THE ROLE OF INTERPOLATION AND APPROXIMATION THEORY IN VARIATIONAL AND PROJECTIONAL METHODS FOR SOLVING PARTIAL DIFFERENTIAL EQUATIONS*

Richard S.VARGA
Department of Mathematics, Kent State University, Kent, Ohio 44242, USA

It is first shown for linear elliptic homogeneous boundary value problems how Galerkin methods for such problems lead naturally to questions concerning approximation-theoretic results for finite-dimensional subspaces. Using the theory of monotone operators, it is shown that the answers of such approximation-theoretic results apply to more general nonlinear boundary value problems. Then, a brief study of spline interpolation and approximation is coupled to the Galerkin error bounds. Included are recent results from Fourier transform methods.

1. HISTORY

It is quite interesting to look back at the theoretical developments in the numerical solution of, say, elliptic boundary value problems in the period 1955–1965, and to contrast them with today's developments. First of all, it may come as a bit of a surprise that a good bit of the research in that area in the period 1955–1965 largely centered about finite difference methods for *second*-order *linear* differential equations in two spatial variables. The reason for this is simple enough. The basic tool for an error analysis used then was the Gerschgorin-Collatz monotone matrix technique (cf. Collatz [8, p.348] and Forsythe and Wasow [10, p.283]), in which a nonsingular discrete matrix A_h, having all the entries of its inverse nonnegative, i.e., $A_h^{-1} \geqslant 0$, played a central role. This technique then essentially restricted attention to problems with positive Green's functions. Nonlinearities were difficult to treat by this approach, and higher order elliptic equations, such as eighth-order elliptic structures problems, were seldom theoretically considered.

Early in the 1960's, however, a renewed interest at General Motors in using spline functions, first pioneered by Schoenberg [13] in 1946, grew, and the results of this effort prompted numerical analysts to *reconsider* the classical Rayleigh-Ritz-Galerkin projectional methods for elliptic boundary value problems. This was the first key ingredient for the current developments in this area of numerical analysis: *the use of piecewise-polynomial or spline subspaces*. The next key ingredient was the development of a new (*and more general*) type of error analysis which wasn't restricted to linear second-order problems. This new error analysis, to be described in sects. 3,4, strongly focuses attention on approximation-theoretic results for spline functions, and is in fact the motivation for the title of this talk.

2. LINEAR ELLIPTIC PROBLEMS IN $\Omega \subset R^n$

For simplicity, consider the following linear elliptic homogeneous boundary value problem in a bounded region Ω in R^n whose boundary $\partial\Omega$ satisfies a restricted cone condition (cf. Agmon [1, p.11]):

$$\mathcal{L}u(x) = f(x)\,, \quad x \in \Omega\,,$$

$$D^\beta u(x) = 0\,, \quad x \in \partial\Omega\,, \quad \text{for all } |\beta| \leqslant m-1\,, \tag{2.1}$$

where

$$\mathcal{L}u(x) \equiv \sum_{|\alpha| \leqslant m} (-1)^\alpha D^\alpha \{p_\alpha(x) D^\alpha u(x)\}\,. \tag{2.2}$$

Here, we are using the usual standard multi-index notation (cf. Yosida [18, p.27]), i.e., $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n)$ is an n-tuple of nonnegative integers, and

$$D^\alpha \equiv \frac{\partial^{\alpha_1+\alpha_2+\ldots+\alpha_n}}{\partial x_1^{\alpha_1}\, \partial x_2^{\alpha_2} \ldots \partial x_n^{\alpha_n}}$$

* Work supported in part by A.E.C. Grant AT (II-I)-2075.

denotes the differential operator of order

$$|\alpha| \equiv \sum_{i=1}^{n} \alpha_i \ .$$

If $C_0^\infty(\Omega)$ is the space of all real-valued functions $u(x) = u(x_1,x_2,\ldots,x_n)$, infinitely differentiable in Ω with compact support in Ω, i.e., $u(x)$ vanishes identically outside some compact set contained in Ω, then for any nonnegative integer s,

$$\|u\|^2_{W_2^s(\Omega)} \equiv \sum_{|\alpha|\leqslant s} \|D^\alpha u\|^2_{L_2(\Omega)} \,, \quad \text{where}$$

$$\|v\|^2_{L_2(\Omega)} \equiv \int_\Omega v^2 \, \mathrm{d}x \,, \tag{2.3}$$

defines a norm on $C_0^\infty(\Omega)$, and the completion of $C_0^\infty(\Omega)$ in this norm serves to define the Sobolev space $\mathring{W}_2^s(\Omega)$. With this notation, we assume that the coefficients $p_\alpha(x)$ in (2.2) are real-valued and bounded in $\bar{\Omega}$, the closure of Ω, and that the *bilinear form* $a(u,v)$ associated with $\mathcal{L}$, defined by

$$a(u,v) \equiv \sum_{|\alpha|\leqslant m} \int_\Omega p_\alpha D^\alpha u D^\alpha v \, \mathrm{d}x \,,$$

$$u, v \in \mathring{W}_2^m(\Omega) \,, \tag{2.4}$$

is $\mathring{W}_2^m$ (R^n)*-elliptic*, i.e., (cf. Céa [6]), there is a constant $\rho > 0$ such that

$$a(u,u) \geqslant \rho\|u\|^2_{W_2^m(\Omega)} \quad \text{for all } u \in \mathring{W}_2^m(\Omega) \,. \tag{2.5}$$

Then, u is said to be a *generalized solution* of (2.1) if

$$a(u,v) = \int_\Omega fv \, \mathrm{d}x \quad \text{for all } v \in \mathring{W}_2^m(\Omega) \,. \tag{2.6}$$

Let S_k be any finite-dimensional subspace of $\mathring{W}_2^m(\Omega)$. Then, u_k is analogously said to be the *Galerkin approximation* of the solution of (2.1) in S_k if

$$a(u_k,w_k) = \int_\Omega fw_k \mathrm{d}x \quad \text{for all } w_k \in S_k \,. \tag{2.7}$$

The existence and uniqueness of the solutions of (2.6) and (2.7) poses no problems, as we shall see in sect. 3.

We now obtain an error bound for the difference $u_k - u$. From (2.6) and (2.7), it follows from the definition of $a(u,v)$ in (2.4) that

$$a(u_k - u, w_k) = 0 \quad \text{for all } w_k \in S_k \,. \tag{2.8}$$

Thus, from (2.5), (2.8), and the boundedness of the p_α in $\bar{\Omega}$, it also follows for some positive constant K (independent of u and S_k) that

$$\rho\|u_k - u\|^2_{W_2^m(\Omega)} \leqslant a(u_k - u, u_k - u) = a(u_k - u, w_k - u)$$

$$\leqslant K\|u_k - u\|_{W_2^m(\Omega)} \cdot \|w_k - u\|_{W_2^m(\Omega)}$$

$$\text{for all } w_k \in S_k \,.$$

Consequently, with $K' \equiv \rho^{-1}K$,

$$\|u_k - u\|_{W_2^m(\Omega)} \leqslant K' \inf_{w_k \in S_k} \|w_k - u\|_{W_2^m(\Omega)} \,. \tag{2.9}$$

This inequality obviously focuses attention on the approximation-theoretic questions of how $\inf_{w_k\in S_k} \|w_k - u\|_{W_2^m(\Omega)}$ depends on the smoothness of the generalized solution u, as well as on the choice of the particular finite-dimensional subspaces S_k of $\mathring{W}_2^m(\Omega)$. These questions will be considered in sect. 4. The next section shows that the basic inequality of (2.9) can be obtained for nonlinear differential equations also.

3. MONOTONE OPERATOR THEORY

In this section, we discuss briefly the theory of *monotone operators*, due to Zarantonello [19], Browder [4], and Minty [12]; see also [7]. Let H be a real Hilbert space with inner product $(\cdot,\cdot)_H$ and norm $\|\cdot\|_H$, and let T be a (possibly nonlinear) mapping from H into H which satisfies the following hypotheses:

(i) T is *finitely continuous*, i.e., for any finite-dimensional subspace H_k of H and any sequence $\{u_n\}_{n=1}^\infty$ of elements of H_k which converges to an element $u \in H$, then the sequence $\{(Tu_n,v)_H\}_{n=1}^\infty$ converges to $(Tu,v)_H$ for any $v \in H$;

(ii) T is *strongly monotone*, i.e., there exists a positive constant ρ for which

$$\rho\|u - v\|_H^2 \leqslant (Tu - Tv, u - v)_H \quad \text{for all } u, v \in H \,. \tag{3.1}$$

Consider then the problem of determining $u \in H$ such that

$$Tu = 0\,, \tag{3.2}$$

i.e.,

$$(Tu,v)_H = 0 \quad \text{for all } v \in H\,. \tag{3.3}$$

The abstract *Galerkin method*, corresponding to $Tu = 0$ in (3.2), consists in finding a u_k in H_k, where H_k is any finite-dimensional subspace of H, which satisfies

$$(Tu_k,v)_H = 0 \quad \text{for all } v \in H_k\,. \tag{3.4}$$

It is known (cf. Browder [4]) that if the mapping T is finitely continuous and strongly monotone, then there exists a unique u satisfying (3.3), and moreover, for *any* finite-dimensional subspace H_k of H, there is a unique u_k satisfying (3.4).

To study the convergence of the Galerkin approximation u_k in H_k to the solution u of (3.2), we now state a result of [7].

THEOREM 1. Let the mapping T: $H \to H$ be finitely continuous, strongly monotone, and *bounded*, i.e., T maps bounded subsets of H into bounded subsets of H. If u is the unique solution of $Tu = 0$, and u_k is its unique Galerkin approximation in H_k (cf. (3.4)), then there exists a positive constant K such that

$$\|u-u_k\|_H^2 \leqslant K \inf_{w_k \in H_k} \|u-w_k\|_H, \tag{3.5}$$

for *any* finite-dimensional subspace H_k of H. Moreover, if T is *Lipschitz continuous for bounded arguments*, i.e., given $M > 0$, there exists a constant $K(M)$ such that

$$\|Tu-Tv\|_H \leqslant K(M)\|u-v\|_H$$
$$\text{for all } u, v \in H \text{ with } \|u\|_H, \|v\|_H \leqslant M\,, \tag{3.6}$$

then there exists a positive constant K such that

$$\|u-u_k\|_H \leqslant K \inf_{w_k \in H_k} \|u-w_k\|_H\,, \tag{3.7}$$

for all finite-dimensional subspaces H_k of H.

The whole point of our discussion here on monotone operators is that the inequality (3.7) is a generalization of the inequality (2.9), and this in fact allows us to treat nonlinear versions of the homogeneous boundary value problem of (2.1). Specifically, if Ω is as before a bounded region in R^n, $n \geqslant 1$, whose boundary $\partial\Omega$ satisfies a restricted cone condition, consider the $2m$-th order nonlinear boundary value problem:

$$\begin{gathered}\sum_{|\alpha|\leqslant m} (-1)^{|\alpha|} D^\alpha\{A_\alpha(x,u(x),\ldots,D^m u(x))\} = 0\,, \quad x \in \Omega\,,\\ D^\beta u(x) = 0\,, \quad x \in \partial\Omega\,, \quad \text{for all } |\beta| \leqslant m-1\,,\end{gathered} \tag{3.8}$$

where $A_\alpha(x,u,\ldots,D^m u)$ denotes a function which can depend upon x and *any* $D^\gamma u$ with $|\gamma| \leqslant m$. Associated with this nonlinear boundary value problem of (3.8) is the *quasi-bilinear form* $a(u,v)$, the analogue of (2.4), defined by

$$\begin{gathered}a(u,v) \equiv \sum_{|\alpha|\leqslant m} \int_\Omega A_\alpha(x,u,\ldots,D^m u)\cdot D^\alpha v\, dx\,,\\ u, v \in \mathring{W}_2^m(\Omega)\,.\end{gathered} \tag{3.9}$$

With suitable growth conditions on the functions A_α, it can be shown [7] that for any $u \in \mathring{W}_2^m(\Omega)$, there exists a positive constant $K = K_u$, depending on u, such that

$$|a(u,v)| \leqslant K_u \cdot \|v\|_{W_2^m(\Omega)} \quad \text{for all } v \in \mathring{W}_2^m(\Omega). \tag{3.10}$$

As a consequence of (3.10), the quasi-bilinear form is, for each $u \in \mathring{W}_2^m(\Omega)$, a bounded linear functional in v on $\mathring{W}_2^m(\Omega)$. As such, the Riesz Representation Theorem [18, p.90] gives us that there is a unique $Tu \in \mathring{W}_2^m(\Omega)$ such that

$$a(u,v) = (Tu,v)_{W_2^m(\Omega)} \quad \text{for all } v \in \mathring{W}_2^m(\Omega)\,, \tag{3.11}$$

where $(v,w)_{W_2^m(\Omega)} \equiv \sum_{|\alpha|\leqslant m} (D^\alpha v, D^\alpha w)_{L_2(\Omega)}$ denotes the usual inner product on $\mathring{W}_2^m(\Omega)$. This then defines an abstract nonlinear mapping T from $\mathring{W}_2^m(\Omega)$ into $\mathring{W}_2^m(\Omega)$. If this mapping T, defined through (3.11), satisfies all the hypotheses of theorem 1, then of course the error bounds of (3.5) and (3.7) are valid, and, as shown in [7], there are interesting cases of nonlinear boundary value problem (3.8) for which the conclusions of Theorem 1 are valid.

4. ERROR BOUNDS FOR SPLINE APPROXIMATION

Splines have certainly grown in popularity since Schoenberg's fundamental results [13] in 1946, and some 600 papers have now been written on splines and their applications. To give some of the current known error bounds for spline interpolation and approximation, we first partition the finite interval $[a,b]$ by means of Δ: $a = x_0 < x_1 < \ldots < x_N = b$. If $\pi \equiv \max_{0 \leq i \leq N-1} (x_{i+1} - x_i)$, $\underline{\pi} \equiv \min_{0 \leq i \leq N-1} (x_{i+1} - x_i)$, let $\mathcal{P}_\sigma(a,b)$ denote all partitions Δ of $[a,b]$ for which $\pi \leq \sigma \underline{\pi}$. Next, consider the m-th order operator

$$Lu(x) \equiv \sum_{j=0}^{m} c_j(x) D^j u(x) , \tag{4.1}$$

where $c_j \in C^j[a,b]$, $j = 0, 1, \ldots, m$, with $c_m(x) \geq \delta > 0$ for all $x \in [a,b]$. Next, let z be any (fixed) positive integer with $1 \leq z \leq m$. Then, $Sp(L,\Delta,z)$, the *L-spline space*, is the collection of all real-valued functions w defined on $[a,b]$ such that

$$L^*L\,w(x) = 0 , \quad \text{for all } x \in (a,b) - \{x_i\}_{i=1}^{N-1} , \text{ with}$$

$$D^k w(x_i-) = D^k w(x_i+) \quad \text{for } 0 \leq k \leq 2m - 1 - z ,$$

$$0 < i < N , \tag{4.2}$$

where L^* is the formal adjoint of L. It can be shown that $Sp(L,\Delta,z)$ is a linear space, with $Sp(L,\Delta,z) \subset W_2^{2m-z}[a,b]$ (cf. [14]). In the important special case $L = D^m$, the elements of $Sp(D^m,\Delta,z)$ are, from (4.2), polynomials of degree $2m-1$ on each subinterval of Δ, and as such are called *polynomial splines.*

We now discuss the possibility of *interpolation* of given functions by elements in $Sp(L,\Delta,z)$. Given any $f \in C^{m-1}[a,b]$, it is easy to see (cf. [14]) that there is a unique $s \in Sp(L,\Delta,z)$ which interpolates f in the sense that

$$D^j(f-s)(x_i) = 0 , \quad j = 0, 1, \ldots, z-1 , \quad \text{if } 0 < i < N ,$$
$$D^j(f-s)(a) = D^j(f-s)(b) = 0 , \quad j = 0, 1, \ldots, m-1 , \tag{4.3}$$

for which the following error bounds are typical (cf. [16] and [17]).

THEOREM 2. Let $f \in W_2^\sigma[a,b]$ with $m \leq \sigma \leq 2m$. If s is the unique element in $Sp(L,\Delta,z)$ interpolating f in the sense of (4.3), then

$$\|D^j(f-s)\|_{L_2[a,b]} \leq \|f-s\|_{W_2^j[a,b]} \leq K\pi^{\sigma-j} \|f\|_{W_2^\sigma[a,b]} \tag{4.4}$$

for any $j = 0, 1, \ldots, m$.

We remark that the quantity σ in Theorem 2 need *not* be an integer; the interpretation of $W_2^\sigma[a,b]$ is then made through the use of *interpolation space theory* (cf. [5]), yet another useful tool to numerical analysts today.

More general forms of Theorem 2 are known, these generalizations coming from more general (Besov) spaces, more general types of interpolation, and more genral types of differential operators. For a survey of such results, see for example [17].

In particular, the inner inequality of (4.4) of Theorem 2 gives us that

$$\inf_{w \in Sp(L,\Delta,z)} \|f-w\|_{W_2^j[a,b]} \leq K\pi^{\sigma-j} \|f\|_{W_2^\sigma[a,b]} , \tag{4.5}$$

and the exponent of π can be shown to be sharp, i.e., it cannot be increased for the function classes considered. It is natural to ask when such optimal approximation holds in higher dimensions as well. This is the topic of the next section. Note that the inequalities of (4.4) and (4.5) then directly apply to the error estimation of spline subspaces $Sp(L,\Delta,z)$ in a Galerkin setting for one-dimensional problems (n=1). In a completely analogous way, these error estimates for spline interpolation can be applied also to tensor products of one-dimensional problems.

5. FOURIER TRANSFORM METHODS

One of the significant difficulties in applying the Galerkin method is the requirement of finding finite-dimensional subspaces whose elements satisfy all *essential* boundary data, (e.g., $D^\beta u(x) = 0$ for all $x \in \partial\Omega$ and all $|\beta| \leq m-1$ in (2.1) and (3.8)). There are several ways around this. If one is given a boundary value problem with Neumann boundary conditions, then one has *no* essential boundary restrictions. Similarly, in the case of problems with periodic boundary conditions defined on hypercubes, or problems defined on all of R^n, both lead to problems in R^n for which there are no essential boundary restrictions. The point is that one can then make strong use of the tool of Fourier transforms, thus following the route of differential equations theory. We shall describe some recent penetrating results by Strang and Fix [15].

Start with a *fixed* function $\phi(x) \in W_2^p(R^n)$ with compact support (written $\phi \in (W_2^p(R^n))_0$), i.e., $\phi(x)$ is identically zero for all

$$|x| = \left(\sum_{i=1}^{n} x_i^2\right)^{\frac{1}{2}}$$

sufficiently large. If Z^n denotes the collection of all n-tuples $j = (j_1, j_2, \ldots, j_n)$ of integers, then from the single ϕ, other functions, viz.,

$$\phi_j^h(x) \equiv h^{-n/2}\phi(\frac{x}{h} - j), \quad j \in Z^n, \quad x \in R^n, \quad (5.1)$$

can be constructed, where h is a positive parameter with $0 < h \leqslant 1$. From this single ϕ, weighted sums

$$\sum_{j \in Z^n} w_j^h \theta_j^h(x)$$

can be formed, and it is of interest to couple the approximation-theoretic behavior of such weighted sums to properties of the single function ϕ. This has been done by Strang and Fix [15], and we state the following result of [15]. For notation,

$$\hat{\phi}(\xi) \equiv \int_{R^n} e^{-ix\xi}\, \phi(x)\, dx \quad (5.2)$$

denotes the Fourier transform of ϕ.

THEOREM 3. Let $\phi \in (W_2^p(R^n))_0$. Then, the following are equivalent:

(i) $\hat{\phi}(0) \neq 0$, but $D^\alpha \hat{\phi}(2\pi j) = 0$ for all $0 \neq j \in Z^n$, $|\alpha| \leqslant p$;

(ii) for any $|\alpha| \leqslant p$,

$$\sum_{j \in Z^n} j^\alpha \phi(t-j)$$

is a polynomial in $t_1, \ldots, t_n$ with leading coefficient Ct^α, $C \neq 0$;

(iii) for any $u \in W_2^{p+1}(R^n)$, there exist weights w_j^h and constants c_s and K, independent of h, such that as $h \to 0$,

$$\|u - \sum_{j \in Z^n} w_j^h \phi_j^h\|_{W_2^s(R^n)} \leqslant c_s h^{p+1-s} \|u\|_{W_2^{p+1}(R^n)} \quad (5.3)$$

for $0 \leqslant s \leqslant p$, with

$$\sum_{j \in Z^n} |w_j^h|^2 \leqslant K \|u\|^2_{W_2^0(R^n)}.$$

The results of Theorem 3 have been used to prove the following improved form of the inequality of (2.9), now with $\Omega = R^n$. If the generalized solution u of (2.6) is in $W_2^{p+1}(R^n)$, with $p+1 \geqslant m$, let the subspace S^h of all sums of the form

$$\sum_{j \in Z^n} w_j^h \phi_j^h(x),$$

where ϕ satisfies the hypotheses of Theorem 3. Then [15], the Galerkin approximation u^h in S^h satisfies

$$\|u-u^h\|_{W_2^s(R^n)} \leqslant Kh^\sigma \|u\|_{W_2^{p+1}(R^n)}, \quad 0 \leqslant s \leqslant m, \quad (5.4)$$

where $\sigma \equiv \min\{p+1-s;\ 2(p+1-m)\}$. Results in other norms have similarly been obtained (cf. [2]).

Finally, just as results from interpolation and approximation theory are used in an essential way in the analysis of Galerkin methods, the same is true of the *least squares methods*, recently investigated by Bramble and Schatz [3]. Moreover, the idea of using interpolation-and-approximation-theoretic results is not confined just to elliptic boundary value problems; recently again, they play an essential role in the analysis of parabolic Galerkin methods (cf. Douglas and Dupont [9], Strang and Fix [15], and [17]).

REFERENCES

[1] S.Agmon, Lectures on Elliptic Boundary Value Problems (Van Nostrand Co., Princeton, New Jersey, 1965).

[2] C.de Boor and G.Fix, Spline approximation by quasi-interpolants (to appear).

[3] J.H.Bramble and A.H.Schatz, Rayleigh-Ritz-Galerkin methods for Dirichlet's problem using subspaces without boundary conditions, Comm. Pure Appl. Math. 23 (1970) 653-676.

[4] F.E.Browder, Existence and uniqueness theorems for solutions of nonlinear boundary value problems, Proc. Amer. Math. Soc. Symposia in Appl. Math. 17 (1965) 24-49.

[5] P.L.Butzer and H.Berens, Semi-Groups of Operators and Approximations (Springer-Verlag, New York, 1967).

[6] J.Céa, Approximation variationnelle des problèmes aux limites, Ann. Inst. Fourier (Grenoble) 14 (1964) 345-444.

[7] P.G.Ciarlet, M.H.Schultz and R.S.Varga, Numerical methods of high-order accuracy for nonlinear boundary value problems, Monotone operator theory, Nummer. Math. 13 (1969) 51–77.

[8] L.Collatz, The Nummerical Treatment of Differential Equations, third edition (Springer-Verlag, Berlin, 1959).

[9] J.Douglas and T.Dupont, Galerkin methods for parabolic equations, SIAM J. Numer. Anal. 7 (1970) 575-626.

[10] G.E.Forsythe and W.R.Wasow, Finite Difference Methods for Partial Differential Equations (John Wiley and Sons, Inc., New York, 1960).

[11] G.W.Hedstrom and R.S.Varga, Application of Besov spaces to spline approximation, J. Approx. Theory (to appear).

[12] G.Minty, Monotone (nonlinear) operators in Hilbert space, Duke Math. J. 29 (1962) 341-346.
[13] I.J.Schoenberg, Contributions to the problem of approximation of equidistant data by analytic functions, Parts A and B, Quart. Appl. Math. 4 (1946) 45-99; 112-141.
[14] M.H.Schultz and R.S.Varga, L-splines, Numer. Math. 10 (1967) 345-369.
[15] G.Strang and G.Fix, Analysis of the Finite Element method (Prentice-Hall, Inc.) (to appear).
[16] B.K.Swartz and R.S.Varga, Error bounds for spline and L-spline interpolation, J. Approx. Theory (to appear).
[17] R.S.Varga, Functional analysis and approximation theory in numerical analysis, AMS-SIAM Publication (1971).
[18] K.Yosida, Functional Analysis (Academic Press, New York, 1965).
[19] E.H.Zarantonello, Solving functional equations by contractive averaging, MRC Technical Report #160 (1960), Math. Research Center, University of Wisconsin, Madison, Wisconsin.