

## On a Dimensional Reduction Method I. The Optimal Selection of Basis Functions\*

By M. Vogelius and I. Babuška

**Abstract.** This paper is the first in a series of three, which analyze an adaptive approximate approach for solving  $(n + 1)$ -dimensional boundary value problems by replacing them with systems of equations in  $n$ -dimensional space. In this approach the unknown functions of  $(n + 1)$  variables are projected onto finite linear combinations of functions of just  $n$  variables.

This paper shows how the coefficients of these linear combinations can be chosen optimally.

**1. Introduction.** Let  $\Omega^h = \omega \times [-h, h]$  be a domain in  $\mathbb{R}^{n+1}$ , and let  $u^h$  be the solution to some elliptic boundary value problem on  $\Omega^h$ . We wish to find—in a very effective way—an approximate solution  $u_{\text{approx}}^h$  satisfying a certain accuracy requirement.

Considering the special structure of  $\Omega^h$ , we expect that  $u^h$  can be approximated well by a linear combination

$$\sum_{j=0}^N \psi_j(y/h) \cdot x_j, \quad y \in [-h, h],$$

of  $N + 1$  functions  $\{x_j\}_{j=0}^N$  on  $\omega$ . Methods built on an assumption of this type and a projection procedure are widely used in engineering. As an example, we mention various theories for plates, beams, etc.; cf. [4], [11], [12]. These methods are also sometimes associated with the name of L. V. Kantorovich; cf. [9].

Our goals are to select the family of functions  $\{\psi_j\}_{j=0}^\infty$  such that

(i) The method is optimally accurate when  $h$  is small and the data sufficiently regular.

(ii) For arbitrary  $h$  and input data the method converges as  $N \rightarrow \infty$ .

(iii) It is possible to derive an a posteriori estimate for the error, and this leads to an effective procedure for the selection of  $N$ .

Another approach, which has also been extensively used in structural mechanics and elsewhere to derive lower-dimensional approximating models, is asymptotic expansion in  $h$ ; cf. [5], [7], [8]. We refer to [17] and references therein for various engineering applications of this approach. It is quite obvious that a method based on an asymptotic expansion in  $h$  does not satisfy the goals (ii) and (iii) stated

---

Received April 18, 1980; revised October 22, 1980.

1980 *Mathematics Subject Classification.* Primary 65N30, 41A30.

\* This work was in part supported by the Office of Naval Research under contract N00014-77-C-0623. In addition the first author was partially supported by the University of Aarhus, Denmark. The computations were supported by the Computer Science Center of the University of Maryland.

© 1981 American Mathematical Society  
0025-5718/81/0000-0102/\$05.00

above. The validity of an approach of this type is dependent on the smallness of  $h$ , while the actual value of  $h$  may simply not be small enough. Even for arbitrarily small  $h$ , the approximation can be unsatisfactory because of rough data. [1] contains an example which shows, for an extremely thin, simply-supported polygonal plate, that the biharmonic equation is not always a good model for a three-dimensional plate. In models based on asymptotic expansion are also often included boundary corrector terms; cf. [10]. These terms are not of the simple form considered here—they are not lower dimensional and for practical purposes they are often very difficult to deal with. Instead of including boundary corrector terms, we suggest increasing  $N$  (in an adaptive way) to achieve the desired accuracy independently of input data.

For the model problem analyzed in this paper we restrict ourselves to projection in the energy. We project the solution  $u^h$  on elements of the form

$$\sum_{j=0}^N \psi_j(y/h)x_j,$$

where  $\{\psi_j\}_{j=0}^\infty$  is a sequence of functions on  $[-1, 1]$ . The variable  $y$  ranges over  $[-h, h]$  and the  $x_j$ 's are arbitrary elements of some linear space (e.g., a function space on  $\omega$ ). The main question addressed here is how to attain the first of our goals, namely that the rate of accuracy be optimal for small  $h$ . In Theorems 3.1 and 4.1, we prove that this requirement almost uniquely determines the sequence  $\{\psi_j\}_{j=0}^\infty$ .

Though the scope of this paper is very much different from that of a formal asymptotic expansion, it is not surprising that particularly some of the techniques used in the proof of Theorem 3.1 are similar. We have chosen to give a detailed proof for two reasons, first of all because elements of it are needed in the proof of Theorem 4.1 and secondly because it is not very long.

In another paper, [15], we show that this method of dimensional reduction fulfills the second of our goals. In addition we also give various estimates of the rate of convergence as  $N \rightarrow \infty$ .

If a singularity is present in the data, we must use a relatively high number of functions  $\psi_j$  (i.e., we must increase  $N$ ). Because such singularities are often localized, it is appropriate to introduce the possibility of using a different  $N$  in different parts of the domain  $\Omega^h$ . The a posteriori error estimation (and the problem of how to design an adaptive algorithm that will produce a good distribution of the  $N$ 's) was briefly discussed in [14]. It will be given a more detailed treatment in a forthcoming paper; cf. [16].

**2. Notation and the Model Problem.** Let  $\mathcal{H}$  be a separable Hilbert space with the inner product  $\langle u, v \rangle$  and the norm  $\|u\| = \langle u, u \rangle^{1/2}$ .

$A$  denotes a selfadjoint, linear (unbounded) operator in  $\mathcal{H}$  with a domain of definition  $\mathcal{D}(A)$ . Furthermore we assume that  $A$  is a strictly positive-definite operator, i.e., there exists  $C > 0$  such that

$$\forall u \in \mathcal{D}(A): C\|u\|^2 \leq \langle Au, u \rangle.$$

$\mathcal{D}(A^{1/2})$  is itself a Hilbert space with the inner product  $\langle u, v \rangle + \langle A^{1/2}u, A^{1/2}v \rangle$ .

Let  $M$  be a selfadjoint bounded linear operator in  $\mathcal{H}$ .  $M$  is also assumed to be a strictly positive-definite operator.

$I$  denotes an interval on the real line.  $L^2(I; \mathfrak{C})$  is then defined as the set of strongly measurable functions  $I \rightarrow \mathfrak{C}$  such that  $\|u(\cdot)\|$  is an element of  $L^2(I)$ ; cf. [6]. The same goes for  $L^2(I; \mathfrak{D}(A^{1/2}))$ .

We also need a Sobolov space of functions with values in  $\mathfrak{C}$ .  $H^1(I; \mathfrak{C})$  denotes the space of functions  $I \rightarrow \mathfrak{C}$  such that  $u(\cdot) \in L^2(I; \mathfrak{C})$  and  $du(\cdot)/dy \in L^2(I; \mathfrak{C})$ ; cf. [2]. The derivative here is taken in the distributional sense.  $H^1(I)$  denotes the standard Sobolev space on  $I$ .

Assume  $a$  and  $b$  are real-valued functions in  $L^\infty([-1, 1])$  such that

$$a_0 \leq a(y), \quad b_0 \leq b(y),$$

for some constants  $a_0 > 0, b_0 > 0$ .  $a_h$  and  $b_h \in L^\infty([-h, h])$  are then defined as

$$a_h(y) = a(y/h), \quad b_h(y) = b(y/h).$$

By  $P_h(d/dy)$  we denote the differential operator  $-(d/dy)(a_h d/dy)$ .

Let  $f$  and  $g$  be two arbitrary vectors from  $\mathfrak{C}$ . We consider the following model problem

$$(1) \quad \left[ \begin{array}{l} P_h\left(\frac{d}{dy}\right)Mu^h + b_h Au^h = 0 \quad \text{in } ]-h, h[, \\ a_h \frac{d}{dy} Mu^h = g \quad \text{for } y = h, \\ a_h \frac{d}{dy} Mu^h = f \quad \text{for } y = -h. \end{array} \right.$$

(Other boundary conditions, e.g., Dirichlet conditions, could just as well have been chosen; we could also consider the inhomogeneous problem. The above selection was simply made for convenience.)

Before we proceed any further, let us give a simple example.

*Example.* Let  $\omega$  be a domain in  $\mathbb{R}^n$  with a Lipschitz boundary. As  $\mathfrak{C}$  we take  $L^2(\omega)$ . Let  $A$  be the Friedrichs extension (cf. [13]) of the operator  $-\text{div } c(\mathbf{x})\text{grad}$  defined on a subspace of  $\dot{H}^1(\omega)$ . ( $\mathbf{x}$  denotes coordinates in  $\omega$ .)  $c$  is a function in  $L^\infty(\omega)$  such that  $\exists c_0 > 0$  with  $c_0 \leq c(\mathbf{x})$ .

If we take  $a = b$  and let  $M$  be the operator of multiplication by  $c(\mathbf{x})$ , the problem (1) becomes

$$\left[ \begin{array}{l} \text{div } d_h(\mathbf{x}, y)\text{grad } u^h = 0 \quad \text{in } \omega \times ]-h, h[, \\ d_h(\mathbf{x}, y) \frac{\partial}{\partial y} u^h = g(\mathbf{x}) \quad \text{for } y = h, \\ d_h(\mathbf{x}, y) \frac{\partial}{\partial y} u^h = f(\mathbf{x}) \quad \text{for } y = -h, \\ u^h = 0 \quad \text{on } \partial\omega \times [-h, h]. \end{array} \right.$$

Here  $d_h(\mathbf{x}, y) = b(y/h) \cdot c(\mathbf{x})$ , and  $\text{div}$  and  $\text{grad}$  are taken with respect to the  $n + 1$  coordinates  $(\mathbf{x}, y)$ .

The precise formulation of (1) is

$$(2) \quad \left[ \begin{array}{l} u^h \in H^1([-h, h]; \mathfrak{C}) \cap L^2([-h, h]; \mathfrak{D}(A^{1/2})), \\ \mathfrak{B}_h(u^h, v) = \langle g, v(h) \rangle - \langle f, v(-h) \rangle, \\ \forall v \in H^1([-h, h]; \mathfrak{C}) \cap L^2([-h, h]; \mathfrak{D}(A^{1/2})), \end{array} \right.$$

where  $\mathfrak{B}_h$  denotes the bilinear form

$$\begin{aligned} \mathfrak{B}_h(u, v) &= \int_{-h}^h a_h \left\langle M^{1/2} \frac{d}{dy} u, M^{1/2} \frac{d}{dy} v \right\rangle dy \\ &\quad + \int_{-h}^h b_h \langle A^{1/2} u, A^{1/2} v \rangle dy. \end{aligned}$$

If  $H^1([-h, h]; \mathfrak{C}) \cap L^2([-h, h]; \mathfrak{D}(A^{1/2}))$  is endowed with the natural norm  $\|\cdot\|$ :

$$\|u\| = \left[ \int_{-h}^h \left\| \frac{d}{dy} u(y) \right\|^2 dy + \int_{-h}^h \|A^{1/2} u(y)\|^2 dy \right]^{1/2},$$

then it is not difficult to prove

**PROPOSITION 2.1.** *There exist  $C_1, C_2$  (independent of  $h$ ) such that*

(i)  $|\mathfrak{B}_h(u, v)| \leq C_1 \|u\| \cdot \|v\|,$

(ii)  $\|u\|^2 \leq C_2 |\mathfrak{B}_h(u, u)|$

$\forall u, v \in H^1([-h, h]; \mathfrak{C}) \cap L^2([-h, h]; \mathfrak{D}(A^{1/2}))$ .

Also one has

**PROPOSITION 2.2.** *If  $x \in \mathfrak{C}$  and  $y_0 \in [-h, h]$ , then*

$$\Lambda: v \rightarrow \langle x, v(y_0) \rangle$$

*is a continuous linear functional on  $H^1([-h, h]; \mathfrak{C})$ .*

*Proof.* From the definition of  $H^1([-h, h]; \mathfrak{C})$  it follows that, if  $v(\cdot) \in H^1([-h, h]; \mathfrak{C})$  and  $\lambda$  is a continuous linear functional on  $\mathfrak{C}$ , then  $\lambda(v(\cdot)) \in H^1([-h, h])$  and

$$\begin{aligned} &\left[ \int_{-h}^h |\lambda(v(y))|^2 dy + \int_{-h}^h \left| \frac{d}{dy} \lambda(v(y)) \right|^2 dy \right]^{1/2} \\ &\leq \|\lambda\|' \left[ \int_{-h}^h \|v(y)\|^2 dy + \int_{-h}^h \left\| \frac{d}{dy} v(y) \right\|^2 dy \right]^{1/2}, \end{aligned}$$

( $\|\cdot\|'$  is the norm in  $\mathfrak{C}$ ).

Hence, with  $\lambda(\cdot) = \langle x, \cdot \rangle$ , we get  $\langle x, v(\cdot) \rangle \in H^1([-h, h])$  and

$$\begin{aligned} &\left[ \int_{-h}^h |\langle x, v(y) \rangle|^2 dy + \int_{-h}^h \left| \frac{d}{dy} \langle x, v(y) \rangle \right|^2 dy \right]^{1/2} \\ &\leq \|x\| \left[ \int_{-h}^h \|v(y)\|^2 dy + \int_{-h}^h \left\| \frac{d}{dy} v(y) \right\|^2 dy \right]^{1/2}. \end{aligned}$$

Using this last estimate together with the standard trace theorem, we finally get that  $\Lambda$  is a continuous linear functional on  $H^1([-h, h]; \mathfrak{C})$ .  $\square$

Propositions 2.1 and 2.2 immediately give

**PROPOSITION 2.3.** *The problem (2) has a unique solution.*

**3. The Direct Result.** We first define exactly what is meant by a dimensionally reduced solution to (2). Let  $\{\psi_j\}_{j=0}^\infty \subseteq H^1([-1, 1])$  be a given sequence of linearly independent functions.

*Definition.* The dimensionally reduced solution  $u_N^h$  of order  $N$  is the projection of  $u^h$  onto the space

$$V_N^h = \left\{ \sum_{j=0}^N \psi_j(y/h)x_j \mid x_j \in \mathcal{D}(A^{1/2}), j = 0, \dots, N \right\}.$$

The projection is with respect to the inner product  $\mathfrak{B}_h(u, v)$ .

The main result of this section is the next theorem, which suggests a way of choosing  $\{\psi_j\}_{j=0}^\infty$ . It also gives an estimate of  $\|u^h - u_N^h\|$  in terms of powers of  $h$ . As will be clear from the formulation of the theorem, this  $h$ -asymptotic estimate is very dependent on to what extent “boundary layers” are present. Since the dimensionally reduced solutions are linear combinations of functions of the type  $\psi(y/h)x$ , we do not include boundary corrector terms, as is often done with formal expansions to improve the asymptotic rate in  $h$ ; cf. [10]. In the case of “boundary layers” we instead increase  $N$ , which is also much more natural insofar as we are completely free to choose  $N$  but have absolutely no control over  $h$ .

We would furthermore like to underline that a particular boundary layer behavior is not relevant to the main scope of this investigation, which is the selection of  $\{\psi_j\}_{j=0}^\infty$ .

Let  $P$  denote the differential operator

$$b^{-1} \frac{d}{dy} a \frac{d}{dy}.$$

In the notation of the previous section  $P = -b^{-1}P_1(d/dy)$ .  $P$  is considered as an operator

$$L^2([-1, 1]) \supseteq \mathcal{D}(P) \rightarrow L^2([-1, 1]).$$

$\mathcal{N}(P^i)$  denotes the null space of the operator  $P^i$ ,  $0 < i$ .  $P^0 = I$  (identity). It is easy to see that  $\mathcal{N}(P^i) \subseteq H^1([-1, 1])$  for all  $i$ .

**THEOREM 3.1.** *There exists a sequence of linearly independent functions  $\{\psi_j\}_{j=0}^\infty$ , with*

$$\mathcal{N}(P^i) = \text{span}\{\psi_j\}_{j=0}^{2i-1}, \quad i > 1,$$

*that has the following property:*

*For any integer  $N \geq 0$  and for any given set of vectors  $f, g \in \mathcal{D}((AM^{-1})^N)$  there exists a constant  $C_N$  (independent of  $h$ ) such that*

$$\|u^h - u_{2N}^h\| < C_N h^{2N+1/2}.$$

*Remarks.* The sequence  $\{\psi_j\}_{j=0}^\infty$  depends only on the operator  $P$ . It is also clear that  $\{\psi_j\}_{j=0}^\infty$  is not uniquely determined by Theorem 3.1. Any other sequence  $\{\eta_j\}_{j=0}^\infty$ , with  $\text{span}\{\eta_j\}_{j=0}^i = \text{span}\{\psi_j\}_{j=0}^i \forall i$ , could have been used.

In order to prove Theorem 3.1, we need a couple of auxiliary results.

By changing variables to  $[-1, 1]$  and introducing  $\tilde{u}^h(y) = u^h(y \cdot h)$  for  $-1 < y < 1$ , we transform (2) into the following equation for  $\tilde{u}^h$ :

$$(3) \quad \begin{cases} \tilde{u}^h \in H^1([-1, 1]; \mathfrak{H}) \cap L^2([-1, 1]; \mathcal{D}(A^{1/2})), \\ \mathfrak{B}_h(\tilde{u}^h, v) = \langle g, v(1) \rangle - \langle f, v(-1) \rangle, \\ \forall v \in H^1([-1, 1]; \mathfrak{H}) \cap L^2([-1, 1]; \mathcal{D}(A^{1/2})). \end{cases}$$

Here the bilinear form  $\tilde{\mathfrak{B}}_h$  is given by

$$\begin{aligned}\tilde{\mathfrak{B}}_h(u, v) &= h^{-1} \int_{-1}^1 a \left\langle M^{1/2} \frac{du}{dy}, M^{1/2} \frac{dv}{dy} \right\rangle dy \\ &\quad + h \int_{-1}^1 b \langle A^{1/2} u, A^{1/2} v \rangle dy.\end{aligned}$$

Now let us define the sequence  $\{\psi_j^0\}_{j=0}^\infty \subseteq H^1([-1, 1])$  by the following equations: For any  $v \in H^1([-1, 1])$

$$(4^0) \quad \int_{-1}^1 a \frac{d\psi_0^0}{dy} \frac{dv}{dy} dy = 0,$$

$$(4^1) \quad \int_{-1}^1 a \frac{d\psi_1^0}{dy} \frac{dv}{dy} dy + \int_{-1}^1 b \psi_0^0 v dy = v(1),$$

and for  $j \geq 2$ :

$$(4^j) \quad \int_{-1}^1 a \frac{d\psi_j^0}{dy} \frac{dv}{dy} dy + \int_{-1}^1 b \psi_{j-1}^0 v dy = 0.$$

The sequence  $\{\psi_j^1\}_{j=0}^\infty \subseteq H^1([-1, 1])$  is defined by the same system of equations, the only difference being that in the right-hand side of  $(4^1)$   $v(1)$  is replaced by  $v(-1)$ .

LEMMA 3.1. *Let  $j_0$  denote an integer  $\geq 0$ . The equations  $(4^j)$ ,  $0 < j < j_0$ , determine the sets  $\{\psi_j^l\}_{j=0}^{j_0}$  uniquely up to a constant in  $\psi_j^l$ ,  $l = 0, 1$ .*

*Proof.* It is sufficient to prove the statement for  $l = 0$ .

The case  $j_0 = 0$  is obvious. We proceed by induction. Hence assume that the equations  $(4^j)$ ,  $0 < j < j_0 - 1$ , determine  $\{\psi_j^0\}_{j=0}^{j_0-1}$  uniquely modulo a constant in  $\psi_{j_0-1}^0$ .

Consider the equations  $(4^j)$ ,  $0 < j < j_0$ . According to the induction hypothesis,  $\{\psi_j^0\}_{j=0}^{j_0-1}$  is determined uniquely up to a constant in  $\psi_{j_0-1}^0$ . Choosing  $v = 1$  in Eq.  $(4^{j_0})$ , we derive the value of  $\int_{-1}^1 b \psi_{j_0-1}^0 dy$ , which means that  $\psi_{j_0-1}^0$  is completely determined. The equation  $(4^{j_0})$  is now nothing but a Neumann problem for  $-(d/dy)(ad/dy)$ , and since  $\int_{-1}^1 b \psi_{j_0-1}^0 v dy$  is equal to the right-hand side of Eq.  $(4^{j_0})$  for any constant  $v$ , this has a solution that is unique modulo a constant. This proves that  $\psi_{j_0}^0$  is determined uniquely up to a constant.  $\square$

Because of the way the two sequences  $\{\psi_j^0\}_{j=0}^\infty$  and  $\{\psi_j^1\}_{j=0}^\infty$  are constructed, we also have

LEMMA 3.2. *For any  $i \geq 0$  the following proper inclusions hold*

$$\mathfrak{U}(P^i) \subset \text{span}\{\psi_j^0, \psi_j^1\}_{j=0}^i \subset \mathfrak{U}(P^{i+1}).$$

*Proof.* The lemma is clearly true if, for any  $i \geq 0$ , we can prove the more detailed statement:

$$\mathfrak{U}(P^i) \subset \text{span}\{\psi_j^0, \psi_j^1\}_{j=0}^i \subset \mathfrak{U}(P^{i+1}) \quad \text{properly,}$$

and, if  $\psi$  is defined by  $\psi = \psi_{i+1}^0 - \psi_{i+1}^1$ , then

$$\text{span}\{\psi_j^0, \psi_j^1\}_{j=0}^i \oplus \text{span}\{\psi\} = \mathfrak{U}(P^{i+1}).$$

The validity of the above statement is easily checked for  $i = 0$ , and we proceed by induction. That is, we assume the statement to be true for  $i = k > 0$ .

From the way  $\psi_{k+1}^0$  and  $\psi_{k+1}^1$  are constructed, it follows that

$$\psi_{k+1}^0, \psi_{k+1}^1 \in \mathcal{U}(P^{k+2}) \setminus \mathcal{U}(P^{k+1}),$$

and, because of the induction hypothesis, we then get

$$\mathcal{U}(P^{k+1}) \subset \text{span}\{\psi_j^0, \psi_j^1\}_{j=0}^{k+1} \text{ properly.}$$

Now, adding the two vectors  $\psi_{k+1}^0$  and  $\psi_{k+1}^1$ , we cannot increase the dimension by more than two. The fact that the codimension of  $\mathcal{U}(P^{k+1})$  in  $\mathcal{U}(P^{k+2})$  is 2, together with the inclusion

$$\text{span}\{\psi_j^0, \psi_j^1\}_{j=0}^k \subset \mathcal{U}(P^{k+1}),$$

hence shows that

$$\text{span}\{\psi_j^0, \psi_j^1\}_{j=0}^{k+1} \subset \mathcal{U}(P^{k+2}) \text{ properly.}$$

That means we have proven the first part of the extended statement for  $i = k + 1$ .

Concerning the second part we consider the linear combination  $\psi_{k+2}^0 - \psi_{k+2}^1$ .

From the construction we know that  $P(\psi_{k+2}^0 - \psi_{k+2}^1) = \psi_{k+1}^0 - \psi_{k+1}^1$ .

Together with the induction hypothesis, this tells that

$$\psi_{k+2}^0 - \psi_{k+2}^1 \in \mathcal{U}(P^{k+2}).$$

On the other hand,  $\psi_{k+2}^0 - \psi_{k+2}^1$  cannot be an element of  $\text{span}\{\psi_j^0, \psi_j^1\}_{j=0}^{k+1}$ . If so, we would have, by application of the operator  $P$ , that

$$\psi_{k+1}^0 - \psi_{k+1}^1 \in \text{span}\{\psi_j^0, \psi_j^1\}_{j=0}^k,$$

which contradicts the induction hypothesis.

This finally proves that  $\psi_{k+2}^0 - \psi_{k+2}^1$ , together with  $\text{span}\{\psi_j^0, \psi_j^1\}_{j=0}^{k+1}$ , span all of  $\mathcal{U}(P^{k+2})$ .

Thus, the second part has been established for  $i = k + 1$ .  $\square$

We are now ready for the

*Proof of Theorem 3.1.* Choose  $\{\psi_j\}_{j=0}^\infty$  such that

$$\text{span}\{\psi_j\}_{j=0}^{2i-1} = \mathcal{U}(P^i) \quad \forall i > 1,$$

and

$$\text{span}\{\psi_j\}_{j=0}^{2i} = \text{span}\{\psi_j^0, \psi_j^1\}_{j=0}^i \quad \forall i > 0.$$

This is possible because of Lemma 3.2. The  $\psi_j$ 's chosen this way obviously have the first property stated in Theorem 3.1.

For any  $N > 0$  and any pair  $f, g \in \mathfrak{D}((AM^{-1})^N)$ , let

$$S_N^h = \sum_{j=0}^N h^{-1+2j} (\psi_j^0(y/h)M^{-1}(AM^{-1})^{j-1}g - \psi_j^1(y/h)M^{-1}(AM^{-1})^{j-1}f).$$

It is clear that  $S_N^h \in V_{2N}^h$ . Because of Proposition 2.1 it follows that there exists a constant  $C$  (independent of  $N$  and  $h$ ) such that

$$\|u^h - u_{2N}^h\| < C \inf_{v \in V_{2N}^h} \|u^h - v\| < C \|u^h - S_N^h\|.$$

We now proceed to estimate  $\|u^h - S_N^h\|$ . This is very straightforward and similar to what can be found in, e.g., [10]. For completeness, we give the details.

By a change of variables and introduction of the bilinear form  $\tilde{\mathfrak{B}}_h$ , it follows that

$$\|u^h - S_N^h\| \leq C [\tilde{\mathfrak{B}}_h(\tilde{u}^h - \tilde{S}_N^h, \tilde{u}^h - \tilde{S}_N^h)]^{1/2},$$

with

$$\tilde{S}_N^h = \sum_{j=0}^N h^{-1+2j} (\psi_j^0(y) M^{-1} (AM^{-1})^{j-1} g - \psi_j^1(y) M^{-1} (AM^{-1})^{j-1} f)$$

(and  $C$  independent of  $N$  and  $h$ ).

Let us now consider  $\tilde{\mathfrak{B}}_h(\tilde{u}^h - \tilde{S}_N^h, v)$  for an arbitrary  $v \in H^1([-1, 1]; \mathfrak{C}) \cap L^2([-1, 1]; \mathfrak{D}(A^{1/2}))$ . We get that

$$\tilde{\mathfrak{B}}_h(\tilde{u}^h - \tilde{S}_N^h, v) = \langle g, v(1) \rangle - \langle f, v(-1) \rangle - \tilde{\mathfrak{B}}_h(\tilde{S}_N^h, v).$$

Concerning the last term, we have

$$\begin{aligned} \tilde{\mathfrak{B}}_h(\tilde{S}_N^h, v) &= \sum_{j=0}^N h^{-1+2j} \tilde{\mathfrak{B}}_h(\psi_j^0(y) M^{-1} (AM^{-1})^{j-1} g, v) \\ &\quad - \sum_{j=0}^N h^{-1+2j} \tilde{\mathfrak{B}}_h(\psi_j^1(y) M^{-1} (AM^{-1})^{j-1} f, v). \end{aligned}$$

Now, because of the properties of the sequence  $\{\psi_j^0\}_{j=0}^\infty$ ,

$$\begin{aligned} &\sum_{j=0}^N h^{-1+2j} \tilde{\mathfrak{B}}_h(\psi_j^0 M^{-1} (AM^{-1})^{j-1} g, v) \\ &= \sum_{j=0}^N h^{-2+2j} \int_{-1}^1 a \frac{d\psi_j^0}{dy} \frac{d}{dy} \langle (AM^{-1})^{j-1} g, v \rangle dy \\ &\quad + \sum_{j=0}^N h^{2j} \int_{-1}^1 b \psi_j^0 \langle (AM^{-1})^j g, v \rangle dy \\ &= \langle g, v(1) \rangle - h^{2N} \int_{-1}^1 a \frac{d\psi_{N+1}^0}{dy} \frac{d}{dy} \langle (AM^{-1})^N g, v \rangle dy. \end{aligned}$$

Similarly,

$$\begin{aligned} &\sum_{j=0}^N h^{-1+2j} \tilde{\mathfrak{B}}_h(\psi_j^1 M^{-1} (AM^{-1})^{j-1} f, v) \\ &= \langle f, v(-1) \rangle - h^{2N} \int_{-1}^1 a \frac{d\psi_{N+1}^1}{dy} \frac{d}{dy} \langle (AM^{-1})^N f, v \rangle dy, \end{aligned}$$

so that altogether

$$\tilde{\mathfrak{B}}_h(\tilde{u}^h - \tilde{S}_N^h, v) = h^{2N} \int_{-1}^1 a \left\langle \frac{dr_N}{dy}, \frac{dv}{dy} \right\rangle dy,$$

with  $r_N$  given by

$$r_N = \psi_{N+1}^0 (AM^{-1})^N g - \psi_{N+1}^1 (AM^{-1})^N f.$$

Using Schwarz's inequality, we see that

$$\left| h^{2N} \int_{-1}^1 a \left\langle \frac{dr_N}{dy}, \frac{dv}{dy} \right\rangle dy \right|$$



is bounded by  $C_N h^{2N+1/2} [\tilde{\mathfrak{B}}^h(v, v)]^{1/2}$ . As a consequence of this, it follows that

$$[\tilde{\mathfrak{B}}(\tilde{u}^h - \tilde{S}_N^h, \tilde{u}^h - \tilde{S}_N^h)]^{1/2} < C_N h^{2N+1/2},$$

and hence  $\|u^h - S_N^h\| \leq C_N h^{2N+1/2}$ .  $\square$

From Theorem 3.1 we immediately get

**COROLLARY 3.1.** *Let  $\{\psi_j\}_{j=0}^\infty$  be a sequence with properties as in Theorem 3.1. Let  $N$  and  $K$  be two nonnegative integers and  $f, g$  a set of elements of  $\mathfrak{D}((AM^{-1})^K)$ . Then there exists a constant  $C$  (dependent on  $K$  and  $N$  but independent of  $h$ ) such that*

$$\|u^h - u_{2N}^h\| < Ch^{2-\min\{K, N\}+1/2}.$$

**4. The Inverse Result.** In Theorem 3.1 we examined a particular choice of the functions  $\psi_j$ . The approximation error was of order  $h^{2N+1/2}$  using the  $2N + 1$  functions  $\{\psi_j\}_{j=0}^{2N}$ , provided the vectors  $f$  and  $g$  were sufficiently regular.

The goal of this section is to prove that the previous choice of functions was by no means arbitrary. That sequence, or any other sequence  $\{\eta_j\}_{j=0}^\infty$  with  $\text{span}\{\eta_j\}_{j=0}^{2i} = \text{span}\{\psi_j\}_{j=0}^{2i}$  for every  $i$ , is the only one that gives this order of approximation.

We formulate this as

**THEOREM 4.1.** *Let  $N$  and  $K$  be two nonnegative integers and  $f$  and  $g$  two linearly independent elements of  $\mathfrak{H}$ . Let  $\{\psi_j\}_{j=0}^\infty$  be the sequence introduced in Theorem 3.1 and  $u^h$  the solution to (2) of Section 2.*

*If  $\{\phi_j\}_{j=0}^K$  is a set of elements of  $H^1([-1, 1])$  with the property that*

$$\inf_{v \in W_K^h} \|u^h - v\| = o(h^{\max\{2N-3/2, -1/2\}}),$$

where  $W_K^h$  denotes the set

$$\left\{ \sum_{j=0}^K \phi_j(y/h)x_j \mid x_j \in \mathfrak{D}(A^{1/2}), 0 < j < K \right\},$$

then

$$\text{span}\{\psi_j\}_{j=0}^{2N} \subseteq \text{span}\{\phi_j\}_{j=0}^K.$$

Theorem 4.1 is actually a little stronger than just an inverse of Theorem 3.1. Let  $N$  be  $> 1$ . Theorem 3.1 then says that with the  $2N + 1$  functions  $\{\psi_j\}_{j=0}^{2N}$  we can obtain an error of order  $h^{2N+1/2}$ . But Theorem 4.1 tells us that, even if we are satisfied with an error of order  $o(h^{2N-3/2})$ , we still have to use all the functions  $\{\psi_j\}_{j=0}^{2N}$ .

*Proof.* We can, without loss of generality, assume that  $f, g \in \mathfrak{D}((AM^{-1})^N)$ . Otherwise we replace  $u^h, f$ , and  $g$  by  $u^{h*} = (A^{-1}M)^N u^h, f^* = (MA^{-1})^N f$  and  $g^* = (MA^{-1})^N g$ , which obviously satisfy the assumptions of the theorem.

Define  $S_N^h, \tilde{S}_N^h$  as in the proof of Theorem 3.1, i.e.,

$$S_N^h = \sum_{j=0}^N h^{-1+2j} (\psi_j^0(y/h)M^{-1}(AM^{-1})^{j-1}g - \psi_j^1(y/h)M^{-1}(AM^{-1})^{j-1}f)$$

and  $\tilde{S}_N^h = S_N^h(y \cdot h)$ .

Then, because of Theorem 3.1 and the assumption of this theorem, we see that  $\exists v^h \in W_K^h$  such that

$$\|S_N^h - v^h\| = o(h^{\max\{2N-3/2, -1/2\}}).$$

By a change of variables, this yields

- (i)  $\int_{-1}^1 \|d(\tilde{S}_N^h - \tilde{v}^h)/dy\|^2 dy = o(h^{\max(4N-2, 0)})$ ,
- (ii)  $\int_{-1}^1 \|A^{1/2} \tilde{S}_N^h - \tilde{v}^h\|^2 dy = o(h^{\max(4N-4, -2)})$ .

The function  $\tilde{v}^h$  has the form

$$\sum_{j=0}^K \phi_j(y) v_j^h, \quad \text{with } v_j^h \in \mathcal{D}(A^{1/2}) \quad \text{for } 0 < j < K.$$

Now, multiplying (ii) through by  $h^2$ , we get, because of the form of  $\tilde{S}_N^h$  and  $\tilde{v}^h$ , that

$$\int_{-1}^1 \left\| \psi_0^0(y) A^{-1/2} g - \psi_0^1(y) A^{-1/2} f - h \sum_{j=0}^K \phi_j(y) A^{1/2} v_j^h \right\|^2 dy \rightarrow 0$$

for  $h \rightarrow 0$ . Since for a fixed  $K$  the set  $\{\sum_{j=0}^K \phi_j(y) x_j | x_j \in \mathcal{IC}\}$  is closed in  $L^2([-1, 1]; \mathcal{IC})$ , we get

$$\psi_0^0(y) A^{-1/2} g - \psi_0^1(y) A^{-1/2} f = \sum_{j=0}^K \phi_j(y) x_j.$$

The fact that  $f$  and  $g$  are linearly independent implies that so are  $A^{-1/2} f$  and  $A^{-1/2} g$ . Hence

$$\text{span}\{\psi_0^0, \psi_0^1\} \subseteq \text{span}\{\phi_j\}_{j=0}^K.$$

This proves Theorem 4.1 for  $N = 0$ . If  $N > 1$ , we are not yet finished. In this case we proceed by induction, i.e., we assume it has been proven that

$$\text{span}\{\psi_j^0, \psi_j^1\}_{j=0}^m \subseteq \text{span}\{\phi_j\}_{j=0}^K$$

for some  $m: 0 \leq m \leq N - 1$ .

Rearranging (i) and dividing through by  $h^{4m+2}$ , we get

$$\begin{aligned} \int_{-1}^1 \left\| \frac{d\psi_{m+1}^0}{dy} M^{-1}(AM^{-1})^m g - \frac{d\psi_{m+1}^1}{dy} M^{-1}(AM^{-1})^m f \right. \\ \left. + \sum_{j=0}^m h^{2U-m-1} \left( \frac{d\psi_j^0}{dy} M^{-1}(AM^{-1})^{j-1} g - \frac{d\psi_j^1}{dy} M^{-1}(AM^{-1})^{j-1} f \right) \right. \\ \left. - h^{-2m-1} \sum_{j=0}^K \frac{d\phi_j}{dy} v_j^h \right\|^2 dy \\ \rightarrow 0 \quad \text{for } h \rightarrow 0. \end{aligned}$$

(Here we use that  $4m + 2 < 4N - 2$ .)

Because of the induction hypothesis, we know that

$$\sum_{j=0}^m h^{2U-m-1} \left( \frac{d\psi_j^0}{dy} M^{-1}(AM^{-1})^{j-1} g - \frac{d\psi_j^1}{dy} M^{-1}(AM^{-1})^{j-1} f \right) = \sum_{j=0}^K \frac{d\phi_j}{dy} x_j^h,$$

and now, using that for a fixed  $K$  the set  $\{\sum_{j=0}^K (d\phi_j/dy) x_j | x_j \in \mathcal{IC}\}$  is closed in  $L^2([-1, 1]; \mathcal{IC})$ , we conclude that

$$\frac{d\psi_{m+1}^0}{dy} M^{-1}(AM^{-1})^m g - \frac{d\psi_{m+1}^1}{dy} M^{-1}(AM^{-1})^m f = \sum_{j=0}^K \frac{d\phi_j}{dy} x_j.$$

The fact that  $f$  and  $g$  are linearly independent implies that so are  $M^{-1}(AM^{-1})^m g$  and  $M^{-1}(AM^{-1})^m f$ . Hence

$$\frac{d\psi_{m+1}^0}{dy}, \frac{d\psi_{m+1}^1}{dy} \in \text{span} \left\{ \frac{d\phi_j}{dy} \right\}_{j=0}^K.$$

From the way  $\psi_0^0$  and  $\psi_0^1$  are constructed, it is easily seen that

$$\text{span} \{ \psi_0^0, \psi_0^1 \} = \{ \text{constant functions} \},$$

and this together with the induction hypothesis allow us to conclude that

$$\text{span} \{ \psi_j^0, \psi_j^1 \}_{j=0}^{m+1} \subseteq \text{span} \{ \phi_j \}_{j=0}^K.$$

The induction proof is now finished, and finally we get, because of the definition of  $\{ \psi_j \}_{j=0}^{2N}$ ,

$$\text{span} \{ \psi_j \}_{j=0}^{2N} = \text{span} \{ \psi_j^0, \psi_j^1 \}_{j=0}^N \subseteq \text{span} \{ \phi_j \}_{j=0}^K. \quad \square$$

By a slight variation of the preceding arguments, we could prove the following version of Theorem 4.1, for the case where  $f$  and  $g$  are not linearly independent, e.g.,  $f = \alpha \cdot g$ .

**THEOREM 4.2.** *Let  $N$  and  $K$  be two nonnegative integers. Let  $f = \alpha \cdot g$ , with  $g \in \mathcal{C} \setminus \{0\}$ , and let  $\{ \psi_j^0 \}_{j=0}^\infty$  and  $\{ \psi_j^1 \}_{j=0}^\infty$  denote the same two sequences as in Lemma 3.1.*

*If  $\{ \phi_j \}_{j=0}^K$  is a set of elements of  $H^1([-1, 1])$  with the property that*

$$\inf_{v \in W_K^h} \| \| u^h - v \| \| = o(h^{2N-1/2}),$$

*then*

$$\text{span} \{ \psi_j^0 - \alpha \psi_j^1 \}_{j=0}^N \subseteq \text{span} \{ \phi_j \}_{j=0}^K.$$

**5. Numerical Examples.** Consider the problem

$$\left[ \begin{array}{ll} \text{div}(a \text{ grad } u^h) = 0 & \text{in } ]0, 1[ \times ]-h, h[, \\ a \frac{\partial}{\partial y} u^h = g(x) & \text{for } y = h, \\ a \frac{\partial}{\partial y} u^h = -g(x) & \text{for } y = -h, \\ u^h = 0 & \text{for } x = 0 \text{ and } x = 1, \end{array} \right.$$

with

$$a(y) = \begin{cases} a_+ & \text{for } y > 0, \\ a_- & \text{for } y < 0. \end{cases}$$

( $a_+$  and  $a_-$  are two positive constants.) This problem clearly falls within the framework of our model problem. Simply choose

$$b(y) = a(y) = \begin{cases} a_+ & \text{for } y > 0, \\ a_- & \text{for } y < 0, \end{cases}$$

$$c(x) = 1, \text{ and}$$

$$A = - \left( \frac{\partial}{\partial x} \right)^2 \text{ with } \mathcal{D}(A) = H^2([0, 1]) \cap \dot{H}^1([0, 1]) \text{ and } \mathcal{H} = L^2([0, 1]).$$

The operator  $P$  in this case is given by

$$P = a^{-1} \frac{d}{dy} a \frac{d}{dy}.$$

Define  $\{\phi_j\}_{j=0}^{\infty} \subseteq H^1([-1, 1])$  as follows

$$\begin{aligned} \phi_0 &= 1, & \phi_1(y) &= a^{-1}y, \\ \phi_{2j}(y) &= \int_{-1}^y l_{2j-1}(t) dt & \text{for } j \geq 1, \end{aligned}$$

and

$$\phi_{2j+1}(y) = a^{-1} \int_{-1}^y l_{2j}(t) dt \quad \text{for } j \geq 1.$$

Here  $l_k$  denotes the Legendre polynomial of degree  $k$ .

It is not difficult to see that with this definition

$$\text{span}\{\phi_j\}_{j=0}^i = \text{span}\{\psi_j\}_{j=0}^i \quad \forall i \geq 0,$$

where  $\{\psi_j\}_{j=0}^{\infty}$  is the sequence introduced in Theorem 3.1.

As before  $u_N^h$  denotes the dimensionally reduced solution of order  $N$ .  $u_N^h$  has the form

$$u_N^h = \sum_{j=0}^N \phi_j(y/h) u_j(x),$$

with  $u_j \in \dot{H}^1([0, 1])$ . The vector  $U = (u_j)_{j=0}^N$  is the solution to a two-point elliptic boundary value problem

$$-h\mathbf{K} \left( \frac{d}{dx} \right)^2 U + 1/h\mathbf{L}U = F, \quad U(0) = U(1) = 0.$$

The matrix  $\mathbf{L}$  is diagonal, and the matrix  $\mathbf{K}$  has a band structure. Both  $\mathbf{L}$  and  $\mathbf{K}$  are independent of  $h$ . We solve this problem numerically by expanding  $U$  in its Fourier series, only maintaining a finite number of terms. Since we are interested in studying the error introduced by the dimensional reduction, we maintain a very high number of terms. The graphs shown here were computed using 400 Fourier coefficients. This ensures that the error introduced by discretization can be neglected compared to the error introduced by dimensional reduction.

Let us start with the case  $a_+ = a_- = 1$  and  $g(x) = \pi/4$ . Figures 1, 2, and 3 show the energy error as a function of  $h$  by dimensional reduction of order 0, 2, and 4, respectively. (Note that the functions  $\{\phi_j\}_{j=0}^{\infty}$  form a basis for the polynomials in this case.) Using interpolation by the  $K$ -method (cf. [3]) we know that  $g(x) \in (\mathcal{C}, \mathcal{D}(A))_{1/4, \infty}$ . An application of Corollary 3.1 hence gives the following conclusions. (The energy error is here not the norm but the difference in energies.)

- (i) The energy error is of order  $h$  with dimensional reduction based on  $\phi_0$ .
- (ii) The energy error is of order  $h^2$  with dimensional reduction based on  $\phi_0, \phi_1$  and  $\phi_2$ .
- (iii) The energy error is of order  $h^2$  with dimensional reduction based on  $\{\phi_j, 0 < j < 4\}$ .

Figures 1, 2, and 3 illustrate the sharpness of the theoretical results. Comparing Figures 2 and 3 we see that

The dependence of the estimate in Theorem 3.1 (or Corollary 3.1) on the regularity of  $f$  and  $g$  is essential. If  $f$  and  $g$  are not sufficiently smooth, higher-order dimensionally reduced models will not improve the asymptotic order of approximation as  $h \rightarrow 0$ . (But they certainly give better approximate results due to the convergence as  $N \rightarrow \infty$ ; cf. [15].)

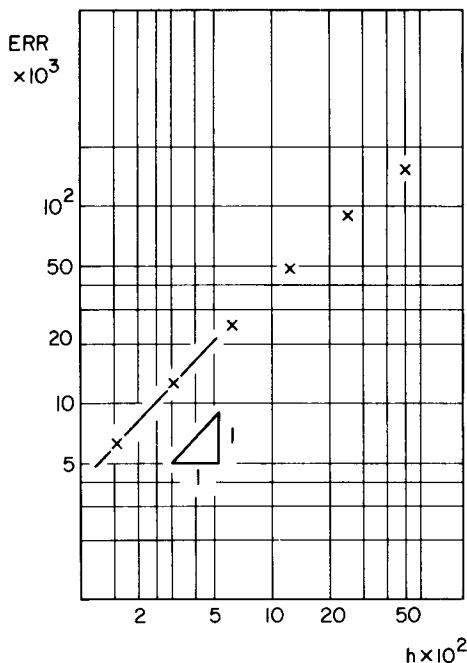


FIGURE 1

Energy error  $\times 10^3$  as a function of  $h$ , using polynomials of degree = 0

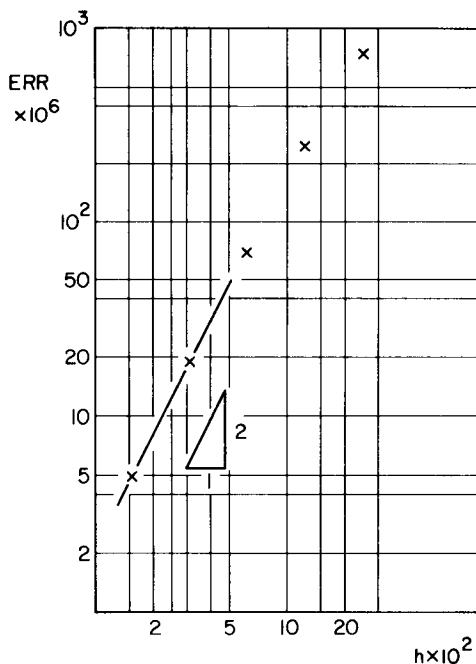


FIGURE 2

Energy error  $\times 10^6$  as a function of  $h$ , using polynomials of degree  $< 2$

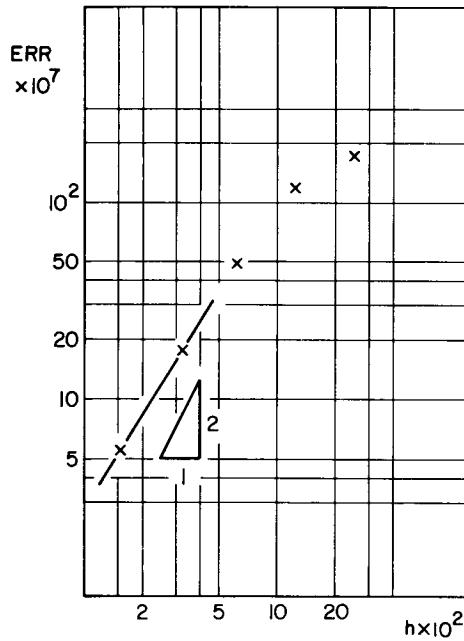


FIGURE 3

Energy error  $\times 10^7$  as a function of  $h$ , using polynomials of degree  $< 4$

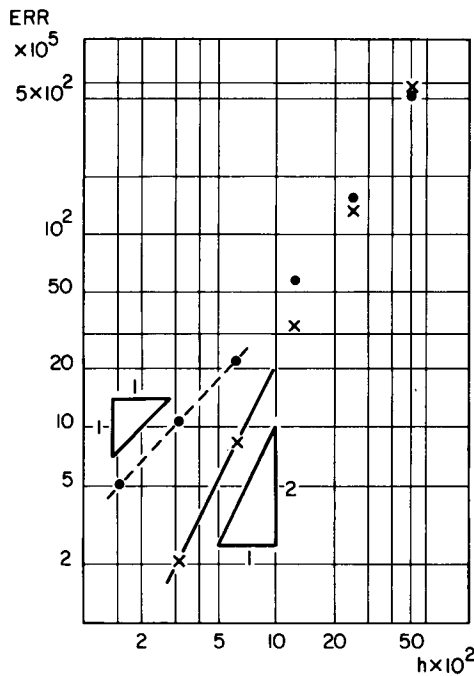


FIGURE 4

Energy error  $\times 10^5$  as a function of  $h$ ,

— • — • — using polynomials of degree  $< 2$ ,

— x — x — using the "special" functions  $\phi_0, \phi_1$  and  $\phi_2$ .

Now consider the case where  $a_+ \neq a_-$ . In our computations  $a_+ = 1$ ,  $a_- = 2$ , and again  $g(x) = \pi/4$ . Figure 4 compares two different dimensionally reduced solutions. For one of the dimensionally reduced solutions the polynomials of degree  $\leq 2$  have been chosen as basis functions in the  $y$ -direction. For the other dimensionally reduced solution the "special" functions  $\phi_0$ ,  $\phi_1$ , and  $\phi_2$ , introduced earlier in this section, have been used. (Note that the "special" functions are piecewise polynomials in this case.)

Applying Corollary 3.1, we get that

The energy error will be of order  $h^2$  with dimensional reduction based on the functions  $\phi_0$ ,  $\phi_1$  and  $\phi_2$ .

Since the "special" function  $\phi_0$  is the constant  $= 1$ , which of course is a polynomial of degree  $\leq 2$ , we also expect that

The energy error will be of order  $h$  with dimensional reduction based on the polynomials of degree  $\leq 2$ .

From Figure 4 it is again evident that there is a very good agreement between the theory and the computational results. Specifically it is seen, by comparing Figures 2 and 4, that

If the dimensional reduction is based on the "special" functions  $\{\phi_j\}_{j=0}^\infty$ , then the asymptotic behavior of the energy error is independent of the regularity of the solution  $u^h$  across the line  $y = 0$ .

A feature very similar to this is well known for optimally constructed finite element meshes.

Institute for Physical Science and Technology  
University of Maryland  
College Park, Maryland 20742

1. I. BABUŠKA & W. C. RHEINBOLDT, "Mathematical problems of computational decisions for the finite element method," in *Mathematical Aspects of Finite Element Methods* (I. Galligani and E. Magenes, Eds.), Lecture Notes in Math., vol. 606, Springer-Verlag, Berlin and New York, 1977, pp. 1–26.
2. V. BARBU, *Nonlinear Semigroups and Differential Equations in Banach Spaces*, Noordhoff, Groningen, 1976.
3. J. BERGH & J. LÖFSTRÖM, *Interpolation Spaces*, Springer-Verlag, Berlin and New York, 1976.
4. V. E. CHEPIGA, "On constructing a theory of multilayer anisotropic shells with prescribed arbitrary accuracy of order  $h^N$ ," *Mech. Solids*, v. 12, 1977, pp. 113–120.
5. P. G. CIARLET & P. DESTUYNDER, "A justification of the two-dimensional linear plate model," *J. Mécanique*, v. 18, 1979, pp. 315–344.
6. N. DUNFORD & J. SCHWARTZ, *Linear Operators*, Part I, Interscience, New York, 1958.
7. K. O. FRIEDRICHS & R. F. DRESSLER, "A boundary layer theory for elastic plates," *Comm. Pure. Appl. Math.*, v. 14, 1961, pp. 1–33.
8. A. L. GOL'DENWEIZER, "Derivation of an approximate theory of bending of a plate by the method of asymptotic integration of the equations of the theory of elasticity," *J. Appl. Math. Mech.*, v. 26, 1962, pp. 1000–1025.
9. L. V. KANTOROVICH & V. I. KRYLOV, *Approximate Methods of Higher Analysis*, Noordhoff, Groningen, 1958.
10. J. L. LIONS, *Perturbations Singulières dans les Problèmes aux Limites et en Contrôle Optimal*, Lecture Notes in Math., vol. 323, Springer-Verlag, Berlin and New York, 1973.

11. P. M. NAGHDI, *Handbuch der Physik*, Band V a/2, Springer-Verlag, Berlin and New York, 1972, pp. 425–640.
12. V. V. PONJATOVSKIĬ, “Theory for plates of medium thickness,” *J. Appl. Math. Mech.*, v. 26, 1962, pp. 335–341.
13. F. RIESZ & B. SZ.-NAGY, *Functional Analysis*, Ungar, New York, 1955.
14. M. VOGELIUS, *A Dimensional Reduction Approach to the Solution of Partial Differential Equations*, Ph. D. Thesis, University of Maryland, Dec. 1979.
15. M. VOGELIUS & I. BABUŠKA, “On a dimensional reduction method. II. Some approximation-theoretic results,” *Math. Comp.*, v. 37, 1981, pp. 47–68.
16. M. VOGELIUS & I. BABUŠKA, “On a dimensional reduction method. III. A posteriori error estimation and an adaptive approach,” *Math. Comp.* (To appear.)
17. I. E. ZINO & E. A. TROPP, *Asymptotic Methods in the Problems of Heat Transfer and Thermoelasticity*, Univ. of Leningrad Publ., Leningrad, 1978. (Russian)