QUADRATURE-GALERKIN APPROXIMATIONS

TO SOLUTIONS OF ELLIPTIC DIFFERENTIAL EQUATIONS

by

Martin H. Schultz

#9   71-8

A popular way of obtaining approximate solutions of a large class of elliptic boundary value problems is Galerkin's method. In practice, these approximate solutions are not actually obtained since the Galerkin formulation involves certain integrals which cannot be evaluated analytically. Instead these integrals are approximated numerically and the resulting equations are solved to give "quadrature approximations" to the Galerkin approximations.

The object of this paper is to investigate a priori bounds for the difference between the solutions and the quadrature approximations. In particular, using a technique due to J. Nitsche, cf. [2], we obtain new bounds in the L^2 - norm. These substantially improve and generalize the weaker results reported in [1] which involved analogous bounds in Sobolev spaces.

Throughout this paper, let $\Omega \subset R^N$ be an open set, m be a positive integer, $W_0^{m,2}(\Omega)$ denote the closure of the real-valued functions $f \in C_0^\infty(\Omega)$ with respect to the norm

$$\| f \|_m \equiv \left(\int_\Omega \sum_{0 \le |\alpha| \le m} |D^\alpha f|^2 \, dx \right)^{1/2} \quad ,$$

where we have used the standard multi-index notation, cf. [5], $W^{m,2}(\Omega)$ denote the closure of the real-valued functions $f \in C^\infty(\Omega)$

with respect to $\|\cdot\|_m$, and H be any closed subspace of $W^{m,2}(\Omega)$ such that $W_0^{m,2}(\Omega) \subset H \subset W^{m,2}(\Omega)$.   If we are considering a problem of order 2m, the a priori bounds reported in [1] are in the Sobolev norm $\|\cdot\|_m$. Such error bounds trivially induce error bounds in the lower order Sobolev norms, $\|\cdot\|_j$, $0 \le j \le m - 1$. But these induced error bounds are not sharp.

Let $a(u,v)$ be a real-valued, bounded, bilinear form over H such that there exist constants $0 < \delta \le \mu$ such that

$$(1) \qquad a(u,v) \le \mu \quad \|u\|_m \quad \|v\|_m$$

and

$$(2) \qquad \delta \|u\|_m^2 \le a(u,u) \quad,$$

for all u and $v \in$ H. Given a real-valued function $g \in W^{0,2}(\Omega)$, our problem is to find $u \in$ H such that

$$(3) \qquad a(u,v) = (g,v)_0 \equiv \int_\Omega g(x)v(x)\,dx, \quad \text{for all } v \in H.$$

If S is any finite dimensional subspace of H, the Galerkin method is to find $u_S \in$ S such that

$$(4) \qquad a(u_S, w) = (g,w)_0, \qquad \text{for all } w \in S.$$

If S is the span of the N linearly independent basis functions

$\{B_i(x) \mid 1 \leq i \leq N\}$, then the coefficients $\{\beta_i \mid 1 \leq i \leq N\}$ in

the expansion $u_S(x) \equiv \sum_{i=1}^{N} \beta_i B_i(x)$ can be characterized as the

solution of the linear system

(5) $A\underline{\beta} = \underline{k}$,

where

(6) $A \equiv [a_{ij}] \equiv [\int_{\Omega} a(B_i, B_j) \, dx]$

and

(7) $\underline{k} \equiv [k_i] \equiv [\int_{\Omega} g(x) B_i(x) dx]$.

We recall two key results, which will be essential,

giving bounds for the difference $u-u_S$. The reader is referred

to [3] and [4] for the details of the proofs.

Theorem 1. If (1) and (2) hold, (3) has a unique solution, u,

(4) has a unique solution, u_S, for every finite dimensional

subspace S of H and

$$(8) \qquad \| u - u_S \|_m \leq \delta^{-1} \mu \inf_{y \in S} \| u - y \|_m.$$

We now make the additional assumption that $a(u,v)$ is strongly coercive over H, i.e., that every solution, u, of (3) is in $W^{2m,2}(\Omega)$ and that there exists a positive constant, ρ, such that

$$(9) \qquad \| u \|_{2m} \leq \rho \| g \|_0, \qquad\qquad \text{for all } g \in W^{0,2}(\Omega).$$

Theorem 2. Let (1) and (2) hold, $a(u,v)$ be strongly coercive over H and C be a collection of finite dimensional subspaces of H such that if $g \in W^{2m,2}(\Omega) \cap H$, then there exists a positive, real-valued function E on C such that

$$(10) \qquad \inf_{y \in S} \| g - y \|_m \leq E(S) \| g \|_p,$$

for some $m \leq p \leq 2m$ and all $S \in C$. Then

$$(11) \qquad \| u - u_S \|_0 \leq \rho \delta^{-1} \mu E(S) \inf_{y \in S} \| u - y \|_m, \qquad \text{for all } S \in C.$$

In practice, instead of solving (4), we solve

$$(12) \qquad a(\hat{u}_S, w) = (\hat{g}, w)_0,$$

for all $w \in S$, where $\hat{g} \in W^{0,2}(\Omega)$ is a "good" approximation to g

such that the integrals $\int_\Omega \hat{g}(x) B_i(x)dx$, $1 \le i \le N$, can be evaluated analytically. First we give a bound for the error $(u - \hat{u}_S)$ in the $W^{m,2}(\Omega)$ - norm, which generalizes Theorem 3 of [1].

<u>Theorem 3.</u> If the hypotheses of Theorem 1 hold,

$$(13) \qquad \|u - \hat{u}_S\|_m \le \delta^{-1} \|g - \hat{g}\|_0 + \delta^{-1}\mu \inf_{y \in S} \|u - y\|_m$$

for all finite dimensional subspaces S of H and all $\hat{g} \in W^{0,2}(\Omega)$.

<u>Proof.</u> By the triangle inequality, we have

$$(14) \qquad \|u - \hat{u}_S\|_m \le \|u_S - \hat{u}_S\|_m + \|u - u_S\|_m .$$

From (4) and (12) we have

$$\delta \|u_S - \hat{u}_S\|_m \le a(u_S - \hat{u}_S, u_S - \hat{u}_S) = (g - \hat{g}, u_S - \hat{u}_S)_0$$

$$\le \|g - \hat{g}\|_0 \|u_S - \hat{u}_S\|_0 \le \|g - \hat{g}\|_0 \|u_S - \hat{u}_S\|_m .$$

Hence,

$$(15) \qquad \|u_S - \hat{u}_S\|_m \le \delta^{-1} \|g - \hat{g}\|_0$$

and the result follows from using (8) and (15) to bound the right-hand side of (14). QED.

Second, using a technique of J. Nitsche, cf. [2] , we give a bound for the error $(u - \hat{u}_S)$ in the $W^{0,2}(\Omega)$ - norm, which extends Theorem 3 of [1].

Theorem <u>4</u>. If the hypotheses of Theorem 2 hold, then

$$(16) \qquad \|u - \hat{u}_S\|_0 \ \leq \ \rho(1 + \mu\delta^{-1}E(S)) \ \|g - \hat{g}\|_0$$

$$+ \ \rho\delta^{-1}\mu^2 E(S) \ \inf_{y \in S} \|u - v\|_m \ ,$$

for all finite dimensional subspaces S of H and all $\hat{g} \in W^{0,2}(\Omega)$.

<u>Proof</u>. Let $w_S \equiv (u - \hat{u}_S) \, / \, \|u - \hat{u}_S\|_0$ and ϕ_S be the solution of $a(u - \hat{u}_S, \phi_S) = (w_S, u - \hat{u}_S)_0 = \|u - \hat{u}_S\|_0$. If $v \in S$ is the orthogonal projection of ϕ_S onto S with respect to the inner-product in $W^{m,2}(\Omega)$, $a(u - \hat{u}_S, \phi_S - v) = (w_S, u - \hat{u}_S)_0 + (\hat{g} - g, v)_0$ and hence $\|u - \hat{u}_S\|_0 = a(u - \hat{u}_S, \phi_S - v) + (g - \hat{g}, v)_0$. Thus, using (1), we have

$$(17) \quad \|u - \hat{u}_S\|_0 \ \leq \ \mu \ \|u - \hat{u}_S\|_m \ \|\phi_S - v\|_m \ + \ \|g - \hat{g}_0\| \ \|v\|_0$$

$$\leq \ \mu \ E(S) \ \|\phi_S\|_p \ \|u - \hat{u}_S\|_m + \|\phi_S\|_m \ \|g - \hat{g}\|_0$$

$$\leq \ \mu \ E(S) \ \|\phi_S\|_{2m} \ \|u - \hat{u}_S\|_m \ + \ \|\phi_S\|_{2m} \ \|g - \hat{g}\|_0$$

$$\leq \ \rho\mu \ E(S) \ \|u - \hat{u}_S\|_m \ + \ \rho \ \|g - \hat{g}\|_0 \ .$$

The result follows by using (13) to bound the right-hand side of (17). QED.

As a rather typical example, let H and $a(u,v)$ be such that if $g \in W^{t,2}(\Omega)$, then the solution $u \in W^{2m+t,2}(\Omega)$ and C be a one-parameter family of finite dimensional subspaces $\{S_h(d) \mid h \geq 0\}$, d a fixed integer $\geq m$, such that $E(S_h(d)) \equiv \inf\limits_{y \in S_h(d)} \|g - y\|_m$ $\leq K_d h^{r-m} \|g\|_r$ and $\inf\limits_{y \in S_h(d)} \|u - y\|_m \leq C_d h^{z-m} \|u\|_z$, where $r \equiv \min(2m, d+1)$ and $z \equiv \min(2m+t, d+1)$. For example, $S_h(d)$ might be a space of polynomial spline functions of degree d with associated mesh length equal to h. Then

$$\|u - \hat{u}_S\|_m \leq \|g - \hat{g}\|_0 + \delta^{-1} \mu C_d h^{z-m} \|u\|_z \qquad \text{and}$$

$$\|u - \hat{u}_S\|_0 \leq \rho (1 + \mu\delta^{-1} K_d h^{r-m}) \|g - \hat{g}\|_0 + \rho\delta^{-1} \mu^2 K_d C_d h^{r+z-2m} \|u\|_z.$$

If in addition, $\hat{g}$ is given by an interpolation mapping of degree p such that

$$\|g - \hat{g}\|_0 \leq M_q h^q \|g\|_q, \qquad \text{where } q \equiv \min(p+1, t),$$

then we have

$$\|u - \hat{u}_S\|_m \leq M_q h^q \|g\|_q + \delta^{-1} \mu C_d h^{z-m} \|u\|_z$$

and

$$\|u - \hat{u}_S\|_0 \leq \rho(1+\mu\delta^{-1}K_dh^{r-m}) M_qh^q \|g\|_q + \rho\delta^{-1}\mu^2K_dC_dh^{r+z-2m}\|u\|_z.$$

In [1], a quadrature scheme was defined to be (m-) consistent if and only if $q \geq z-m$, i.e., if and only if the error in the $W^{m,2}(\Omega)$ - norm due to the quadrature scheme is of the same order of magnitude as the error due to Galerkin's method. However, here we see that if we demand consistency in the $W^{0,2}(\Omega)$ - norm, we must have $q \geq r+z-2m$. But $r+z-2m \geq (m+1) + z -2m \geq z-m+1$, which is a more stringent condition on the regularity of g and the degree of the interpolation mapping. In the still further special case of $r=2m$ and $z=d+1$ we have m-consistency if and only if $q \geq d+1-m$ and 0-consistency if and only if $q \geq d+1$. Thus, for 0-consistency we must have $t \geq d+1$ and $p \geq d$.

REFERENCES

[1]   Herbold, R. J., M. H. Schultz, and R. S. Varga,
      Quadrature Schemes for the Numerical Solution of
      Boundary Value Problems. _Aequationes Mathematicae_ 3,
      247-270 (1969).

[2]   Nitsche, J., Ein kriterium fur die quasi-optimalitat
      des Ritzschen verfohrens. _Num. Math._ 11, 346-348 (1968).

[3]   Schultz, M. H., Multivariate spline functions and
      elliptic problems. _Approximations with Special Emphasis
      on Spline Functions_, 279-347. New York: Academic Press (1969).

[4]   Schultz, M. H., L^2 - Error bounds for the Rayleigh-Ritz-
      Galerkin method. To appear in _SIAM J. Num. Anal_.

[5]   Yosida, K., Functional Analysis (458 pp.) New York:
      Academic Press (1965).