REDUCED FINITE ELEMENT SYSTEMS IN DYNAMICS WHICH RETAIN FULL RATE OF CONVERGENCE

T.J.R. HUGHES, R.L. TAYLOR, H. HILBER
Division of Structural Engineering and Structural Mechanics,
Department of Civil Engineering, University of California, Berkeley, California 94720, U.S.A.

SUMMARY

Engineers have discussed for some time the relative merits of various descriptions of the inertial terms in finite element equations (e.g., lumped vs. consistent mass matrices) as well as the advisability of reducing the number of equations in eigenvalue and initial values problems. In addition, there has been an "engineering intuition" that rotatory degrees of freedom, which arise in beam, plate and shell finite element models, were "inefficient" in dynamics. These points are of extreme importance in the practical application of finite element methods to real engineering problems as they are intimately affiliated with the basic issue of accuracy attainable for a fixed expenditure. Recent analytical work by I. Fried ("Numerical Integration in the Finite Element Method", Computers and Structures, 4(1974), 921), among others, has shed considerable light on this subject. He has obtained results, based upon a theorem for numerical integration techniques, which indicates in what ways one may modify the mass matrix (among other things) and still retain order of convergence. This result augers a considerable change in attitude towards what is considered an appropriate (if not "consistent") matricial representation of the inertial properties of finite elements.

In our paper we merge the developments above with a method recently proposed by the first author for reducing the size of finite element eigenvalue problems. Briefly, this technique is based upon a variational theorem in which it is admissible to describe the inertial properties of continua by way of independent displacement, velocity and momentum fields. By way of this theorem we are able to design, in the present paper, finite element systems of reduced size, for the entire spectrum of dynamic problems in solid continua, *which retain the full order 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.*

To illustrate these developments, we examine numerically the spectral properties of structural models designed by various discretization techniques, e.g., consistent mass, lumped mass, the present approach, various ad-hoc representations, etc. Computational considerations, as well as some other numerical applications, are also discussed.

1. Introduction.

In this paper we investigate a finite element method for dynamics problems in which there exists some flexibility in the definition of the mass matrix. The technique emanates from variational theorems in which the displacement, velocity and momentum fields are taken to be independent. In Section 2 we present some variational theorems of this sort for application to linear elastodynamics. Based upon specific forms of these theorems, in which the momentum field is eliminated, we set up the spatially discretized (finite element) equations in Section 3. If the velocity interpolation is of lower order than the displacement interpolation, then the finite element equations may be combined to form "reduced systems" involving fewer degrees of freedom. In Section 4 we perform an error analysis of the scheme for eigenvalue problems. Our 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 order of the complete polynomial contained in the displacement (velocity, resp.) interpolation assumption.

Although we have not included the analyses here, we have obtained the same error estimate for a wide class of semi-discrete formulations of hyperbolic and parabolic problems.

These results 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.

The motivation for investigating schemes of this sort is that either they may possess some special features that are exploitable in practical circumstances or that they may engender some economic advantage in engineering problems. In this regard we wish to point out that there is some numerical evidence which indicates that the overall spectrum of the reduced system is better behaved than that for consistent mass. This will be further discussed at the conference presentation of the paper, at which time the results of numerical experiments will also be presented. Another interesting property is that the spectral radius of the reduced system is, in general, a small fraction of that of the consistent mass system. If one contemplates solving an initial-value problem by way of a conditionally stable ordinary differential equation algorithm, then much larger time steps can be employed in conjunction with the reduced system. It may be possible to save some computational labor because of this.

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 paper 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, $dx = dx_1\, dx_2\, dx_3$ 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 coördinates (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:

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

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

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

where $i = \sqrt{-1}$ and ω is the circular frequency, is stationary, i.e.,

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

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

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

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

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

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)

$$I = \int_\Omega \{ \frac{1}{2} \rho v_i v_i - i\omega\, \rho v_i u_i + \frac{1}{2} c_{ijk\ell}\, u_{i,j} u_{k,\ell}\}\, dx, \tag{8}$$

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*

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

where $\underset{\sim}{u}_e$ and $\underset{\sim}{v}_e$ are the eth 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:

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

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:

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

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)):

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

Henceforth we shall work only with (12). $\underset{\sim}{V}$ may be eliminated from (12) in the obvious way (assuming det $\underset{\sim}{W} \neq 0$):

$$(\underset{\sim}{K} - \omega^2\, \underset{\sim}{A}^T \underset{\sim}{W}^{-1} \underset{\sim}{A})\, \underset{\sim}{U} = \underset{\sim}{0}. \tag{13}$$

---

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

The matrix $\underset{\sim}{M} = \underset{\sim}{A}^T \underset{\sim}{W}^{-1} \underset{\sim}{A}$ is the mass matrix in the present theory. Note that if the velocity degrees of freedom are not coupled from element to element, then M can be directly assembled from the element contributions $\underset{\sim}{m}_e = \underset{\sim}{a}_e^T \cdot \underset{\sim}{w}_e^{-1} \underset{\sim}{a}_e$. 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 a reduced system (i.e., a system involving fewer degrees of freedom). For example, assuming det $\underset{\sim}{K} \neq 0$, we can eliminate $\underset{\sim}{U}$:

$$(\underset{\sim}{W} - \omega^2 \underset{\sim}{A} \underset{\sim}{K}^{-1} \underset{\sim}{A}^T) \underset{\sim}{V} = \underset{\sim}{0}. \tag{14}$$

In (14), $\underset{\sim}{W}$ is banded, whereas $\underset{\sim}{A}\underset{\sim}{K}^{-1}\underset{\sim}{A}^T$ is full.

4. Error Analysis.

In this section we establish the error estimates for the reduced systems. We focus our attention here on eigenvalue problems. However, we note that we have been able to obtain similar estimates for semi-discrete Galerkin formulations of initial-boundary-value problems of hyperbolic and parabolic type. 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, $0 \leq \ell \leq k$, 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 (distributional) derivatives of order ℓ, $0 \leq \ell \leq s$, 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 $c_1 \, ||u||_m^2 \leq (Au,u) \leq c_2 ||u||_m^2$, 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

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

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

$$0 \leq \lambda_1 \leq \lambda_2 \leq \ldots,$$

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

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

and

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

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 throught 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]):

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

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 a 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 $\tilde{u}^h = u^h$.

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 $||\tilde{e}^h||_m \leq ch^{k+1} \; |\tilde{u}^h|_{\hat{k}+1}$.
Proof. Subtracting (22) from (20) we get

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

By adding and subtracting $q\tilde{u}$, 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

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

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. Strange and Fix [13]):

$$\begin{aligned} \lambda_\ell \leq \lambda_\ell^h &\leq \lambda_\ell + c \; h^{2(k+1-m)} \; \lambda_\ell^{(k+1)/m} , \\ ||u_\ell - u_\ell^h|| &\leq c \; h^{k+1} \quad \lambda_\ell^{(k+1)/2m} , \\ ||u_\ell - u_\ell^h||_m &\leq c \; h^{k+1-m} \; \lambda_\ell^{(k+1)/2m} . \end{aligned} \tag{26}$$

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

$$a(\tilde{u}^h_\ell, w^h) = \tilde{\lambda}^h_\ell \, (v^h_\ell, w^h), \tag{27}$$

$$(v^h_\ell, x^h) = (\tilde{u}^h_\ell, x^h), \tag{28}$$

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

$$\tilde{\lambda}^h_\ell = \min_{S^h_\ell \subset S^h} \; \max_{w^h \text{ in } S^h_\ell} \; \tilde{R}(w^h), \tag{29}$$

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 $\tilde{\lambda}^h_\ell \ge \lambda^h_\ell$. 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., $\tilde{\lambda}^h_\ell \ge \lambda^h_\ell \ge \lambda_\ell$ for each $\ell = 1,2,\ldots, \min(N,\hat{N})$.

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

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

$$\begin{aligned} a(\tilde{u}^h_\ell, \tilde{u}^h_p) &= \tilde{\lambda}^h_\ell \, (v^h_\ell, \tilde{u}^h_p), \\ (v^h_\ell, v^h_p) &= (\tilde{u}^h_\ell, v^h_p). \end{aligned} \tag{30}$$

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

$$\begin{aligned} a(\tilde{u}^h_p, \tilde{u}^h_\ell) &= \tilde{\lambda}^h_p \, (v^h_p, \tilde{u}^h_\ell), \\ (v^h_p, v^h_\ell) &= (\tilde{u}^h_p, v^h_\ell). \end{aligned} \tag{31}$$

Combining (30) and (31) we obtain

$$(\tilde{\lambda}^h_\ell - \tilde{\lambda}^h_p) \; (v^h_\ell, v^h_p) = 0,$$

which implies $v^h_\ell \perp v^h_p$ in L_2. Since we also have

$$a(\tilde{u}^h_\ell, \tilde{u}^h_p) = \tilde{\lambda}^h_\ell \, (v^h_\ell, v^h_p),$$

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

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

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

4.13 Remark. Proposition 4.12 teels 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 $\sigma^h_\ell = \max_{w^h \text{ in } e^h_\ell} |(\hat{\pi} w^h - w^h, w^h)|$ where e^h_ℓ is the set of all unit vectors con- in E^h_ℓ, the ℓ-dimensional subspace of S^h spanned by $u^h_1, u^h_2, \ldots, u^h_\ell$. Then $\tilde{\lambda}_\ell \le \lambda_\ell \, (1-\sigma^h_\ell)^{-1}$.

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

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

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

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

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

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

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

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

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

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

Proof. By Lemmas 4.14 and 4.15 we have immediately that

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

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

4.17 Remark. Comparing this result with $(26)_1$, we see that if $\hat{k} \ge k-m$, 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 <u>lack-of-consistency</u> $\tilde{u}_\ell^h - v_\ell^h$; viz., let $w^h = \tilde{e}_\ell^h$ in (33), then

$$||\tilde{e}_\ell^h||_m \leq c\,\{\lambda_\ell^h||\tilde{e}_\ell^h|| + |\lambda_\ell^h - \tilde{\lambda}_\ell^h| + \tilde{\lambda}_\ell^h||\tilde{u}_\ell^h - v_\ell^h||\}. \tag{34}$$

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

<u>Proof.</u> 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):

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

4.20 <u>Lemma</u>. Assume that the multiplicity of λ_i is one. Then

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

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

<u>Proof</u>. 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 $||u_i^h|| = 1$, $1 \leq i \leq N$, and $||\tilde{u}_j^h|| = 1$, $1 \leq j \leq \hat{N}$. 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 <u>Lemma</u>. $||\tilde{e}_i^h|| \leq 2||u_i^h - \beta\tilde{u}_i^h||$.

<u>Proof.</u> By the triangle inequality

$$||\tilde{e}_i^h|| = ||u_i^h - \tilde{u}_i^h|| \le ||u_i^h - \beta\tilde{u}_i^h|| + ||\beta\tilde{u}_i^h - \tilde{u}_i^h|| ,$$

$$= ||u_i^h - \beta\tilde{u}_i^h|| + ||(\beta-1)\tilde{u}_i^h|| ,$$

$$\le ||u_i^h - \beta\tilde{u}_i^h|| + |\beta-1| .$$

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

$$1 = ||u_i^h|| \le ||u_i^h - \beta\tilde{u}_i^h|| + ||\beta\tilde{u}_i^h||,$$

$$= ||u_i^h - \beta\tilde{u}_i^h|| + \beta.$$

Combining this with the previous result completes the proof.□

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

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

and

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

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

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

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

4.23 Remark. Comparing these results with $(26)_3$, we see that if $\hat{k} \ge k\text{-}m$ 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 R, where R is a positive integer > 0. Then (36) and (37) still hold.

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

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

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

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

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

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

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

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

$$||u_{i+r} - \tilde{u}^h_{i+r}|| \leq ||u_{i+r} - u^h_{i+r}|| + ||u^h_{i+r} - \hat{u}^h_{i+r}||$$

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

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):

$$\begin{aligned} ||\hat{u}^h_{i+r} - v^h_{i+r}|| &= ||\sum_{j=0}^{R} \alpha_{ij}\ \hat{\pi} u^h_{j+r} - v^h_{i+r}|| , \\ &= ||\sum_{j=0}^{R} \alpha_{ij}\ (\hat{\pi} u^h - \sum_{k=0}^{R} \beta_{jk}\ v^h_{r+k})|| , \\ &\leq c \sum_{r=0}^{R} ||\tilde{u}^h_{i+r} - v^h_{i+r}||. \end{aligned}$$

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 $2n$, $n \leq m$, with smooth coefficients, then the condition for maintaining the full rate of convergence for eigenvalues and energy is that $\hat{k} \geq k-m+n$. 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.

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 $k \geq 1$, 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*.

---

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

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 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.

References.

1. Bogner, F. K., Fox, R. L., and Schmidt, L. A., "The Generation of Interelement 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-Paterson 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," in review.

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).