NISEE/COMPUTER APPLICATIONS
DAVIS HALL
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720
(415) 642-5113

Report no.
UC SESM 75 – 9

STRUCTURAL ENGINEERING AND STRUCTURAL MECHANICS

A REDUCTION SCHEME FOR PROBLEMS OF STRUCTURAL DYNAMICS

by

THOMAS J.R. HUGHES
HANS M. HILBER
ROBERT L. TAYLOR

SEPTEMBER 1975

DEPARTMENT OF CIVIL ENGINEERING
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA

A REDUCTION SCHEME FOR PROBLEMS OF STRUCTURAL DYNAMICS

by

Thomas J. R. Hughes
Hans M. Hilber
Robert L. Taylor

Division of Structural Engineering and
Structural Mechanics
Department of Civil Engineering
University of California
Berkeley, California 94720

ABSTRACT

A method for reducing the size of finite element systems in dynamics is presented. The technique is based upon a variational theorem in which it is admissible to describe the inertial properties of structures by way of independent displacement, velocity and momentum fields. This theorem allows us to construct reduced systems for problems in structural mechanics which retain the full rate of convergence of systems employing "consistent" mass matrices. In particular, we are able to make precise the engineering intuition regarding the "inefficiency" of rotatory degrees of freedom in dynamics, i.e., for the common beam, plate and shell elements, rotatory degrees of freedom may be entirely eliminated while retaining full rate of convergence. An error analysis of the scheme and numerical examples are presented.

TABLE OF CONTENTS

ACKNOWLEDGEMENT

This research was partially supported by funds from Faculty Research Grant 615, University of California, Berkeley.

1. Introduction.

In this paper we investigate a finite element method for dynamics problems in which there exists considerable flexibility in the definition of the mass matrix. The technique emanates from a variational formulation in which the displacement, velocity and momentum fields may be taken to be independent. In the next section we present some variational theorems of this sort for application to linear elastodynamics. Based upon specific forms of these results, in which the momentum field is eliminated, we set up the spatially discretized finite element equations in Section 3. If the number of velocity degrees of freedom is less than the number of displacement degrees of freedom, then the finite element equations may be combined to form reduced systems involving fewer unknowns. An essentially identical approach can be used in reducing the size of problems of heat transfer. In Section 4 we include an error analysis of the scheme for eigenvalue problems as well as a wide class of hyperbolic and parabolic problems. The main result is that the rate of convergence of the consistent mass model is maintained as long as $\hat{k} \geq k - m$, where $2m$ is the order of the spatial differential operator and k ($\hat{k}$, resp.) is the degree of the complete polynomial contained in the displacement (velocity, resp.) interpolation assumption. In addition, the boundedness property of frequencies, characteristic of the consistent mass model, is maintained by the reduced system.

These developments imply, in particular, that reduced systems may be constructed, for the common beam, plate and shell elements, in which rotational degrees of freedom are eliminated, and for which the rate of convergence of the consistent mass matrix is retained. Some examples along these lines are discussed in Section 5. Numerical corroboration of these results is presented in Section 6.

2. Variational Theorems for Linear Elastodynamics.

Variational theorems can be constructed for problems of linear elastodynamics (and, in fact, other and more general theories) in which displacement, velocity and momentum fields are taken to be independent. Ideas along these lines go back, at least, to the work of Livens (see Section 26.2 of Pars [12] and Appendix I of Lanczos [10])pertaining to the dynamics of mass points and rigid bodies. In the present work we analyze further the technique suggested in [9]. However, we have recently realized that our approach may be viewed under the general heading of dual complementary variational principles as extensively developed by Oden and Reddy [11]. We do not exploit the techniques or results of this subject in the present work.

We consider here the standard initial-boundary-value problem of linear elastodynamics. Namely let u_i represent the displacements, v_i the velocities, p_i the momenta, $c_{ijk\ell}$ the elastic stiffness coefficients, f_i the body force and ρ the mass density. Define a functional

$$F = \int_0^T \{\int_\Omega (-\frac{1}{2}\rho v_i v_i - p_i(\dot{u}_i - v_i) + \frac{1}{2} c_{ijk\ell}\, u_{i,j} u_{k,\ell} - f_i u_i)\, dx - \int_{\partial\Omega_T} \bar{T}_i u_i\, da\}\, dt, \tag{1}$$

where Ω is a bounded region in R^3 with nice boundary $\partial\Omega$, $\partial\Omega_T$ is that part of $\partial\Omega$ on which there exists prescribed tractions $\bar{T}_i$, $\partial\Omega_u = \partial\Omega \sim \partial\Omega_T$ is the complement of $\partial\Omega_T$ in $\partial\Omega$, upon which displacements u_i are prescribed, dx is the volume element for Ω, da is the area element for $\partial\Omega$, t denotes time, $T > 0$, a superposed dot indicates time differentiation (i.e., $\dot{u}_i = \partial u_i/\partial t$) a comma indicates differentiation with respect to the coordinates (i.e., $u_{i,j} = \partial u_i/\partial x_j$) and, finally, the summation convention is employed for repeated indices. In (1) u_i, v_i and p_i are considered to be independent. Assume $u_i = \bar{u}_i$ on $\partial\Omega_u$. Then the first variation of F is zero, i.e.,

$$0 = \int_0^T \{ \int_\Omega (- (\rho v_i - p_i)\beta_i - (\dot{u}_i - v_i)\gamma_i + (\dot{p}_i - (c_{ijk\ell}\, u_{k,\ell}),_j - f_i)\,\alpha_i) dx + \int_{\partial\Omega_T} (n_j\, c_{ijk\ell}\, u_{k,\ell} - \bar{T}_i)\alpha_i da \}\, dt, \tag{2}$$

where n_j denotes the unit outward normal vector with respect to $\partial\Omega$, for all α_i, β_i and γ_i such that $\alpha_i = 0$ on $\partial\Omega_u$ and at $t=0$ and $t=T$, if and only if u_i, v_i and p_i satisfy the equations of linear elastodynamics:

$$\left.\begin{aligned} p_i &= \rho v_i \quad , \\ v_i &= \dot{u}_i \quad , \\ \dot{p}_i &= (c_{ijk\ell}\, u_{k,\ell}),_j + f_i \quad , \end{aligned}\right\} \text{ in } \Omega \tag{3}$$

and

$$\bar{T}_i = n_j c_{ijk\ell}\, u_{k,\ell}, \text{ on } \partial\Omega_T \quad .$$

(1) can be generalized in the usual way to include the initial conditions as Euler-Lagrange equations (see [8]). However, this is peripheral to our main purpose here.

A suitable functional for the case of free vibration may be deduced from (1). Namely, assume harmonic dependence, $f_i = 0$ and homogeneous boundary conditions; then

$$G = \int_\Omega \{ \tfrac{1}{2}\omega^2 \rho v_i v_i + \omega^2 p_i\,(u_i - v_i) + \tfrac{1}{2} c_{ijk\ell}\, u_{i,j}\, u_{k,\ell} \}\, dx, \tag{4}^*$$

where ω is the circular frequency, is stationary, i.e.,

$$0 = \int_\Omega \{ \omega^2(\rho v_i - p_i)\beta_i + \omega^2(u_i - v_i)\gamma_i - (-\omega^2 p_i + (c_{ijk\ell}\, u_{k,\ell}),_j)\alpha_i \}\, dx + \int_{\partial\Omega_T} n_j\, c_{ijk\ell}\, u_{k,\ell}\, \alpha_i\, da, \tag{5}$$

for all α_i, β_i and γ_i such that $\alpha_i = 0$ on $\partial\Omega_u$, if and only if u_i, v_i and p_i satisfy the equations of free vibrations:

*Strictly speaking, one should introduce time-independent functions $\bar{u}_i$, (cont'd.)

$$\left.\begin{aligned} p_i &= \rho v_i \quad , \\ v_i &= u_i \quad , \\ 0 &= -\omega^2 p_i + (c_{ijk\ell}\, u_{k,\ell})_{,j} \end{aligned}\right\} \text{ in } \Omega \qquad (6)$$

and

$$0 = n_j\, c_{ijk\ell}\, u_{k,\ell}, \text{ on } \partial\Omega_T.$$

In the sequel we consider the special case in which it is assumed that $p_i = \rho v_i$ *ab initio*. Substituting this constraint into (1), we obtain

$$H = \int_0^T \{ \int_\Omega (\tfrac{1}{2} \rho v_i v_i - \rho v_i \dot{u}_i + \tfrac{1}{2} c_{ijk\ell}\, u_{i,j} u_{k,\ell} - f_i u_i)\, dx - \int_{\partial\Omega_T} \bar{T}_i u_i\, da \}\, dt, \qquad (7)$$

Assuming the same conditions which lead to (2), we obtain all but $(3)_1$ as Euler-Lagrange equations for H.

In a similar fashion, we obtain from (4)

$$I = \int_\Omega \{ -\tfrac{1}{2} \omega^2 \rho v_i v_i + \omega^2 \rho v_i u_i + \tfrac{1}{2} c_{ijk\ell}\, u_{i,j} u_{k,\ell} \}\, dx, \qquad (8)$$

for which all but $(6)_1$ are Euler-Lagrange equations.

Similar results for beam, plate and shell theories are easily obtained.

With independent interpolatory assumptions for u_i and v_i, we are able to create finite element methods with mass matrices other than those which have been termed "consistent."

3. Finite Element Formulation.

The variational theorems presented in the preceding section may be used to derive finite element models in which alternative descriptions of the mass matrix are possible. Consider an individual element. Select shape functions*

$\bar{v}_i$ and $\bar{p}_i$, defined by $u_i = \bar{u}_i \sin \omega t$, $v_i = \omega \bar{v}_i \cos \omega t$ and $p_i = \omega \bar{p}_i \cos \omega t$, when discussing the case of free vibration. To keep down the proliferation of notations, we shall retain the use of u_i, v_i and p_i for this case also. There should be no confusion as the context makes the meaning of the variables clear.

*Warning: We shall not introduce new notations for the approximate finite element fields which we employ in the present section.

$$\begin{Bmatrix} u_1 \\ u_2 \\ u_3 \end{Bmatrix} = \underset{\sim}{\phi}_e \, \underset{\sim}{u}_e \quad , \quad \begin{Bmatrix} v_1 \\ v_2 \\ v_3 \end{Bmatrix} = \underset{\sim}{\psi}_e \, \underset{\sim}{v}_e \quad , \tag{9}$$

where $\underset{\sim}{u}_e$ and $\underset{\sim}{v}_e$ are the e^{th} element's nodal displacement and velocity vectors, respectively. Note that $\underset{\sim}{\phi}_e$ and $\underset{\sim}{\psi}_e$ are in general not the same. Assume, for simplicity, that F_i and $\bar{T}_i$ are zero. Substitute (9) into (7) and perform the integrations:

$$H = \sum_e \int_0^T \{ \tfrac{1}{2} \underset{\sim}{v}_e^T \underset{\sim}{w}_e \underset{\sim}{v}_e - \underset{\sim}{v}_e^T \underset{\sim}{a}_e \dot{\underset{\sim}{u}}_e + \tfrac{1}{2} \underset{\sim}{u}_e^T \underset{\sim}{k}_e \underset{\sim}{u}_e \} \, dt, \tag{10}$$

where $\sum_e$ indicates summation over all elements, Ω_e is the volume of the e^{th} element, $\underset{\sim}{w}_e = \int_{\Omega_e} \rho \underset{\sim}{\psi}_e^T \underset{\sim}{\psi}_e \, dx$, $\underset{\sim}{k}_e$ is the element stiffness matrix, and $\underset{\sim}{a}_e = \int_{\Omega_e} \rho \underset{\sim}{\psi}_e^T \underset{\sim}{\phi}_e \, dx$. Employing the obvious notation, the global equations, including the imposed kinematic constraints, are obtained by setting the first variation of H to zero:

$$\left.\begin{aligned} \underset{\sim}{A} \dot{\underset{\sim}{U}} &= \underset{\sim}{W} \underset{\sim}{V}, \\ \underset{\sim}{A}^T \dot{\underset{\sim}{V}} + \underset{\sim}{K} \underset{\sim}{U} &= \underset{\sim}{0}. \end{aligned}\right\} \tag{11}$$

In a similar fashion, we can use I to generate the matrix equations of free vibration (or, equivalently, we can assume harmonic dependence in (11)):

$$\left.\begin{aligned} \underset{\sim}{A} \underset{\sim}{U} &= \underset{\sim}{W} \underset{\sim}{V}, \\ -\omega^2 \underset{\sim}{A}^T \underset{\sim}{V} + \underset{\sim}{K} \underset{\sim}{U} &= \underset{\sim}{0} \end{aligned}\right\} \tag{12}$$

$\underset{\sim}{V}$ may be eliminated from (11) and (12) in the obvious way (assuming det $\underset{\sim}{W} \neq 0$):

$$\left.\begin{aligned} \underset{\sim}{M}^* \ddot{\underset{\sim}{U}} + \underset{\sim}{K} \underset{\sim}{U} &= \underset{\sim}{0} \quad , \\ \text{and} \qquad & \\ (\underset{\sim}{K} - \omega^2 \underset{\sim}{M}^*) \underset{\sim}{U} &= \underset{\sim}{0} \quad , \end{aligned}\right\} \tag{13}$$

respectively, where $\underset{\sim}{M}^* = \underset{\sim}{A}^T \underset{\sim}{W}^{-1} \underset{\sim}{A}$ is the mass matrix in the present theory. Note that if the velocity degrees of freedom are not coupled from element to element, then $\underset{\sim}{M}^*$ can be directly assembled from the element contributions $m_e^* = a_e \, w_e^{-1} \, a_e$.

The case of main interest to us is when the number of entries in $\underset{\sim}{U}$ exceeds the number in $\underset{\sim}{V}$. In this case it is possible to define reduced systems (i.e., ones involving fewer degrees of freedom). For example, assuming $\det \underset{\sim}{K} \neq 0$, we can eliminate $\underset{\sim}{U}$:

$$\left.\begin{aligned} \underset{\sim}{K}^* \ddot{\underset{\sim}{V}} + \underset{\sim}{W} \underset{\sim}{V} = \underset{\sim}{0} \quad , \\ (\underset{\sim}{W} - \omega^2 \underset{\sim}{K}^*) \underset{\sim}{V} = \underset{\sim}{0} \quad , \end{aligned}\right\} \qquad (14)$$

where $\underset{\sim}{K}^* = \underset{\sim\sim}{A K^{-1}} \underset{\sim}{A}^T$ is the reduced stiffness. In (14), W is banded whereas $\underset{\sim}{K}^*$ is full. The solution of these equations may be obtained by any number of available algorithms.

Another possibility for constructing reduced systems is to employ <u>global</u> approximations for $\underset{\sim}{v}$ in conjunction with the usual finite element approximations of $\underset{\sim}{u}$. The reduced systems would have the same form as those in (14). An approach such as this might be useful when the geometric complexity of the structure in question necessitates a fine discretization to define the stiffness, but only very low mode response is of interest.

To convert the reduced systems to forms which can be solved by standard algorithms, the following procedures may be employed.

For implicit algorithms for $(14)_1$ or generalized eigenvalue form algorithms for $(14)_2$:

(i) Factor $\underset{\sim}{K}$: $\underset{\sim}{K} = \underset{\sim}{L} \, \underset{\sim}{L}^T$, where $\underset{\sim}{L}$ is lower triangular.

(ii) Solve $\underset{\sim}{L} \, \underset{\sim}{Z} = \underset{\sim}{A}^T$ for $\underset{\sim}{Z}$.

(iii) Form $\underset{\sim}{K}^* = \underset{\sim}{Z}^T \underset{\sim}{Z}$, making use of the symmetry of $\underset{\sim}{K}^*$. Solve $(14)_1$ or $(14)_2$.

(iv) To recover $\underset{\sim}{U}$, solve:

$\underset{\sim}{L}^T\underset{\sim}{U} = -\underset{\sim}{Z}\dot{\underset{\sim}{V}}$ for $(14)_1$;

or $\underset{\sim}{L}^T\underset{\sim}{U} = \omega^2\underset{\sim}{Z}\underset{\sim}{V}$ for $(14)_2$.

For explicit algorithms for the time-dependent case or standard eigenvalue form algorithms, repeat steps (i) and (ii) above, then:

(iii)' Reduce $\underset{\sim}{Z}$: $\underset{\sim}{Z} = \underset{\sim}{Q}\underset{\sim}{R}$ where $\underset{\sim}{Q}^T\underset{\sim}{Q} = \underset{\sim}{I}$ and $\underset{\sim}{R}$ is upper triangular. Note $\underset{\sim}{K}^* = \underset{\sim}{R}^T\underset{\sim}{R}$.

(iv)' Solve $\underset{\sim}{R}^T\underset{\sim}{S}\underset{\sim}{R} = \underset{\sim}{W}$ for $\underset{\sim}{S}$.

(v)' Solve for $\underset{\sim}{Y}$:

$$\ddot{\underset{\sim}{Y}} + \underset{\sim}{S}\underset{\sim}{Y} = \underset{\sim}{0},$$

$$\text{or } (\underset{\sim}{S} - \omega^2\underset{\sim}{I})\underset{\sim}{Y} = \underset{\sim}{0},$$

$$\text{where } \underset{\sim}{Y} = \underset{\sim}{R}\underset{\sim}{V}.$$

(vi)' Recover $\underset{\sim}{U}$ by solving:

$$\underset{\sim}{L}^T\underset{\sim}{U} = -\underset{\sim}{Q}\dot{\underset{\sim}{Y}},$$

$$\text{or } \underset{\sim}{L}^T\underset{\sim}{U} = \omega^2\underset{\sim}{Q}\underset{\sim}{Y}.$$

4. Error Analysis.

In this section we establish the error estimates for the reduced systems. Ample background for the ensuing analyses is provided by the book of Strang and Fix [13].

Throughout this section we adopt much of the standard error analysis notation. The way the preceding variational formulations fit into the general scheme to follow should be obvious. In the sequel, c denotes a general constant whose value may change from line-to-line in the inequality in question.

By C^k we mean the space of functions $u:\Omega\to R^n$ whose (classical) derivatives of order ℓ, $0\le\ell\le k$, exist and are continuous throughout Ω. Here we assume Ω is a bounded region in R^n with boundary $\partial\Omega$ of class C^∞.

Let L_2 denote the space of (equivalence classes of) mappings $u:\Omega\to R^n$ which are Lebesgue square integrable, i.e., $\int_\Omega u\cdot u\, dx < \infty$. The L_2 inner product and

norm are defined in the usual way: $(u,v) = \int_\Omega u\cdot v\, dx$, and $||u|| = (u,u)^{1/2}$, respectively.

Let H^s denote the Sobolev space of mappings $u:\Omega\to R^n$ which have (distributational) derivatives of order ℓ, $0\le\ell\le s$, in L_2. H^s is a Hilbert space with inner product and norm: $(u,v)_s = \sum_{\ell=0}^{s} (D^\ell u, D^\ell v)$, and $||u||_s = (u,u)_s^{1/2}$, respectively, where D^ℓ indicates the total derivative of order ℓ.

Let A be a linear partial differential operator of order 2m, with smooth (i.e., C^∞) coefficients, having dense domain D_A in L_2. We assume A is elliptic and that there exist positive constants c_1 and c_2 such that $c_1 ||u||_m^2 \le (Au,u) \le c_2||u||_m^2$, for all u in D_A.

To fix ideas we shall consider the case of the boundary value problem

$$Au + qu = f, \tag{15}$$

where q is a smooth positive function defined on the closure of Ω, f is in L_2 and u is required to satisfy appropriate conditions on $\partial\Omega$. Without loss of generality, we may assume these boundary conditions to be homogeneous. In this case it is well known from the theory of partial differential equations that

$$||u||_{s+m} \le c\, ||f||_s,$$

where c is a constant. In particular, if f is in C^∞ then u is in C^∞.

We are primarily interested in the eigenvalue problem

$$Au = \lambda u, \tag{16}$$

where again u is required to satisfy the boundary conditions. For this case it is well known that there exists an infinite sequence of real, positive eigenvalues,

$$0<\lambda_1 \le \lambda_2 \le \ldots,$$

and corresponding orthonormal eigenvectors $u_1, u_2, \ldots$, of class C^∞.

The energy inner product is defined by integration by parts:

$$a(u,v) = (Au,v),$$

where u,v satisfy the boundary conditions. The Galerkin equations corresponding to (15) and (16) are

$$a(u,v) + (qu,v) = (f,v), \tag{17}$$

$$a(u,v) = \lambda\,(u,v), \tag{18}$$

respectively, where u and v are in $E \equiv \{u\colon u \text{ is in } H^m$ and satisfies certain essential boundary conditions$\}$. A weak solution of (15) or (16) is a function u in E which satisfies (17) or (18), respectively, for all v in E. The Galerkin equations are the basis of finite element approximations to (15) and (16).

Let S^h and $\hat{S}^h$ be closed, finite-dimensional subspaces of E. Let $N = \dim S^h$ and $\hat{N} = \dim \hat{S}^h$. These spaces are to be thought of as finite-element spaces with mesh parameter h. Let

$$\pi : L_2 \to S^h \quad \text{and} \quad \hat{\pi} : L_2 \to \hat{S}^h,$$

denote orthogonal projection operators with respect to the L_2 inner product. We assume $S^h \supset P_k$ and $\hat{S}^h \supset P_{\hat{k}}$, where P_k is the space of complete polynomials of degree k. In addition we assume that the following approximation theorems hold for S^h and $\hat{S}^h$ (cf. Ciarlet-Raviart [3]):

$$\begin{aligned} |v - \pi v|_\ell &\le c_1 h^{k+1-\ell}\, |v|_{k+1}, \\ |v - \hat{\pi} v|_\ell &\le c_2 h^{\hat{k}+1-\ell}\, |v|_{\hat{k}+1}, \end{aligned} \tag{19}$$

for all v in E, where $|v|_\ell = (D^\ell v, D^\ell v)^{1/2}$.

4.1 Definition. u^h in S^h is the consistent finite element approximation to u, the solution of (15), if and only if

$$a(u^h,w^h) + (qu^h,w^h) = (f,w^h), \tag{20}$$

for all w^h in S^h.

4.2 Remark. The standard error estimate for the consistent approximation is (see Strang-Fix [13]):

$$||e^h||_m \leq c\, h^{k+1-m}\, |u|_{k+1}, \tag{21}$$

where $e^h = u - u^h$.

4.3 Definition. $\tilde{u}^h$ in S^h is the reduced finite element approximation to u, the solution of (15), if and only if

$$a(\tilde{u}^h, w^h) + (v^h, w^h) = (f, w^h), \tag{22}$$

and

$$(v^h, x^h) = (q\,\tilde{u}^h,\, x^h), \tag{23}$$

for all w^h in S^h, x^h in $\hat{S}^h$, where v^h is in $\hat{S}^h$.

4.4 Proposition. Assume $S^h \subset \hat{S}^h$. Then $\hat{u}^h = u^h$.

Proof. In this case we may select $x^h = w^h$ in (23). Thus $\tilde{u}^h$ satisfies the same equation as u^h. □

4.5 Remark. This proposition establishes the intuitively obvious fact that using higher-order finite element spaces for lower-order terms does not improve in any way upon the consistent approximation.

4.6 Theorem. Let $\tilde{e}^h = u^h - \tilde{u}^h$. Then $||\tilde{e}^h||_m \leq c h^{\hat{k}+1} |\tilde{u}^h|_{\hat{k}+1}$.

Proof. Subtracting (22) from (20) we get

$$a(\tilde{e}^h, w^h) + (qu^h - v^h, w^h) = 0.$$

By adding and subtracting $q\tilde{u}^h$, in the second term, we obtain:

$$a(\tilde{e}^h, w^h) + (q\tilde{e}^h, w^h) = -\,(q\tilde{u}^h - v^h, w^h).$$

Since $\tilde{e}^h$ is in S^h, we may choose $w^h = \tilde{e}^h$ in the above. By the assumptions on A and q, we have then that

$$\begin{aligned} ||\tilde{e}^h||_m^2 &\leq c\,\{\, a(\tilde{e}^h, \tilde{e}^h) + (q\,\tilde{e}^h,\, \tilde{e}^h)\,\}, \\ &= -c\,(q\tilde{u}^h - v^h,\, \tilde{e}^h), \\ &\leq c\,||q\tilde{u}^h - v^h||\;\; ||\tilde{e}^h||, \\ &\leq c\,||q\tilde{u}^h - v^h||\;\; ||\tilde{e}^h||_m. \end{aligned}$$

We have employed the Schwartz inequality in obtaining the third line. Thus we have

$$||\tilde{e}^h||_m \leq c\ ||q\tilde{u}^h - v^h||.$$

From (23) we see that $v^h = \hat{\pi}(qu^h)$. Combining this fact with the approximation theorem, (19), we obtain

$$||\tilde{e}^h||_m \leq c\ h^{\hat{k}+1}|\tilde{u}|_{\hat{k}+1}. \square$$

4.7 Remark. Combining this result with the standard error estimate for the consistent approximation, (21), we see that the full rate of convergence for energy is maintained as long as $\hat{k} \geq k-m$. This result can be trivially generalized to the case where (qu,v) is replaced by a positive definite bilinear form $b(u,v)$. For example, if b corresponds to a differential operator B of order $2n$, $n \leq m$, with smooth coefficients, then we have the estimate

$$||\tilde{e}^h||_m \leq c\ h^{\hat{k}+1-n}\ |\tilde{u}|_{\hat{k}+1},$$

from which it follows that the full rate of convergence is maintained if $\hat{k} \geq k-m+n$.

We shall now consider the eigenvalue problem.

4.8 Definition. u_ℓ^h in S^h and λ_ℓ^h in R are the consistent finite element approximations to u_ℓ, the ℓ^{th} eigenvector, and λ_ℓ, the ℓ^{th} eigenvalue, respectively, of (16) if and only if

$$a(u_\ell^h, w^h) = \lambda_\ell^h\ (u_\ell^h, w^h), \tag{24}$$

for all w^h in S^h, and

$$\lambda_\ell^h = \min_{S_\ell^h \subset S^h} \left\{ \max_{w^h \text{ in } S_\ell^h} R(w^h) \right\}, \tag{25}$$

where S_ℓ^h is any ℓ-dimensional subspace of S^h and $R(w^h) = a(w^h,w^h)/(w^h,w^h)$, the Rayleigh quotient.

4.9 Remark. The error estimates for (24) are standard (cf. Strang and Fix [13]):

$$\lambda_\ell \leq \lambda_\ell^h \leq \lambda_\ell + c\ h^{2(k+1-m)}\ \lambda_\ell^{(k+1)/m}, \tag{26}$$

$$||u_\ell - u_\ell^h|| \le c\, h^{k+1}\, \lambda_\ell^{(k+1)/2m} \quad ,$$
$$||u_\ell - u_\ell^h||_m \le c\, h^{k+1-m}\, \lambda_\ell^{(k+1)/2m} \quad . \qquad (26)\ (\text{Cont.})$$

4.10 Definition. $\tilde{u}_\ell^h$ in S^h and $\tilde{\lambda}_\ell^h$ in R are the reduced finite element approximations to u_ℓ and λ_ℓ, respectively, if and only if

$$a(\tilde{u}_\ell^h, w^h) = \tilde{\lambda}_\ell^h\, (v_\ell^h, w^h), \qquad (27)$$

$$(v_\ell^h,\ x^h) = (\tilde{u}_\ell^h,\ x^h), \qquad (28)$$

for all w^h in S^h, x^h in $\hat{S}^h$, where v_ℓ^h is in $\hat{S}^h$, and

$$\tilde{\lambda}_\ell^h = \min_{S_\ell^h \subset S^h} \left\{ \max_{w^h \text{ in } S_\ell^h} \tilde{R}(w^h) \right\}, \qquad (29)$$

where $\tilde{R}(w^h) = a(w^h, w^h)/(\hat{\pi} w^h, w^h)$.

4.11 Remark. It is immediate from (25), (29) and the fact that projections decrease norm (i.e. $||\hat{\pi}w|| \le ||w||$ for all w in E), that $\tilde{\lambda}_\ell^h \ge \lambda_\ell^h$. In other words, the reduced eigenvalue approximations are bounded from below by the corresponding consistent eigenvalue approximations, which are in turn bounded from below by the exact values, i.e., $\tilde{\lambda}_\ell^h \ge \lambda_\ell^h \ge \lambda_\ell$ for each $\ell = 1,2,\ldots,$ min $(N,\hat{N})$.

We note also that if $\tilde{\lambda}_\ell^h \ne \tilde{\lambda}_p^h$, then $v_\ell^h \perp v_p^h$ with respect to the L_2 inner product and $\tilde{u}_\ell^h \perp \tilde{u}_p^h$ with respect to the energy inner product. These are easy to establish.

Let $w^h = \tilde{u}_p^h$ in (27) and let $x^h = v_p^h$ in (28):

$$a(\tilde{u}_\ell^h,\ \tilde{u}_p^h) = \tilde{\lambda}_\ell^h\, (v_\ell^h,\ \tilde{u}_p^h),$$
$$(v_\ell^h,\ v_p^h) = (\tilde{u}_\ell^h,\ v_p^h). \qquad (30)$$

Now replace ℓ by p in (27) and (28), and let $w^h = \tilde{u}_\ell^h$ and $x^h = v_\ell^h$:

$$a(\tilde{u}_p^h,\ \tilde{u}_\ell^h) = \tilde{\lambda}_p^h\, (v_p^h,\ \tilde{u}_\ell^h),$$
$$(v_p^h, v_\ell^h) = (\tilde{u}_p^h,\ v_\ell^h). \qquad (31)$$

Combining (30) and (31) we obtain

$$(\tilde{\lambda}_\ell^h - \tilde{\lambda}_p^h)\ (v_\ell^h,\ v_p^h) = 0,$$

which implies $v_\ell^h \perp v_p^h$ in L_2. Since we also have

$$a(\tilde{u}_\ell^h,\ \tilde{u}_p^h) = \tilde{\lambda}_\ell^h\ \ (v_\ell^h,\ v_p^h),$$

and $\tilde{\lambda}_\ell^h > 0$, it follows that $\tilde{u}_\ell^h \perp \tilde{u}_p^h$ with respect to the energy inner product.

4.12 Proposition. Assume $S^h \subset \hat{S}^h$. Then $\tilde{u}_\ell^h = u_\ell^h$ and $\tilde{\lambda}_\ell^h = \lambda_\ell^h$.

Proof. Under this assumption $\hat{\pi}$ restricted to S^h is the identity map. Therefore (29) is equivalent to (25), i.e., $\tilde{\lambda}_\ell^h = \lambda_\ell^h$. Selecting $x^h = w^h$ in (28) and using this and $\tilde{\lambda}_\ell^h = \lambda_\ell^h$ in (27) implies $\tilde{u}_\ell^h = u_\ell^h$. □

4.13 Remark. Proposition 4.12 tells us that, within the present scheme, we cannot improve upon the consistent mass matrix. However, as we shall see below, we can define alternative mass descriptions which retain the full rate of convergence of the consistent mass matrix, and are of smaller size.

4.14 Lemma. Let $\sigma_\ell^h = \max_{w^h \text{ in } e_\ell^h} |(\hat{\pi}w^h - w^h, w^h)|$ where e_ℓ^h is the set of all unit vectors contained in E_ℓ^h, the ℓ-dimensional subspace of S^h spanned by u_1^h, u_2^h, ..., u_ℓ^h. Then $\tilde{\lambda}_\ell \le \lambda_\ell\ (1-\sigma_\ell^h)^{-1}$.

Proof. By (29) we have

$$\tilde{\lambda}_\ell^h \le \max_{w^h \text{ in } S_\ell^h} \tilde{R}(w^h),$$

$$= \max_{w^h \text{ in } e_\ell^h} \frac{a(w^h,w^h)}{(\hat{\pi}w^h,w^h)}\ .$$

Assuming w^h is in e_ℓ^h, a simple calculation yields:

$$(\hat{\pi}w^h,w^h) = (w^h - (w^h - \hat{\pi}w^h),\ w^h),$$

$$= (w^h,w^h) - (w^h - \hat{\pi}w^h,w^h),$$

$$\ge 1 - \sigma_\ell^h.$$

Combining the above results and using (25) gives us that

$$\tilde{\lambda}_\ell^h \leq (1 - \sigma_\ell^h)^{-1} \max_{w^h \text{ in } e_\ell^h} a(w^h, w^h),$$

$$= \lambda_\ell^h (1 - \sigma_\ell^h)^{-1}. \square$$

4.15 Lemma. $\sigma_\ell^h \leq c\, h^{2(\hat{k}+1)} |w^h|^2_{\hat{k}+1}$.

Proof. By definition of the projection $\hat{\pi}$, we have that $w^h - \hat{\pi}w^h$ is orthogonal to $\hat{S}^h$. Using this and the approximation estimate (19) we obtain

$$(w^h, w^h - \hat{\pi}w^h) = (w^h - \hat{\pi}w^h,\ w^h - \hat{\pi}w^h)$$

$$= ||w^h - \hat{\pi}w^h||^2$$

$$\leq c\, h^{2(\hat{k}+1)} |w^h|^2_{\hat{k}+1}. \square$$

4.16 Theorem. Assume h is small enough so that $\sigma_\ell^h \leq 1/2$. Then we obtain our error estimate for the eigenvalues of the reduced problem:

$$\tilde{\lambda}_\ell^h \leq \lambda_\ell^h (1 + c\, h^{2(\hat{k}+1)}). \tag{32}$$

Proof. By Lemmas 4.14 and 4.15 we have immediately that

$$\tilde{\lambda}_\ell^h \leq \lambda_\ell^h (1 + 2\sigma_\ell^h),$$

$$\leq \lambda_\ell^h (1 + c\, h^{2(\hat{k}+1)}). \square$$

4.17 Remark. Comparing this result with $(26)_1$, we see that if $\hat{k} \geq k-m$, then the full rate of convergence for eigenvalues of the consistent approximation is maintained by the reduced approximation. The situation for eigenvectors is similar, as we shall now show.

4.18 Lemma. Let $\tilde{e}_\ell^h = u_\ell^h - \tilde{u}_\ell^h$. Then

$$a(\tilde{e}_\ell^h, w^h) = \lambda_\ell^h (\tilde{e}_\ell^h, w^h) + (\lambda_\ell^h - \tilde{\lambda}_\ell^h)(\tilde{u}_\ell^h, w^h) + \tilde{\lambda}_\ell^h (\tilde{u}_\ell^h - v_\ell^h, w^h), \tag{33}$$

for all w^h in S^h.

Proof. Subtracting (27) from (24) we obtain

$$a(\tilde{e}_\ell^h, w^h) = (\lambda_\ell^h u_\ell^h - \tilde{\lambda}_\ell^h v_\ell^h, w^h),$$
$$= (\lambda_\ell^h(u_\ell^h - \tilde{u}_\ell^h) + (\lambda_\ell^h - \tilde{\lambda}_\ell^h)\tilde{u}_\ell^h + \tilde{\lambda}_\ell^h(\tilde{u}_\ell^h - v_\ell^h), w^h). \square$$

This identity enables us to estimate the H^m norm of $\tilde{e}_\ell^h$ in terms of the L_2 norm of $\tilde{e}_\ell^h$, the previously obtained estimate for $\lambda_\ell^h - \tilde{\lambda}_\ell^h$, and the L_2 norm of the lack-of-consistency $\tilde{u}_\ell^h - v_\ell^h$; viz., let $w^h = \tilde{e}_\ell^h$ in (33), then

$$||\tilde{e}_\ell^h||_m \le c\{\lambda_\ell^h||\tilde{e}_\ell^h|| + |\lambda_\ell^h - \tilde{\lambda}_\ell^h| + \tilde{\lambda}_\ell^h||\tilde{u}_\ell^h - v_\ell^h||\}. \qquad (34)$$

4.19 Lemma. $(\tilde{\lambda}_j^h - \lambda_i^h)(\hat{\pi}u_i^h, \tilde{u}_j^h) = \lambda_i^h(u_i^h - \hat{\pi}u_i^h, \tilde{u}_j^h)$ for all i and j such that $1 \le i,j \le \min(N,\hat{N})$.

Proof. The term $-\lambda_i^h(\hat{\pi}u_i^h, \tilde{u}_j^h)$ appears on both sides so it remains to show that $\tilde{\lambda}_j^h(\hat{\pi}u_i^h, \tilde{u}_j^h) = \lambda_i^h(u_i^h, \tilde{u}_j^h)$. To do this we employ (24) and (27):

$$\tilde{\lambda}_j^h(\hat{\pi}u_i^h, \tilde{u}_j^h) = \tilde{\lambda}_j^h(\hat{\pi}u_i^h, \hat{\pi}\tilde{u}_j^h),$$
$$= \tilde{\lambda}_j^h(u_i^h - (u_i^h - \hat{\pi}u_i^h), \hat{\pi}\tilde{u}_j^h),$$
$$= \tilde{\lambda}_j^h(u_i^h, \hat{\pi}\tilde{u}_j^h) - \tilde{\lambda}_j^h(u_i^h - \hat{\pi}u_i^h, \hat{\pi}\tilde{u}_j^h),$$
$$= a(\tilde{u}_j^h, u_i^h);$$
$$\lambda_i^h(u_i^h, \tilde{u}_j^h) = a(u_i^h, \tilde{u}_j^h). \square$$

4.20 Lemma. Assume that the multiplicity of λ_i is one. Then

$$||\hat{\pi}u_i^h - \beta v_i^h|| \le c||u_i^h - \hat{\pi}u_i^h||,$$

where $\beta = (\hat{\pi}u_i^h, v_i^h)$.

Proof. Note that $\{u_i^h\}_1^N$ and $\{v_i^h\}_1^{\hat{N}}$ constitute orthogonal bases for S^h and $\hat{S}^h$, respectively. For convenience we assume $||u_i^h|| = 1$, $1 \le i \le N$, and $||\tilde{u}_j^h|| = 1$, $1 \le j \le \hat{N}$. Since $\hat{\pi}u_i^h$ is in $\hat{S}^h$, we may expand it in terms of $\{v_i^h\}_1^{\hat{N}}$:

$$\hat{\pi}u_i^h - \beta v_i^h = \sum_{j\neq i}^{\hat{N}} (\hat{\pi}u_i^h, v_j^h)v_j^h.$$

The estimates $(26)_1$ and (32) and the fact that λ_i is isolated imply that there exists a constant ρ such that

$$\frac{\lambda_i^h}{|\tilde{\lambda}_j^h - \lambda_i^h|} \leq \rho \ , \text{ for all } j \neq i, \tag{35}$$

whenever h is small enough.

By the definition of $\hat{\pi}$, $(\hat{\pi}u_i^h, \tilde{u}_j^h) = (\hat{\pi}u_i^h, \hat{\pi}\tilde{u}_j^h)$. Now using Lemma 4.19 and the preceding relations, we have that

$$\begin{aligned} ||\hat{\pi}u_i^h - \beta v_i^h||^2 &\leq \sum_{j\neq i}^{\hat{N}} (\hat{\pi}u_i^h, v_j^h)^2 \ , \\ &= \sum_{j\neq i}^{\hat{N}} \left\{ \frac{\lambda_i^h}{|\tilde{\lambda}_j^h - \lambda_i^h|} \right\}^2 (u_i^h - \hat{\pi}u_i^h, \tilde{u}_j^h)^2, \\ &\leq \rho^2 \, ||u_i^h - \hat{\pi}u_i^h||^2. \ \square \end{aligned}$$

4.21 Lemma. $||\tilde{e}_i^h|| \leq 2||u_i^h - \beta\tilde{u}_i^h||$.

Proof. By the triangle inequality

$$\begin{aligned} ||\tilde{e}_i^h|| = ||u_i^h - \tilde{u}_i^h|| &\leq ||u_i^h - \beta\tilde{u}_i^h|| + ||\beta\tilde{u}_i^h - u_i^h|| \ , \\ &= ||u_i^h - \beta\tilde{u}_i^h|| + ||(\beta-1)\tilde{u}_i^h|| \ , \\ &\leq ||u_i^h - \beta\tilde{u}_i^h|| + |\beta-1| \ . \end{aligned}$$

We may choose the sign of v_i^h such that $\beta \geq 0$. Using the fact that u_i^h and $\tilde{u}_i^h$ are unit vectors we get

$$\begin{aligned} 1 = ||u_i^h|| &\leq ||u_i^h - \beta\tilde{u}_i^h|| + ||\beta\tilde{u}_i^h||, \\ &= ||u_i^h - \beta\tilde{u}_i^h|| + \beta. \end{aligned}$$

Combining this with the previous result completes the proof. $\square$

4.22 Lemma. Assume λ_i is isolated. Then we have the estimates

$$||\tilde{e}_i^h|| \leq c \, h^{\hat{k}+1}, \tag{36}$$

and

$$||\tilde{e}_i^h||_m \leq c\, h^{\hat{k}+1}. \tag{37}$$

Proof. Applying the triangle inequality to the result of Lemma 4.21, we get

$$||\tilde{e}_i^h|| \leq 2\{||u_i^h - \hat{\pi}u_i^h|| + ||\hat{\pi}_i^h - \beta v_i^h|| + \beta||v_i^h - \tilde{u}_i^h||\}.$$

Using Lemma 4.20 and the approximation estimates (19), we obtain $||\tilde{e}_i^h|| \leq c\, h^{\hat{k}+1}$. Employing this result in (34) yields that $||\tilde{e}_i^h||_m \leq c\, h^{\hat{k}+1}$. □

4.23 Remark. Comparing these results with $(26)_3$, we see that if $\hat{k} \geq k-m$ the H^m rate of convergence for eigenvectors is maintained.

We shall now remove the restriction that λ_i be isolated. The argument is tedious, but not essentially different than before, so we only sketch the main points.

4.24 Lemma. Let λ_i have multiplicity Q, where Q is a positive integer > 0. Then (36) and (37) still hold.

Sketch of proof. Let $\lambda_i = \lambda_{i+1} = \ldots = \lambda_{i+Q}$. There is still a separation constant between these eigenvalues and the others (cf. (35)). Under these circumstances the analog of Lemma 4.20 is

$$||\hat{\pi}u_{i+r}^h - \sum_{j=0}^{Q} \beta_{rj}\, v_{i+j}^h|| \leq c||\tilde{u}_{i+r}^h - v_{i+r}^h||, \tag{38}$$

where $\beta_{rj} = (\hat{\pi}u_{i+r}^h, v_{i+j}^h)$, $0 \leq r,j \leq Q$.

Let $\underset{\sim}{\alpha} = \underset{\sim}{\beta}^{-1}$, where $\underset{\sim}{\beta} = [\beta_{ij}]$, $0 \leq i,j \leq Q$. We define linear combinations of the eigenvectors as follows:

$$U_{i+r} = \sum_{j=0}^{Q} \alpha_{ij}\, u_{j+r},$$

$$U_{i+r}^h = \sum_{j=0}^{Q} \alpha_{ij}\, u_{j+r}^h,$$

where $0 \leq r \leq R$. Let $\hat{U}_{i+r}^h = \hat{\pi}U_{i+r}^h = \sum_{j=0}^{Q} \alpha_{ij}\, \hat{\pi}u_{j+r}^h$.

Employing the triangle inequality, we can estimate the difference between U_{i+r} and its reduced approximation, $\tilde{u}^h_{i+r}$:

$$||U_{i+r} - \tilde{u}^h_{i+r}|| \le ||U_{i+r} - U^h_{i+r}|| + ||U^h_{i+r} - \hat{U}^h_{i+r}||$$

$$+ ||\hat{U}^h_{i+r} - v^h_{i+r}|| + ||\hat{\pi}\tilde{u}^h_{i+r} - \tilde{u}^h_{i+r}||.$$

The first term on the right-hand side can be estimated using $(26)_2$; the second and fourth can be taken care of by the approximation estimate (19); for the third term we employ (38):

$$||\hat{U}^h_{i+r} - v^h_{i+r}|| = ||\sum_{j=0}^{Q} \alpha_{ij}\, \hat{\pi}u^h_{j+r} - v^h_{i+r}||\ ,$$

$$= ||\sum_{j=0}^{Q} \alpha_{ij}\, (\hat{\pi}u^h - \sum_{k=0}^{Q} \beta_{jk}\, v^h_{r+k})||\ ,$$

$$\le c \sum_{r=0}^{Q} ||\tilde{u}^h_{i+r} - v^h_{i+r}||.$$

Applying (19) completes the L_2 estimate, from which the H^m estimate follows. □

4.25 Remark. All of the preceding results extend to the generalized eigenproblem in which (u,v) in (16) is replaced by a positive-definite bilinear form b(u,v). For example if b(u,v) = (Bu,v), where B is a linear differential operator of order 2n, $n \le m$, with smooth coefficients, then the condition for maintaining the full rate of convergence for eigenvalues and energy is that $\hat{k} \ge k-m+n$. The proofs go as before except, instead of $\hat{\pi}$, one must employ $\hat{P}$, the orthogonal projection onto $\hat{S}^h$ with respect to b.

We shall now consider time dependent problems. Let a superposed dot indicate time differentiation and let u(t) denote the function obtained from $u: R \times \Omega \to R^n$ by freezing t in R. We assume for simplicity that the coefficients of A are independent of t and that F is a C^∞ mapping from R to L_2, i.e., F is C^∞ in t. There are two important cases:

(hyperbolic case) $$\ddot{u} + Au = f \,, \tag{39}$$

(parabolic case) $$\dot{u} + Au = f \,, \tag{40}$$

where u is required to satisfy the boundary conditions and appropriate initial conditions. In the former case there are initial conditions on u and $\dot{u}$, whereas in the latter only u need be specified. The corresponding Galerkin equations are

$$(\ddot{u},v) + a(u,v) = (f,v) \,, \tag{41}$$

and

$$(\dot{u},w) + a(u,v) = (f,v) \,, \tag{42}$$

respectively, where u and v are in E. If u_0 and $\dot{u}_0$ are given initial data then

$$(u(o),v) = (u_0,v), \tag{43}$$

and

$$(\dot{u}(o),v) = (\dot{u}_0,v), \tag{44}$$

are the weak forms of the initial conditions for the hyperbolic case.

In the parabolic case only (43) is required. We assume u_0 and $\dot{u}_0$ are in L_2 and H^m, respectively.

4.25 Definition. u^h in S^h is the consistent finite element approximation to u, the solution of (39) if and only if

$$(\ddot{u}^h,w^h) + a(u^h,w^h) = (f,w^h), \tag{45}$$

$$(u^h(o),w^h) = (u_0,w^h), \tag{46}$$

and

$$(\dot{u}^h(o),w^h) = (\dot{u}_0,w^h), \tag{47}$$

for all w^h in S^h. From the theory of ordinary differential equations u^h is a C^∞ mapping from R to H^m.

4.26 Remark. The error for the hyperbolic case is conveniently measured in terms of the energy

$$E = K(e^h) + U(e^h), \tag{48}$$

where $K(e^h) = 1/2(\dot{e}^h,\dot{e}^h)$ and $U(e^h) = 1/2\ a(e^h,e^h)$. Because of our hypotheses on the operator A, $E \to 0$ is equivalent to $||\dot{e}^h|| \to o$ and $||e^h||_m \to o$. The standard error estimate for the consistent approximation is (see [13]):

$$E^{1/2} \leq c(h^{k+1-m} + h^{k+1}t) \tag{49}$$

4.27 Definition. $\tilde{u}^h$ in S^h is the reduced finite element approximation to u, the solution of (39), if and only if

$$(\dot{v}^h,w^h) + a(\tilde{u}^h,w^h) = (f,w^h), \tag{50}$$

$$(v^h,x^h) = (\dot{\tilde{u}}^h,x^h), \tag{51}$$

$$(\tilde{u}^h(o),w^h) = (u_o,w^h), \tag{52}$$

and

$$(v^h(o),x^h) = (\dot{u}_o,x^h), \tag{53}$$

for all w^h in S^h, x^h in $\hat{S}^h$, where v^h is in $\hat{S}^h$. As in the case of the consistent approximation, $\tilde{u}^h$ is C^∞ in t.

4.28 Proposition. Assume $S^h \subset \hat{S}^h$. Then $\tilde{u}^h = u^h$.

Proof. Pick $x^h = w^h$ in (51) and (53). Then (51) evaluated at t=0, when combined with (53), shows $\tilde{u}^h$ satisfies (47). Time differentiating (51) and substituting the result in (50) yields that $\tilde{u}^h$ satisfies (45). Since (52) is equivalent to (46), $\tilde{u}^h$ satisfies the same equations as u^h. □

4.29 Theorem. Let $\tilde{E} = K(\dot{\tilde{e}}^h) + U(\tilde{e}^h)$ Then $\tilde{E} \leq \tilde{E}_o + ch^{\hat{k}+1}t$.

Proof. Subtracting (50) from (45) we get

$$(\ddot{u}^h - \dot{v}^h,w^h) + a(\tilde{e}^h,w^h) = 0$$

Adding and subtracting $\ddot{\tilde{u}}^h$ in the first term, and observing that $\dot{v}^h = \hat{\pi}\ddot{\tilde{u}}^h$ from (51), results in

$$(\ddot{\tilde{e}}^h,w^h) + a(\tilde{e}^h,w^h) = -(\ddot{u}^h - \hat{\pi}\ddot{\tilde{u}}^h,w^h).$$

Selecting $w^h = \dot{\tilde{e}}^h$ in the above yields

$$\frac{d}{dt}\ \tilde{E} \le ||\ddot{\tilde{u}}^h - \hat{\pi}\ddot{u}^h\ ||\ ||\dot{e}^h\ ||,$$

$$= ||\ddot{\tilde{u}}^h - \hat{\pi}\ddot{u}^h\ ||\ \sqrt{2}\ K(\tilde{e}^h)^{1/2},$$

$$\le ||\ddot{\tilde{u}}^h - \hat{\pi}\ddot{u}^h\ ||\ \sqrt{2}\ \tilde{E}^{1/2}.$$

The approximation theorem, (19), when applied to this result yields that

$$\frac{d}{dt}\ \tilde{E} \le c\ h^{\hat{k}+1}\ \tilde{E}^{1/2}.$$

Integrating this relation over 0 to t yields

$$\tilde{E}^{1/2} \le \tilde{E}_o^{1/2} + c\ h^{\hat{k}+1} t,$$

which was to be proved. □

4.30 Remark. $\tilde{E}_o^{1/2}$ is of order $\min(\hat{k}+1,\ k+1-m)$. Therefore, if $\hat{k} \ge k-m$ the full rate of convergence of the consistent approximation is maintained, at least for short times.

We shall now consider the parabolic case.

4.31. Definition. u^h in S^h is the consistent finite element approximation to u, the solution of (40), if and only if

$$(\dot{u}^h, w^h) + a(u^h, w^h) = (f, w^h), \tag{54}$$

$$(u^h(o), w^h) = (u_o, w^h), \tag{55}$$

for all w^h in S^h; u^h is C^∞ in t.

4.32 Remark. The error estimate for the consistent approximation is (see [13]):

$$||e^h||_m \le c\ h^{k+1-m}\ ||u||_{k+1},$$

for all t.

4.33 Definition. $\tilde{u}^h$ in S^h is a reduced finite element approximation to u, the solution of (40), if and only if

$$(v^h,w^h) + a(\tilde{u}^h,w^h) = (f,w^h), \tag{56}$$

$$(v,x^h) = (\dot{\tilde{u}},x^h), \tag{57}$$

and

$$(\tilde{u}(o),w^h) = (u_o,w^h), \tag{58}$$

for all w^h in S^h, x^h in $\hat{S}^h$, where v^h is in $\hat{S}^h$; $\tilde{u}^h$ is C^∞ in t.

4.34 Proposition. Assume $S^h \subset \hat{S}^h$. Then $\tilde{u}^h = u^h$.

Proof. The proof goes along the same lines as that for Prop. 4.28. □

4.35 Theorem. The error $||\tilde{e}^h||_m$ in the reduced approximation is of order min $(k+1, \hat{k}+1)$.

4.36 Proof. Subtracting (56) from (54), using (57), and taking $w^h = \tilde{e}^h$ yields

$$(\dot{e}^h,\tilde{e}^h) + a(\tilde{e}^h,\tilde{e}^h) = -(\dot{\tilde{u}}^h - \hat{\pi}\dot{\tilde{u}}^h,\tilde{e}^h). \tag{59}$$

It follows directly that

$$\frac{d}{dt}\, 1/2\ ||\tilde{e}^h||^2 + \lambda_1||\tilde{e}^h||^2 \leq ||\dot{\tilde{u}}^h - \hat{\pi}\dot{\tilde{u}}^h||\ ||\tilde{e}^h||.$$

Cancelling $||\tilde{e}^h||$ from the above results in

$$\frac{d}{dt}\ ||\tilde{e}^h|| + \lambda_1\ ||\tilde{e}^h|| \leq ||\dot{\tilde{u}} - \hat{\pi}\dot{\tilde{u}}^h||.$$

Multiplying this inequality by exp $(\lambda_1 t)$, and integrating from o to t yields

$$\begin{aligned} ||\tilde{e}^h(t)|| &\leq \exp(-\lambda_1 t)\ \ ||\tilde{e}^h(o)|| + \\ &\qquad \int_o^t \exp(\lambda_1(\tau-t))\ \ ||\dot{\tilde{u}}^h(\tau) - \hat{\pi}\dot{\tilde{u}}^h(\tau)||d\tau, \\ &\leq c\{h^{\hat{k}+1}\ ||u_o||_{k+1}\ \exp(-\lambda_1 t) \\ &\quad + h^{\hat{k}+1} \int_o^t \exp(\lambda_1(\tau-t))\ \ ||\dot{\tilde{u}}(\tau)||d\tau\}, \end{aligned} \tag{60}$$

where we have applied (19) in deriving the second line.

To obtain an estimate for $a(\tilde{e}^h,\tilde{e}^h)$, we integrate (59) from t_1 to t_2:

$$\begin{aligned} 2\int_{t_1}^{t_2} a(\tilde{e}^h,\tilde{e}^h)d\tau &= ||\tilde{e}^h(t_1)||^2 - ||\tilde{e}^h(t_2)|| \\ &\quad - 2\int_{t_1}^{t_2} (\dot{\tilde{u}}^h - \hat{\pi}\dot{\tilde{u}}^h,\tilde{e}^h)d\tau. \end{aligned}$$

Applying (19), (60) and the mean value theorem to this result yields that

$a(\tilde{e}^h,\tilde{e}^h)$ is order $2\min(k+1,\hat{k}+1)$, from which the assertion of the theorem follows. □

4.36 Remark. Thus if $\hat{k} \geq k-m$ the rate of convergence of the consistent approximation is maintained (cf. Remark 4.32).

5. Discussion.

The previous developments enable us to design reduced finite element systems for dynamics which retain the rate of convergence of systems employing consistent mass matrices. Some examples are illustrative of the nature of the reduced system.

5.1 Beam Element.

For the standard cubic beam element ($k = 3$, $m = 2$) the full rate of convergence is maintained as long as $\hat{k} \geq 1$. The optimal choice is then a linear element interpolation for the velocity field ($\hat{k} = 1$) which may be made continuous at the nodes. This model, aside from the effects of boundary conditions, results in a reduced system of one-half the number of degrees of freedom as that of the standard consistent mass system.

5.2 Plate Bending Elements.

A survey of the standard error estimates for plate bending elements has been given by Ciarlet [2]. There are several basic plate bending elements which contain a full cubic displacement function ($k = 3$, $m = 2$) and are thus of quadratic convergence rate in the H^2 norm (e.g., the 16 degree of freedom rectangular element of Bogner, Fox and Schmidt [1], the 16 degree of freedom quadrilateral of Fraeijs de Veubeke [5], the 12 degree of freedom triangle of Clough and Tocher [4], etc.). To retain the rate of convergence of consistent mass for these cases one needs that $\hat{k} \geq 1$, i.e., the velocity field must contain a polynomial of the first degree. In the case of triangles this is achieved most simply by assuming a linear velocity field with nodal degrees of freedom at the vertices. For quadrilaterals it seems the most appropriate scheme is to employ a bilinear velocity field, also with nodal

degrees of freedom at the vertices. These procedures will result in reduced systems of approximately 1/4 the size for the Bogner, Fox and Schmidt rectangle, 1/5 for the Fraeijs de Veubeke quadrilateral and 1/6 for the Clough and Tocher triangle.*

A reduced system for the compatible 9 degree of freedom triangle ($k = 2$, $m = 2$) of Clough and Tocher can also be constructed, as above, with a linear velocity field. The reduced system would be approximately 1/3 the size of the consistent mass system and would also maintain the first-order convergence rate of this element. However, it may be preferable in this case to simply use lumped mass, i.e., assign one-third the total mass to each translatory degree of freedom. The standard result on numerical integration techniques (see Fried [6] and references therein) guarantees that the lumped mass matrix (which is exact only for uniform translation) retains the first-order rate of convergence of this element. A similar argument may be made for several other slowly converging plate bending elements (see Ciarlet [2] for examples).

5.3 Classical Elasticity.

Classical linear elasticity involves a second-order elliptic differential operator so that $m = 1$. Thus, to retain the convergence rate of consistent mass for compatible elements in which the displacement interpolations contain complete polynomials of degree k, one needs the velocity interpolations to contain complete polynomials of degree $\hat{k} = k - 1$. For the standard families of triangular and quadrilateral elements (see [14]) the velocity fields could be taken to be one order lower than the displacement fields. For example, for the quadratic triangle ($k = 2$) the velocity field could be taken to be linear and defined in terms of the three vertex degrees of freedom. The reduced system for this case approaches 1/4 the size of the consistent

*These ratios for the Fraeijs de Veubeke and Clough-Tocher elements are limiting values for infinite rectangular meshes.

mass system.

For this class of problems an alternative scheme has been proposed by Fried and Malkus [7] which is remarkably simple and produces a diagonal mass matrix. They choose the locations of the nodal degrees of freedom to coincide with the so-called Lobatto points. Numerical integration formulas are then derived employing these points, which insure the maximal rate of convergence. One unpleasant feature of the scheme is that for certain higher-order elements some negative masses occur. Zero masses may occur also, but these may be eliminated by static condensation, reducing the size of the system.

6. Numerical Examples.

To verify the results of the error analysis, spectral properties of three structural models were studied:

6.1 Quadratic Rod Element.

The differential equation for this model is

$$u'' + \omega^2 u = 0,$$

where u represents the longitudinal displacement of the rod and ω is the natural frequency; thus $m = 1$. The boundary conditions studied were fixed-free and fixed-fixed. The displacement field is assumed to vary quadratically within each element ($k = 2$), thus the element has 3 degrees of freedom; one at each end and one at the midpoint. Results are presented for three cases: a consistent mass matrix, a reduced system involving a linear velocity approximation for each element ($\hat{k} = 1$) and a diagonal mass matrix in which 1/6 the mass is lumped at the end-point degrees of freedom and 2/3 at the midpoint. The theoretical rate of convergence of consistent mass, which is quartic, is verified numerically in Fig. 1; $\bar{\omega}$ denotes the numerically

computed frequencies. The slope of the curves indicates the convergence rate of the fundamental frequency. It is also seen in Fig. 1 that the reduced system retains the convergence rate of the consistent mass matrix. This corroborates the present theory ($1 = \hat{k} \geq k-m = 1$). The diagonal mass matrix is also seen to retain the convergence rate of the consistent case. This matrix was arrived at via experimentation, but can be constructed following the theory of Fried and Malkus [7].

In Fig. 2, the complete spectra for the fixed-fixed case is presented; n is the mode number and N is the number of elements. These spectra are invariant and apply for all N. The upper bound property of the consistent and reduced systems is evident. In addition, the maximum relative error of any frequency is seen to be a minimum for the reduced case (approximately 15%). These properties, as will be shown in the following examples, are common to the cases we have investigated.

6.2 Beam Element.

The differential equation for this example is

$$u^{iv} - \omega^2 u = 0,$$

where u represents the tranverse displacement of the beam. For this case $m = 2$ and the standard cubic displacement function is employed ($k = 3$). Simply-supported boundary conditions are considered. Results are presented for four cases: a consistent mass matrix, a reduced system involving a linear velocity approximation for each element ($\hat{k} = 1$, cf. § 5.1), a diagonal mass matrix in which 1/2 the mass is lumped at each translatory degree of freedom and zero is assigned to the rotatory degrees of freedom, and a linear displacement function used for the mass matrix only. The last case results in a mass matrix of the form

$$\frac{M}{6}\left[\begin{array}{cc|c} 2 & 1 & \\ 1 & 2 & \underset{\sim}{0} \\ \hline & \underset{\sim}{0} & \underset{\sim}{0} \end{array}\right],$$

where M is the element mass. As can be seen from Fig. 3, this ad-hoc approach has an adverse effect on convergence rate. For all other cases the rate of convergence is quartic, which verifies the standard error estimate for the consistent case and the present theory for the reduced case (i.e., $1 = \hat{k} \geq k-m = 1$). The lumped mass matrix scheme used here is the best diagonal mass matrix for the beam.

In Fig. 4 spectra for these cases are presented. Again, the upper bound property of the consistent and reduced cases is evident, as is the fact that the reduced spectrum produces the least maximum relative error; approximately 12%.

6.3 Plate Bending Element.

Here the differential equation is

$$\nabla^4 u - \omega^2 u = 0,$$

where ∇^4 indicates the biharmonic operator and u denotes the transverse displacement of the plate; thus $m = 2$. The plate is square and simply-supported boundary conditions are employed. The Bogner-Fox-Schmidt element [1], which contains a complete cubic displacement function ($k = 3$), was chosen for study. Results for three cases are presented: consistent mass,

a reduced system for which a bilinear velocity assumption is made ($\hat{k} = 1$, cf. § 5.2) and a diagonal mass matrix in which 1/4 the plate mass is lumped at each translatory degree of freedom. As can be seen in Fig. 5 each case exhibits quartic rate of convergence. This corroborates the standard error estimate for the consistent case and the present estimate for the reduced case.

In Fig. 6 spectral results are presented for the above cases. The qualitative aspects of the previous problems are again in evidence.

Examples involving elements of unequal size were run for problems 6.1 and 6.2 to see if any of the results were special for regular meshes. In all cases the same rates of convergence were observed, although the spectral lines changed somewhat in each case. We believe the high rates of convergence of the lumped mass models are somewhat accidental; there is no analytical evidence extant which indicates that lumped models can be constructed for arbitrary bending elements which retain the full rate of convergence of consistent mass.

REFERENCES.

1. Bogner, F. K., Fox, R. L., and Schmidt, L. A., "The Generation of Inter-element Compatible Stiffness and Mass Matrices by the Use of Interpolation Formulas," Proc. Conf. on Matrix Methods Struct. Mech., AFFDL-TR-66-80, 397-443, Wright-Patterson AFB, Ohio (1965).

2. Ciarlet, P. G., "Quelques Methodes d'Elements Finis pour le Probleme d'une Plaque Encastree," in Computing Methods in Applied Sciences and Engineering, Part 1, 156-176, R. Glowinski and J. L. Lions, editors, Springer-Verlag, Berlin-Heidelberg - New York (1974).

3. Ciarlet, P. G. and Raviart, P.-A., "General Lagrange and Hermite Interpolation in R^n with Applications to Finite Element Methods," Arch. Rational Mech. Anal., 46, 177-199 (1972).

4. Clough, R. W., and Tocher, J. L., "Finite Element Stiffness Matrices for Analysis of Plate Bending," Proc. Conf. Matrix Methods Struct. Mech., AFFDL-TR-66-80,515-546, Wright-Patterson AFB, Ohio (1965).

5. Fraeijs de Veubeke, B. M., "Bending and Stretching of Plates-Special Models for Upper and Lower Bounds," Proc. Conf. Matrix Methods Struct. Mech., AFFDL-TR-66-80, 863-886, Wright-Patterson AFB, Ohio (1965).

6. Fried, I., "Numerical Integration in the Finite Element Method," Computers and Structures, 4, 921-932 (1974).

7. Fried, I. and Malkus, D. S., "Finite Element Mass Matrix Lumping by Numerical Integration With No Convergence Rate Loss," Int. J. Solids Struct.

8. Gurtin, M. E., "Variational Principles for Linear Elastodynamics," Arch. Rational Mech. Anal., 16, No. 1, 34-50 (1964).

9. Hughes, T. J. R., "Reduction Scheme For Some Structural Eigenvalue Problems by a Variational Theorem," to appear in Int. J. Num. Meth. Engng.

10. Lanczos, C., The Variational Principles of Mechanics, third edition, University of Toronto Press, Toronto (1966).

11. Oden, J. T., and Reddy, J. N., "On Dual Complementary Variational Principles in Mathematical Physics," Int. J. Engng Sci., 12, 1-29 (1974).

12. Pars, L. A., A Treatise on Analytical Dynamics, Wiley, New York (1965).

13. Strang, G. and Fix, G. J., An Analysis of the Finite Element Method, Prentice-Hall, Englewood Cliffs, N.J. (1973).

14. Zienkiewicz, O. C., The Finite Element Method in Engineering Science, McGraw-Hill, London (1971).

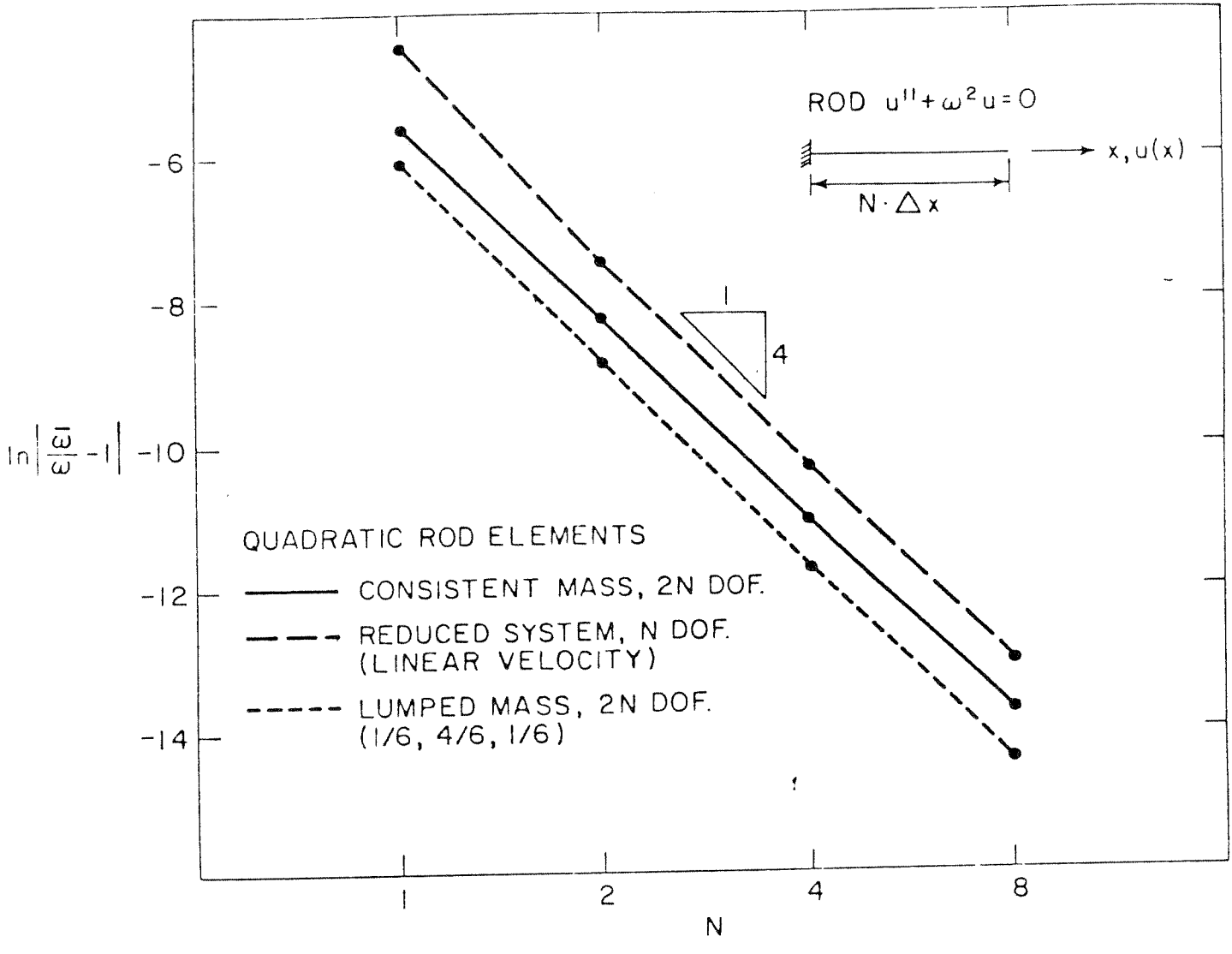

Fig. 1 Convergence rates for the fundamental frequency of a fixed-free rod

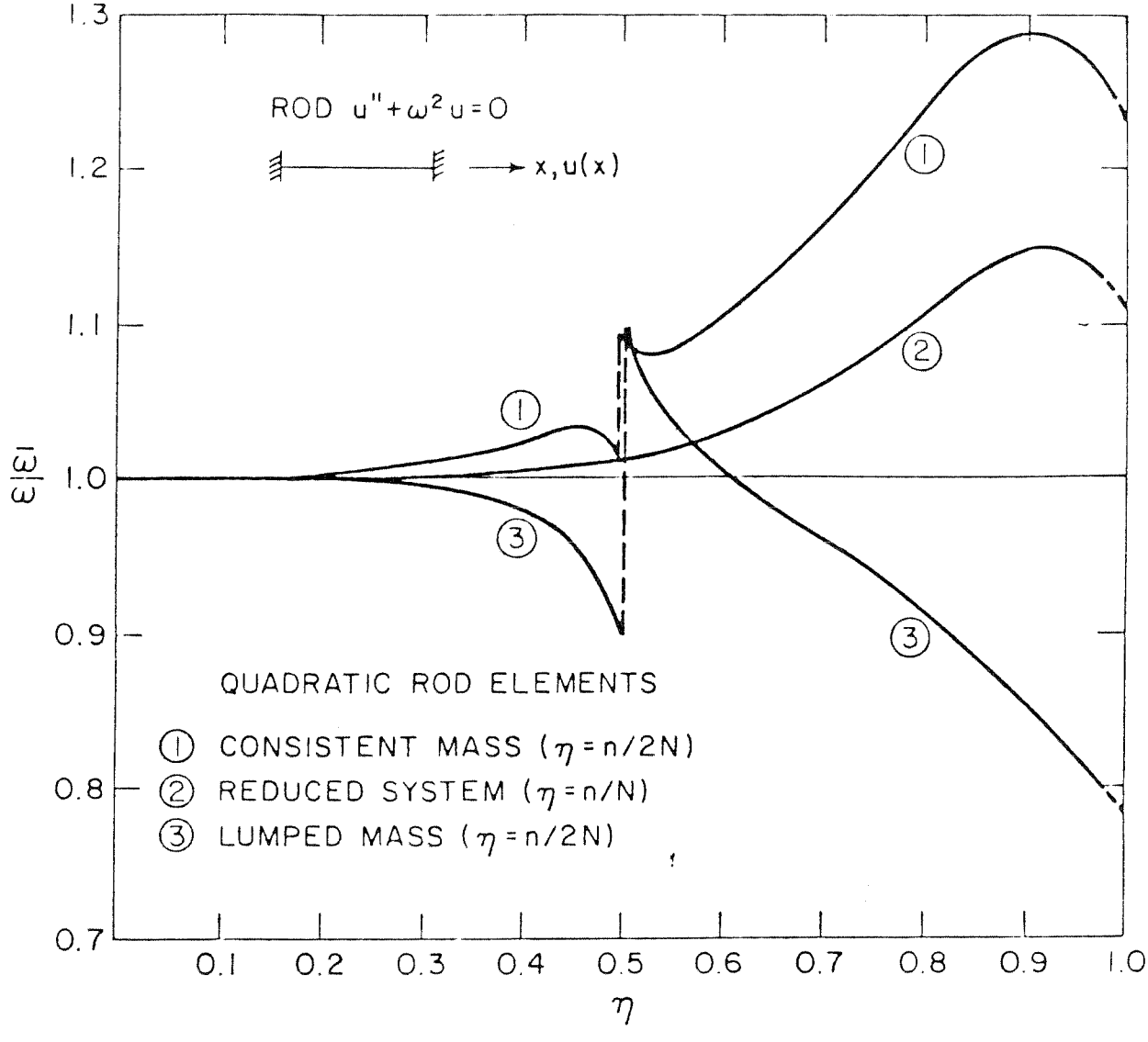

Fig. 2 Spectra for a fixed-fixed rod

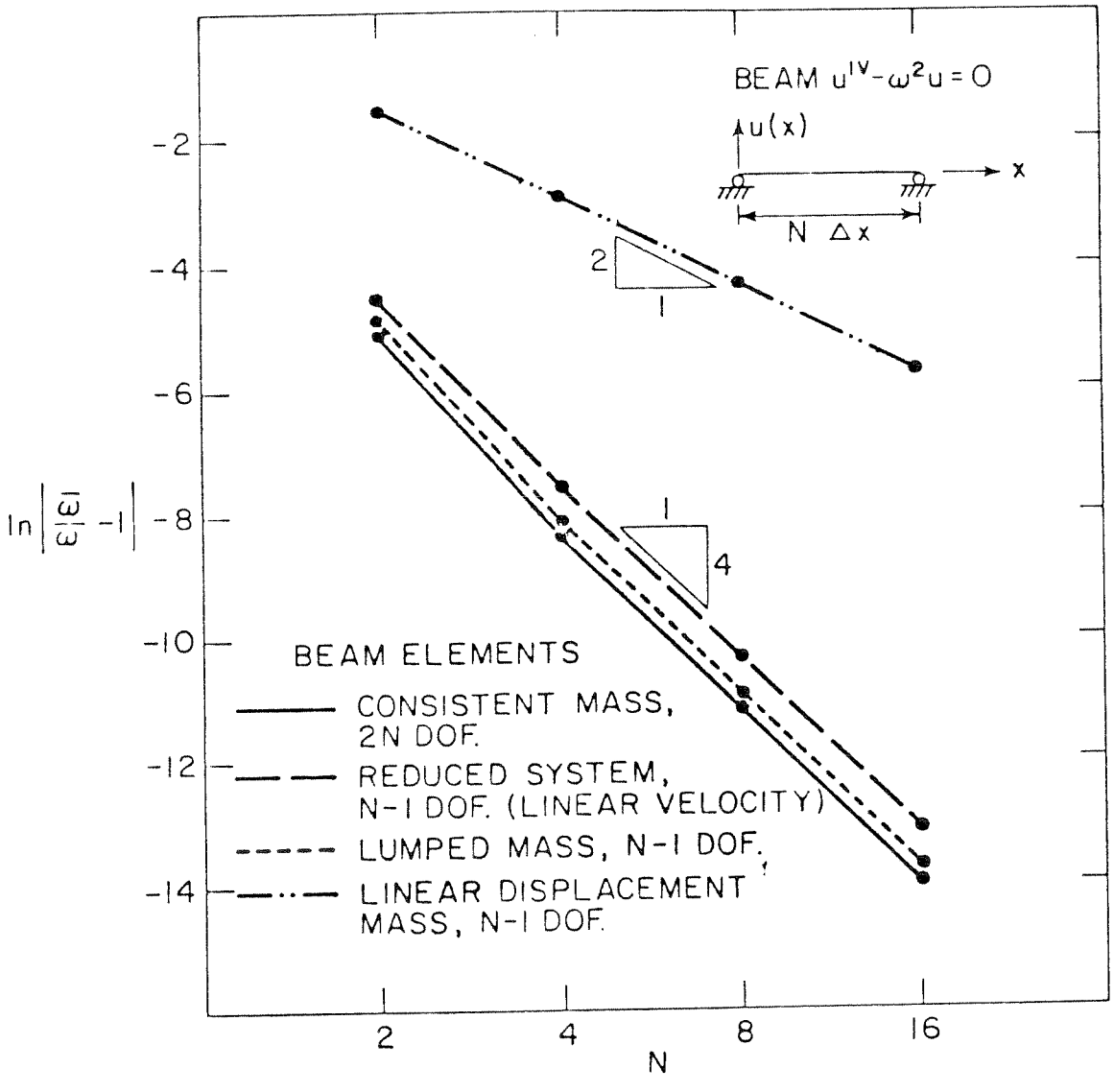

Fig. 3 Convergence rates for the fundamental frequency of a simply-supported beam

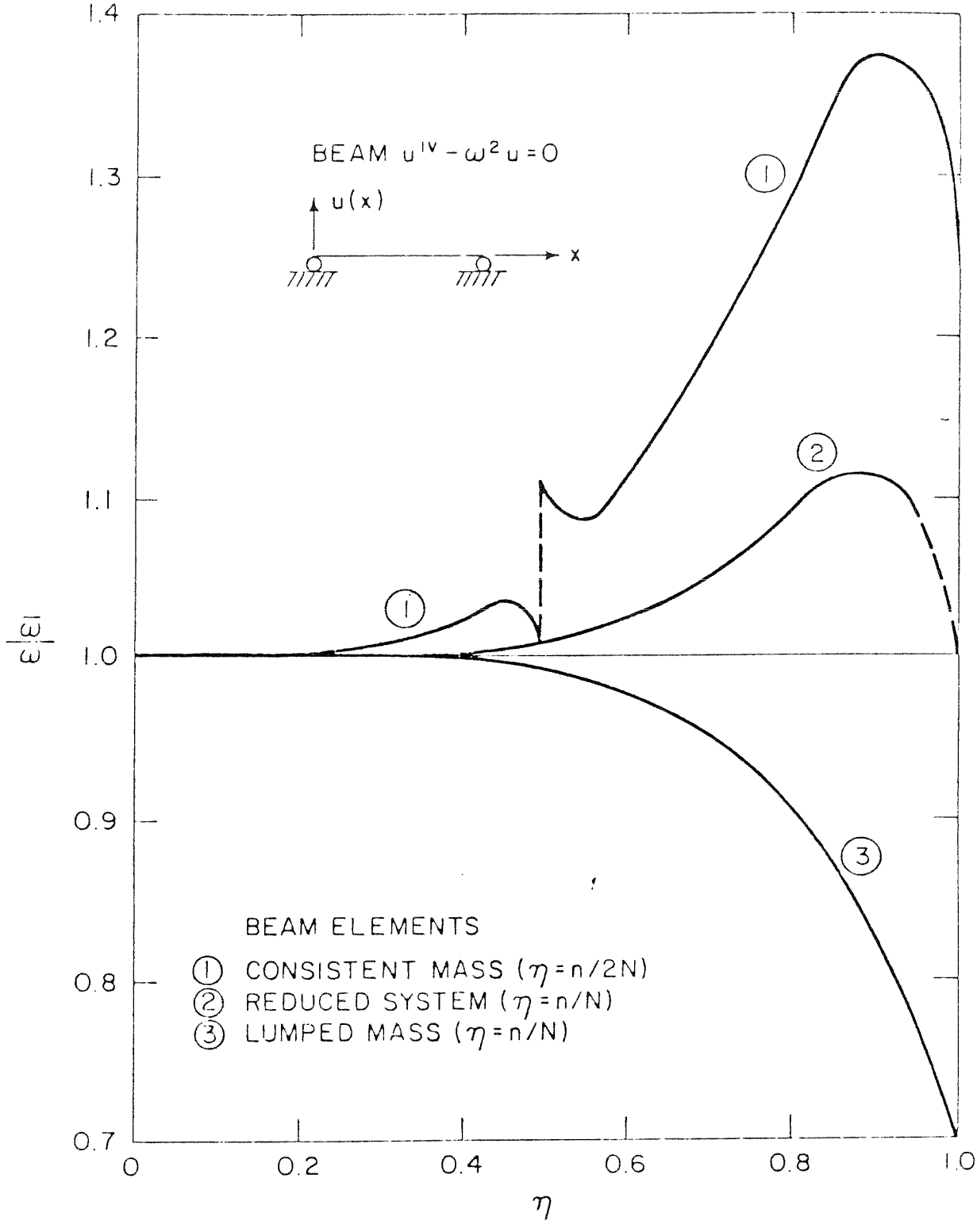

Fig. 4 Spectra for a simply-supported beam

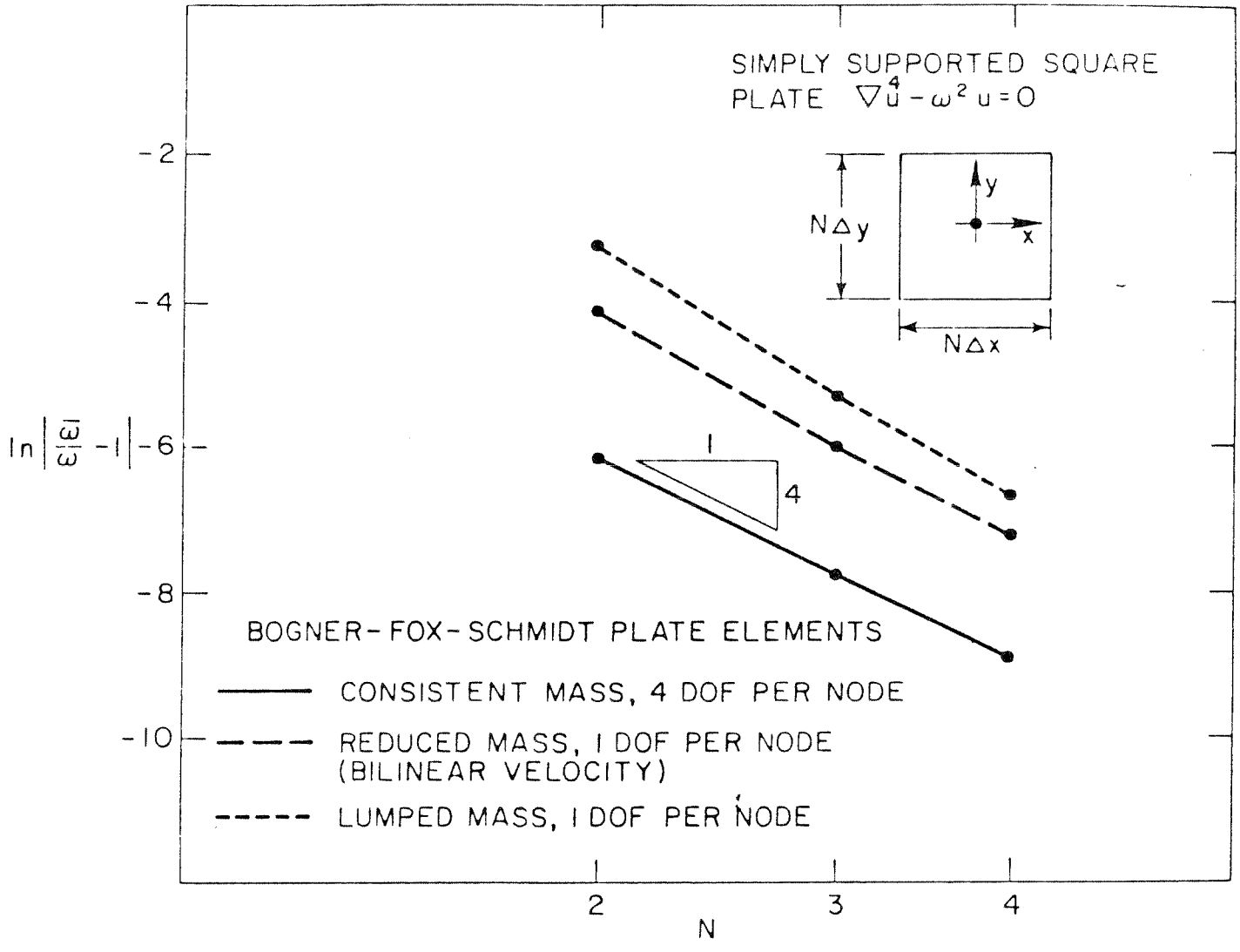

Fig. 5 Convergence rates for the fundamental frequency of a simply-supported plate

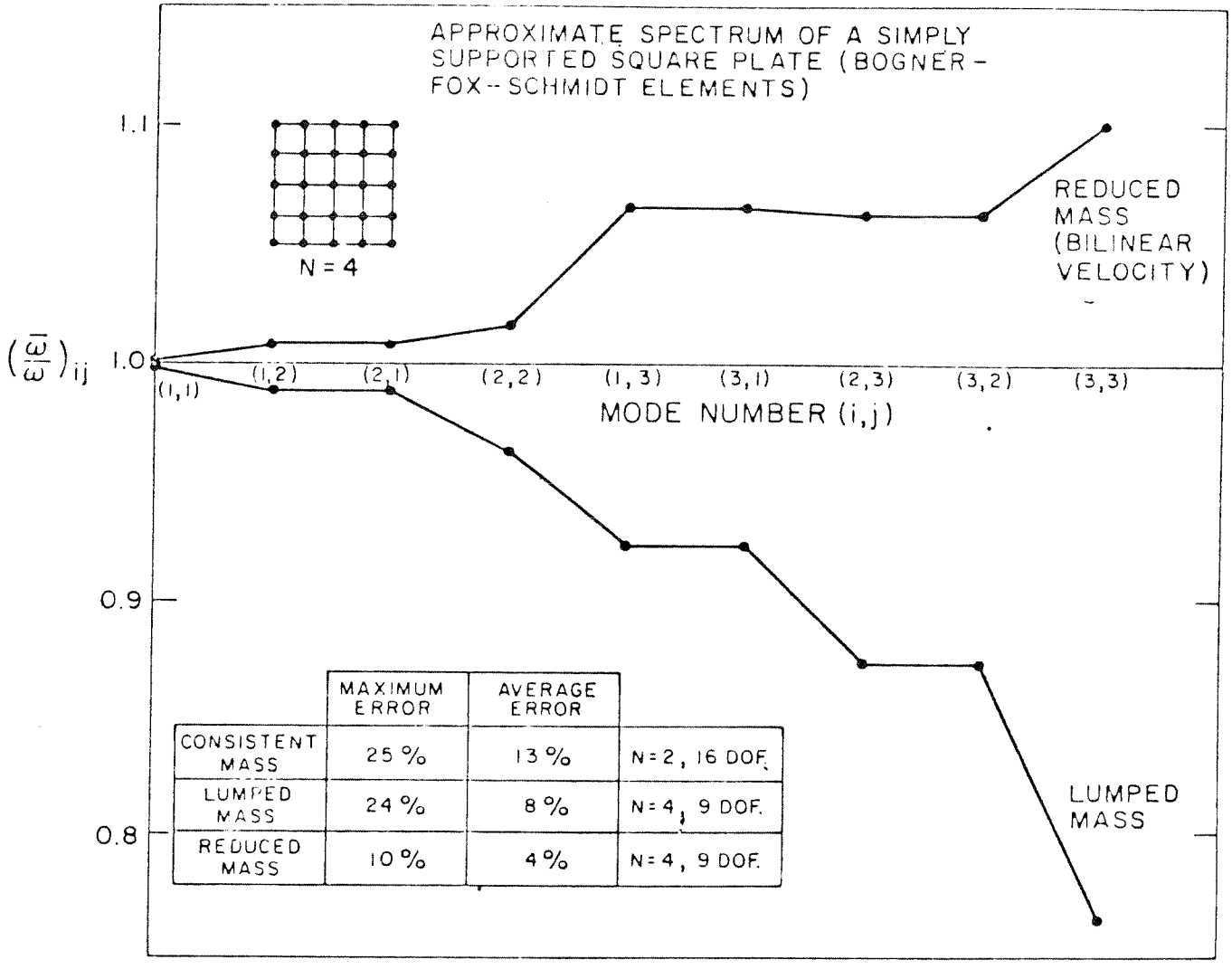

Fig. 6 Spectra for a simply-supported plate