mathscr{H}_t$

$$\begin{aligned} &\| [(1 - T_n(z))^{-1} - (1 - \bar{T}(z))^{-1}]h \|_t \\ &\quad = \| (1 - T_n(z))^{-1}(T_n(z) - \bar{T}(z))(1 - \bar{T}(z))^{-1} h \|_t \\ &\quad \leqq \| (1 - T_n(z))^{-1} \|_t \, \| (T_n(z) - \bar{T}(z))(1 - \bar{T}(z))^{-1} h \|_t \\ &\quad \xrightarrow[n\to\infty]{} 0 \, . \end{aligned}$$

Further, for $g \in \mathscr{H}$,

$$\lim_{n\to\infty} \| (P_n B - B)g \|_t = 0$$

and hence

$$\lim_{n\to\infty} \| f_n - f \|_t = 0$$

where

$$\begin{aligned} f = -(1 - \bar{T}(z))^{-1} B g &= -(1 - z\bar{B}_t + \bar{C})^{-1} B g \\ &= (z - B^{-1}(1 + \bar{C}))^{-1} g \\ &= (z - A_t)^{-1} g \qquad \text{(Lemma 2)} \, . \end{aligned}$$

Assertion of the theorem follows by observing that $||\,.\,||_t \geqq ||\,.\,||$. For a symmetric t, $s = [b, \infty)$ with some $b > -\infty$, $\bar{C} = 0$ and $A_t = B^{-1}$ is self-adjoint.

In the following, f_n will stand for $\sum_{j=1}^{n} \alpha_j \phi_j$ as defined by equation (2).

COROLLARY 1. *Let A be densely defined sectorial operator and z be at a nonzero distance from its sector, $\{\phi_i\}$ be a* l.i. *basis in $\mathscr{D}(A)$. We have that* $\lim_{n\to\infty} ||f_n - (z - A_F)^{-1} g|| = 0$.

Proof. Define t of Theorem 1 by $t(u, v) = (u \mid Av)$, $u, v \in \mathscr{D}(A)$. t is closable from Theorem 1.27, Chapter 6 of [1]. Since $\{\phi_i\}$ is a l.i. basis in $\mathscr{D}(A)$ and $\mathscr{D}(A)$ is dense in $\mathscr{D}(\bar{t}) = \mathscr{H}_t$, it is a l.i. basis in $\mathscr{H}_t$. The result now follows from the fact that A_t of Theorem 1 now becomes A_F [1, pp. 325-6].

COROLLARY 2. *Let A be symmetric, bounded below by b, z be at a nonzero distance from $[b, \infty)$ and $\{\phi_i\}$ be a* l.i. *basis in $\mathscr{D}(A)$. Then* $\lim_{n\to\infty} ||f_n - (z - A_F)^{-1} g|| = 0$.

Proof. The result follows from Corollary 1, by noticing that the sector of A is $[b, \infty)$.

If the set $\{\phi_i\}$ is taken to be $\{A^i h\}$ for some $h \in \mathscr{D}(A^i)$ for $i = 0, 1, 2, \cdots$; the Bubnov-Galerkin method is called the method of moments [7]. Since $\{A^i h\}$ satisfies the conditions of Corollaries 1 and 2, the convergence of the method of moments also is established by these results. The result R.1 [2] thus is a special case of Corollary 2.

In Corollaries 1 and 2 we have considered the case of a densely defined A. In these results one can replace this condition by requiring that the form domain of A be dense. However since the Friedrichs extension is defined only for a densely defined A, the limit operator A_t may not be A_F. This situation is of a particular interest in Physics which we describe in brief.

Let A be given, formally, by $A = A_1 + A_2$, where A_1 and A_2 are symmetric but $\mathscr{D}(A) = \mathscr{D}(A_1) \cap \mathscr{D}(A_2)$ is not dense. However if the form domain of A is dense, the self-adjoint operator A_t associated with the form $t(u, v) = (u \mid (A_1 + A_2)v)$ is a legitimate operator to describe a physical system [5]. This construction enables one to include a larger class of interactions in the treatment than the requirement that A be densely defined [5]. It is obvious that the Bubnov-Galerkin method enables one to compute the resolvent of A_t in this case also, which is of prime importance in Physics.

ACKNOWLEDGEMENT. The author is thankful to Professor J. Nuttall for helpful discussions and his hospitality.

REFERENCES

1. T. Kato, *Perturbation Theory for Linear Operators*, Springer Verlag, N.Y., 1966.
2. D. Masson and P. Thewarapperuma, *On a connection between the method of moments and the Friedrichs extension*, University of Toronto Preprint.
3. S. G. Mikhlin, *Variational Methods in Mathematical Physics*, Pergamon Press, N.Y., 1964, ch. IX.
4. F. Riesz and B. Sz. Nagy, *Functional Analysis*; Unger, N.Y., 1971.
5. B. Simon, *Quantam Mechanics for Hamiltonians Defined as Quadratic forms*, Princeton University Press, 1971, Ch. 2.
6. S. R. Singh and A. D. Stauffer, Nuovo Cimento **22B** (1974) pp. 139-52.
7. Yu. V. Vorobyev, *Method of Moments in Applied Mathematics*, Gordon and Breach, N.Y., 1965.

Received February 10, 1976. Work supported in part by the National Research Council of Canada and the Centre for Interdisciplinary Studies in Chemical Physics, University of Western Ontario.

UNIVERSITY OF WESTERN ONTARIO