VARIATIONAL PRINCIPLES AND FINITE-ELEMENT MODELS

J. T. Oden, M. ASCE
This preprint has been provided for the purpose of convenient distribution of information at the Meeting. To defray, in part, the cost of printing, a meeting price of 50¢ to all registrants has been established. The post-meeting price, when ordered from ASCE headquarters will be 50¢ while the supply lasts. For bulk orders (of not less than 200 copies of one preprint) please write for prices.

No acceptance or endorsement by the American Society of Civil Engineers is implied; the Society is not responsible for any statement made or opinion expressed in its publications.

Reprints may be made on condition that the full title, name of author, and date of preprinting by the Society are given.

VARIATIONAL PRINCIPLES AND FINITE ELEMENT MODELS

by J. T. Oden,* M. ASCE

1. INTRODUCTION

This paper contains an account of the finite element method viewed as a technique for generating approximations to weak solutions of boundary-value problems. The principal aim here is unification and generalization of the theory behind the method, and the establishment of the relationship of the method to the older averaging techniques of Ritz and Galerkin. As such, the discussion is largely expository and, in fact, tutorial. Applications to various problems in continuum mechanics are also discussed.

2. WEAK SOLUTIONS OF BOUNDARY-VALUE PROBLEMS

The finite-element method belongs to a large class of approximate methods designed to produce approximations to weak solutions of boundary- and initial-value problems of mathematical physics.

2.1 Weak and Strong Solutions

To make clear the distinction between weak and strong solutions, consider, for example, the space $C^p(R)$ of functions with continuous partial derivatives of order $m$ everywhere in a bounded, open region $R$ of $n$-dimensional euclidean space $E^n$. Let $u(x)$ be a typical element of this space, $x = (x_1, x_2, \ldots, x_n)$ being a point in $R$. If $f(x)$ is a prescribed function of sufficient differentiability on $R$, and if $L$ is a linear partial-differential operator, we consider the problem of finding those functions $u(x)$ for which

$$Lu = f$$

at every $x \in R$. To make possible the existence of unique solutions to (2.1), we must, of course, also impose conditions on $u(x)$ and various of its derivatives at points $x$ on the boundary (closure) $\partial R$ of the form

*Professor and Chairman of Engineering Mechanics, University of Alabama in Huntsville.
where $A$ is also a linear operator. The problem of finding functions $u$ which simultaneously satisfy (2.1) and (2.2) is a linear boundary-value problem. The functions $u^*(\omega)$ which satisfy (2.1) and (2.2) at every $\omega$ in $R$ and $\partial R$ are called strong solutions of the boundary-value problem.

Now, in general, we can expand the class of functions in which we seek solutions to boundary-value problems by regarding $u(\omega)$ as an element of a Hilbert space $H$. In other words, we can define for any pair of functions $u, v \in H$, a real number, denoted $(u, v)$, that satisfies all of the rules required of inner-products. For example,

\[ (u, v) = \int_{\mathbb{R}} u(v) \, dv \]  

wherein Lebesque integration is implied. The associated norm of a function $u \in H$ is then

\[ \|u\|^2 = (u, u) < \infty \]  

and the resulting space (generally denoted $L^2_0(\mathbb{R})$) is complete in the norm (2.4).

Let $h(\omega)$ denote an arbitrary element in $H$ with support in $R$ (i.e., the closure of the set of points on which $h(\omega) \neq 0$ is contained in $R$). Then an element $u(\omega)$ is a weak (or "generalized") solution of the boundary-value problem (2.1) if and only if

\[ \langle u, h \rangle = \langle f, h \rangle \]  

The class of all weak solutions of the boundary-value problem is often much larger than that of the strong solutions, since (2.5) requires only that the integral of $h\Delta u$ be the same as $fh$.

2.2 Dual and Conjugate Spaces.

The notion of weak solutions of boundary-value problems can be put in a different setting by introducing the notion of dual spaces. Let $U$ denote a linear vector space, the elements of which may be regarded as functions $u(x)$ of a certain type (say, integrable), defined on the region $R \subset \mathbb{R}^n$. Let $V$ denote another linear vector space defined over the same field, and suppose that there exists a mapping $s: U \otimes V \to \mathbb{R}$ (i.e., a mapping of ordered pairs $[u, v]$ of vectors into the real numbers) such that

\begin{align*}
(1) & \quad (u, \alpha v + \beta w) = \alpha (u, v) + \beta (u, w) \\
(2) & \quad (\alpha u + \beta v, w) = \alpha (u, w) + \beta (v, w) \\
(3) & \quad (u, v^*) = 0, \text{ for fixed } v^* \text{ and all } u = 0 \\
(4) & \quad (u^*, v) = 0, \text{ for fixed } u^* \text{ and all } v = 0
\end{align*}

where $\alpha$ and $\beta$ are scalars and $(u, v)$ is the real number associated with the pair of vectors $u$ and $v$. Then $V$ is called the dual space of $U$, and the mapping $s[u, v] = (u, v)$ is called the scalar product of $u$ and $v$.

A linear mapping $t$ of $U$ into $V$ is called a linear functional, and the set $V^*$ of all linear functionals on $U$ is itself a linear vector space called the conjugate space of $U$. In 1934, H. B. Fesh showed that for every continuous linear functional $t$ on a real space $U$ there exists an element $v \in V$ such that $t(u) = (u, v)$ and the vector $v$ is uniquely determined by $t$. As a consequence, we can generally treat the dual space $V$ as algebraically the same (isomorphic to) the conjugate space $U^*$.

Now let $U'$ denote a subspace of $U$. The set of elements in $V$ which have the property that

\[ (u, v) = 0 \]  

is called the orthogonal complement of $U$ and is denoted $U^\perp$. In (2.7), $v$ is said to be orthogonal to $u$. Generally, if $U$ is the direct sum of two subspaces, $U = U_1 \oplus U_2$, $V$ is the direct sum $V^\perp \oplus V^\perp$, where $U^\perp_1$ and $U^\perp_2$ are the orthogonal complements of $U_1$ and $U_2$, respectively. Then it can be shown to follow that $U^\perp_1$, $U^\perp_2$ and $U^\perp_1$, $U^\perp_2$ are dual pairs.
To now interpret the idea of weak solutions in the context of dual spaces, let $\mathcal{L}: \mathcal{H} \to \mathcal{U}$ denote a linear mapping of a function space $\mathcal{H}$ into $\mathcal{U}$.

Let $\Delta u$ denote the image of $u$ under the mapping and let $f$ denote a fixed element in $\mathcal{U}$. The function

$$ r = \Delta u - f $$

is called the residual of $\Delta u$ with respect to $f$. Clearly, the residual $r = r(u)$ belongs to $\mathcal{U}$ since $\mathcal{U}$ is, by hypothesis, a linear space.

An element $u \in \mathcal{H}$ is a weak solution of (2.1) if the residual of $\Delta u$ with respect to $f$ is orthogonal to the entire dual space $\mathcal{V}$; i.e., $u$ is a weak solution if and only if

$$ (r, h) = 0 $$

for all $h \in \mathcal{V}$. In other words, $u$ is a weak solution if the residual is orthogonal to the entire space $\mathcal{V}$. According to condition (iii) of (2.6), this means $r = 0$.

We can go directly from the idea of the scalar product (2.6) to the inner-product (2.3) by making use of the Riesz theorem and the isomorphism of $\mathcal{U}$ and $\mathcal{V}$ (or $\mathcal{U}^*$). Accordingly, if $\xi(u)$ is a linear functional on $\mathcal{U}$, then there exists an element $\omega \in \mathcal{H}$ such that $\xi(u) = (\omega, u)$. Likewise, every linear functional on $\mathcal{U}$ can be written $\xi(u) = (\omega, u)$, where $\omega \in \mathcal{H}^*$. Hence $(\omega, u) = (\omega, u)$. Now let $\kappa: \mathcal{H} \to \mathcal{H}^*$ denote a one-to-one and onto mapping of $\mathcal{H}$ onto $\mathcal{H}^*$ such that $(\omega, u) = (\omega, u)$. Then the bilinear form $(\kappa^* u, \cdot)$ is an inner product on $\mathcal{H}^*$ since it is symmetric and positive definite.

2.3 Green's Theorem.

The idea of weak solutions can, in certain cases, be expressed in a more convenient form in terms of the adjoint of the operator $\mathcal{L}$. Let $\mathcal{H}$ and $\mathcal{V}$ denote two inner-product spaces and suppose $\mathcal{L}$ denotes a linear mapping of $\mathcal{H}$ into $\mathcal{V}$. If $\mathcal{H}^*$ and $\mathcal{V}^*$ are the respective dual spaces of $\mathcal{H}$ and $\mathcal{V}$, the mapping $\mathcal{L}$ induces a dual mapping $\mathcal{L}^*: \mathcal{H}^* \to \mathcal{V}^*$, related to $\mathcal{L}$ according to

$$ (\mathcal{L}^* u, v) = (u, \mathcal{L} v) $$

wherein $v^* \in \mathcal{V}^*$ and $u \in \mathcal{H}$. The operator $\mathcal{L}^*$ is called the adjoint of $\mathcal{L}$. In most cases, $\mathcal{H}$ and $\mathcal{H}^*$ are either isomorphic to one another or they coincide, and $\mathcal{L}^*$ is a mapping of $\mathcal{H}$ into $\mathcal{H}$.

The class of problems in which $\mathcal{U}$ is a Hilbert space, the elements of which are functions $u(\mathcal{X})$, $\mathcal{X} \in \mathcal{X}$, is of vital importance. Suppose that the inner product on $\mathcal{U}$ is defined as in (2.3), and let $\mathcal{L}$ be a linear operator on $\mathcal{U}$. Then the operator $\mathcal{L}^*$ for which

$$ v\mathcal{L} u - u\mathcal{L} v = b(u, v) $$

is called the formal adjoint of $\mathcal{L}$, where $b(u, v)$ is a bilinear form in $u$ and $v$ called the bilinear concomitant. Generally $b(u, v)$ is the divergence of another bilinear form $\mathcal{J}(u, v)$, so that we may define

$$ \Gamma(u, v) = \int b(u, v) dR = \int \text{div} \mathcal{J}(u, v) dR = \oint \mathcal{J}(u, v) dS $$

Then, with the inner product given by (2.3),

$$ (v, \mathcal{L} u) - (u, \mathcal{L} v) = \Gamma(u, v) $$

This result is called Green's Theorem. It assumes a more recognizable form if we examine specific examples of $\mathcal{L}$. For example, if $\mathcal{L}u = \varphi u$, (2.12) becomes

$$ \int (v \varphi u - u \varphi v) dR = \oint (\varphi \frac{\partial u}{\partial n} - u \frac{\partial \varphi}{\partial n}) dS $$

If $\mathcal{L}u = \varphi^2 u$, (2.12) yields

$$ \int (v \varphi^2 u - u \varphi^2 v) dR = \oint \varphi^2 \frac{\partial u}{\partial n} u - \varphi \frac{\partial \varphi}{\partial n} - \varphi^2 \frac{\partial \varphi u}{\partial n} dS $$

etc.

We are frequently faced with cases in which the conditions on $\mathcal{H}$ are such that the bilinear functional $\Gamma(u, v)$ in (2.12) vanishes. Then $\mathcal{L}^*$ is the
adjoint of \( \mathcal{L} \) and
\[
\langle v, \mathcal{L}_u \rangle = \langle u, \mathcal{L}v \rangle
\tag{2.14}
\]
Accordingly, if \( u \) is a weak solution to (2.5), we also have
\[
\langle u, \mathcal{L}h \rangle = \langle f, h \rangle
\]

We also encounter frequently boundary- and initial-value problems of the type \( M u = f \) where \( M \) is a separable operator of the form \( M = \mathcal{L}^* \mathcal{L} \). An example of such an operator is that encountered in the transverse vibration of elastic plates; i.e.
\[
M u = -\frac{\partial^2 u}{\partial t^2} + \mathcal{L}^* \mathcal{L} u = f(x,t)
\]
which is a composition of the linear operators, \( \mathcal{L} = \partial^2 / \partial t^2 - \mathcal{L}^* \mathcal{L} u \). The adjoint of \( \mathcal{L} \) is clearly \( \mathcal{L}^* = (\mathcal{L}^*)^* \mathcal{L} \) so that (2.14) now assumes the form \( \langle M u, h \rangle = \langle \mathcal{L}(\mathcal{L}^* u), h \rangle = \langle \mathcal{L}, \mathcal{L}^* u \rangle \). We then have, instead of (2.5), \( \langle \mathcal{L}, \mathcal{L}^* u \rangle = \langle f, h \rangle \).

### 2.4 Variational Formulations

An alternate and rewarding way to view the idea of weak solutions of boundary-value problems is in the context of variational principles. To review quickly some of the important features of variational methods, let \( \mathcal{D} \) denote an operator mapping a Hilbert space \( \mathcal{H} \) into \( \mathcal{H}^* \), \( \mathcal{D} \) being not necessarily linear. If \( \mathcal{D} \) is a scalar and \( h \) is an arbitrary element of \( h \), the Gateaux differential of \( \mathcal{D} \) at \( u \) is the function \( \mathcal{D}_h(u, h) \) such that
\[
\lim_{\alpha \to 0} \frac{1}{\alpha} \| \mathcal{D}(u + \alpha h) - \mathcal{D}(u) - \mathcal{D}_h(u, h) \| = 0 \tag{2.15}
\]
It is meaningful to refer to \( \mathcal{D}_h(u, h) \) as the Gateaux derivative of \( \mathcal{D} \) in the "direction" \( h \), or to the operator \( \mathcal{D}_h(u) \) on \( h \) as the Gateaux derivative of \( \mathcal{D} \) at \( u \). If \( \mathcal{H} = \mathbb{R}^n \), the real numbers, \( \mathcal{D}(u) \) is the ordinary derivative of a real-valued function \( \mathcal{D}(u) = p(x) \) (i.e., \( \mathcal{D}_h(u, h) = \frac{dp}{dx}(x) \)). If \( \mathcal{D} : \mathbb{R}^n \to \mathbb{R} \) and \( h = (1,0,0,...,0) \), then \( \mathcal{D}_h(u, h) = \mathcal{D} x \); etc. However, in more general settings the Gateaux differential of a continuous operator need not be continuous and it need not be linear in \( h \). If, on the other hand, \( \mathcal{D}_h(u, h) \) exists in some neighborhood \( \| u - u_0 \| < r \) of \( u_0 \), is continuous in \( u \) in this neighborhood, and if it is continuous in \( h \) at the zero element \( h = 0 \), then \( \mathcal{D}_h(u, h) \) is linear in \( h \). We shall henceforth assume that \( \mathcal{D}_h(u, h) \) exists and is continuous in \( u \) and \( h \) and is linear in \( h \).

A functional \( K : \mathcal{H} \to \mathbb{R} \) takes elements of the space \( \mathcal{H} \) into real numbers. If a given functional is Gateaux differentiable at \( u \), we can compute
\[
\mathcal{D}_K(u, h) = \lim_{\alpha \to 0} \frac{1}{\alpha} [ K(u + \alpha h) - K(u) ] \tag{2.16}
\]

Now \( \mathcal{D}_K(u, h) \) is, for each \( u \), a linear functional on \( \mathcal{H} \) which, according to our earlier observations, can be written in the form \( \langle \mathcal{D}(u), h \rangle \), \( \mathcal{D}(u) \) being a possibly nonlinear mapping from \( \mathcal{H} \) into \( \mathcal{H}^* \) which is equivalent to the Gateaux derivative of \( K(u) \). Consequently, the operator \( \mathcal{D}(u) \) given by
\[
\langle \mathcal{D}(u), h \rangle = \frac{1}{\alpha} \frac{d}{d\alpha} K(u + \alpha h) \bigg|_{\alpha=0} \tag{2.17}
\]
is called the gradient of the functional \( K(u) \), and we write \( \mathcal{D}(u) = \text{grad } K(u) \).

If at a particular point \( u_0 \), \( \text{grad } K(u_0) = 0 \), then \( u_0 \) is called a critical point of \( K(u) \) and we say that \( K(u) \) assumes a stationary value at \( u_0 \).

If, for a given \( \mathcal{D}(u) \), there exists a functional \( K(u) \) such that \( \mathcal{D}(u) = \text{grad } K(u) \), then \( \mathcal{D}(u) \) is referred to as a potential operator. If \( \mathcal{D}(u) \) is continuous and has a linear Gateaux differential \( \mathcal{D}_h(u, h) \) at every \( u \in \mathcal{B} \subset \mathcal{H} \), then a necessary and sufficient condition that \( \mathcal{D}(u) \) be potential on \( \mathcal{H} \) is that it be symmetric; i.e.
\[
\langle \mathcal{D}_h(u, k), h \rangle = \langle \mathcal{D}_h(u, k), h \rangle \tag{2.17}
\]

We can now appreciate the concept of variational formulations of boundary-value problems of mathematical physics. Take, for example, the case of the boundary-value problem
\[
\mathcal{D}(u) = \mathcal{L} - f = 0 \tag{2.18}
\]
wherein \( f \) is fixed and \( \mathcal{D}(u) \) is a potential operator. By definition, there exists a functional \( K(u) \) for which \( \mathcal{D}(u) = \text{grad } K(u) \). Indeed \( \mathcal{D}_K(u, h) = \langle \mathcal{D}(u), h \rangle = \langle \mathcal{L} - f, h \rangle \). If \( u_0 \) is a critical point of \( K(u) \), then
that critical points of the functional $K(u)$ which has the property $\nabla K(u)$ are weak solutions of the problem $\Theta(u) = 0$. This, in fact, is the essence of the variational method: to obtain weak solutions of boundary-value problems by determining critical points of an associated functional.

It is clear that the inverse problem of the calculus of variations, (i.e., given a potential operator $\Theta(u)$, find a functional $K(u)$ such that $\nabla K(u) = \Theta(u)$) is of crucial importance in applying the variational formulation.

This problem is handled by an important theorem due to Vainberg:

**Theorem:** Let $\Theta(u)$ be potential on $\mathcal{U}$. Then, there exists a unique functional whose value at $u_0$ is $K_0$ and whose gradient is $\Theta(u)$ which is given by

$$K(u) = \int_0^1 \langle \Theta(u) + s(u - u_0), u - u_0 \rangle ds + K_0$$

(2.20)

where $s$ is a real parameter.

Thus, so long as $\Theta(u)$ is potential (i.e., so long as $\Theta(u)$ satisfies (2.17)), we can construct the functional associated with the variational statement of the problem $\Theta(u) = 0$ by simply introducing $\Theta(u)$ into (2.20) and performing the indicated integration.

The question arises as to whether or not the variational method can be used in cases in which $\Theta(u)$ is not potential? Fortunately, the answer to this question is affirmative, and it involves the concept of imbedding the given problem in a functional setting different from that implied by (2.18).

To be more specific, suppose that $\Theta(u)$ is not potential and that $u$ is not necessarily a strong solution of (2.18). Then, as in (2.9), we may construct the residual $r(u) = \Theta(u) - \mathcal{L}u - f$. Now let $J(v)$ be a functional on $\mathcal{U}^*$ which has the following properties:

1. $J(v)$ has a continuous, linear Gateaux differential in a neighborhood $\mathcal{B}$ of $v = 0$ ($0$ being the null element of $\mathcal{B}$).
2. $\nabla J(0) = 0$
3. $J(0 + h) \geq J(0)$ for every $h \in \mathcal{B}$.

Then $J(v)$ not only assumes a stationary value at $v = 0$, but also $J(v)$ is a minimum there (we remark that property (iii) is only introduced for convenience and need not actually hold for our imbedding techniques to be employed). Consequently, we may seek critical points of the functional $I(u) = J(r(u))$ for if at a point $u \in \mathcal{U}$, $DJ(r,h) = \langle r(u), h \rangle = 0$ then $u$ is clearly a weak solution of (2.18).

Remark: Notice that the function $u$ is not varied in computing $DJ(r,h)$.

Examples of functional imbeddings are sometimes surprisingly familiar. For instance, take $J_1(v) = \|v\|^2 = \langle v, v \rangle$. Then (1)-(iii) above are satisfied and setting $DJ_1(u,h) = 0$ gives the well-known idea of "least squares".

In fact, $DJ_1(r,h) = 2 \langle r, h \rangle$. Alternately, let $\mathcal{L}$ denote any self-adjoint linear, positive definite, strongly elliptic operator such that $\langle \mathcal{L}v, v \rangle \geq c\|v\|^2$, $\mathcal{L}(0) = 0$. Then we can also use the imbedding $J_\mathcal{L}(r) = \langle \mathcal{L}r, r \rangle$ since $DJ_\mathcal{L}(r,h) = 2 \langle \mathcal{L}r, h \rangle$ and $DJ_\mathcal{L}(r,h) = 0 = \mathcal{L}r = 0$ which, in turn, implies that $r = 0$.

3. AVERAGING METHODS OF APPROXIMATIONS

3.1 Galerkin's Method.

Among the most important methods for obtaining approximate solutions to (2.17) are Galerkin's method and the closely related methods of weighted residuals. In Galerkin's method, we seek approximations to solutions of (2.5) in a finite-dimensional subspace $\mathcal{B}_0$ of $\mathcal{U}$ spanned by a linearly independent set of basis functions $w_1(x), w_2(x), \ldots, w_n(x)$. Every element $w(x)$ in $\mathcal{B}_0$ is, therefore, of the form

$$w(x) = \sum_{i=1}^n c_i w_i(x)$$
where \( a_1, a_2, \ldots, a_k \) are scalars. Galerkin's method consists of selecting the coefficients \( a_i \) so that \( w(x) \) approximates a weak solution of (2.1); i.e.

\[
\text{w}(x) \text{ is a Galerkin approximation of the solution } \text{u}(x) \text{ of (2.1) if }
\]

\[
\langle \Delta \text{w}, f, w \rangle = 0
\]
for all \( w \in \mathcal{S}_G \). If \( \mathcal{L} \) is linear, (3.2) leads to a system of linear equations for the coefficients \( a_i \). To wit,

\[
\left( \sum_{i=1}^{G} a_i \varphi_i - f, \varphi_j \right) = 0
\]

for each \( j \), so that

\[
\left( \sum_{i=1}^{G} a_i \varphi_i - f, \varphi_j \right) = 0
\]

Consequently, the coefficients \( a_i \) of the Galerkin approximation \( \tilde{w}(x) \) of \( u(x) \) must satisfy

\[
\sum_{i=1}^{G} L_{ij} a_i - f_j = 0
\]

(3.4)

where \( L_{ij} = \langle \Delta \varphi_i, \varphi_j \rangle \) and \( f_j = \langle f, \varphi_j \rangle \). If \( \mathcal{L} \) is nonlinear, the Galerkin approximation leads to a system of nonlinear equations in \( a_i \).

For \( L_0 \)-approximations, it is easy to show that the Galerkin approximation \( \tilde{w} \) is the best \( L_0 \)-approximation of the solution \( u^* \) of (3.1); i.e.

\[
\| u^* - \tilde{w} \| = \inf_{w \in \mathcal{S}_G} \| u^* - w \|
\]

Moreover, it is clear that Galerkin’s method chooses the coefficients \( a_i \) so that the residual (error)

\[
r(x) = \Delta \tilde{w} - f
\]

(3.6)
is orthogonal to (lies in the orthogonal complement of) the linear manifold \( \mathcal{S}_G \).

### 3.2 Semidiscretizations

In problems involving equations of evolution such as

\[
\begin{align*}
\frac{\partial u(x,t)}{\partial t} &= \mathcal{L}u(x,t) + \mathcal{N}(u(x,t), x) + \mathcal{Q}(u(x,t), x), \quad x \in \Omega, \\
\frac{\partial u(x,0)}{\partial t} &= 0
\end{align*}
\]

(3.7)

where \( \mathcal{L} \) is a partial differential operator involving only derivatives with respect to \( x \) (e.g., \( \mathcal{L}u(x,t) = -\nabla \cdot (\mu(x) \nabla u(x,t)) + f(x) \)), then it is often convenient to use semidiscrete Galerkin approximations of the solution of (3.7); that is, we assume that the coefficients \( a_i \) in (3.1) are continuous real-valued functions defined for all \( t \geq 0 \): \( u(x) = \sum_{i=1}^{G} a_i(t) \varphi_i(x) \). Then \( \tilde{w}(x,t) \) is the semidiscrete Galerkin approximation of \( u(x,t) \) of (3.7) if

\[
\left( \frac{\partial^2 \tilde{w}}{\partial t^2}, \varphi_j \right) - \langle \mathcal{L} \tilde{w}, \varphi_j \rangle = \langle f, \varphi_j \rangle
\]

(3.8)

Thus, the coefficients \( a_i \) are solutions of the system of ordinary differential equations

\[
\sum_{i=1}^{G} C_{ij} \frac{d^2 a_i}{dt^2} - \sum_{i=1}^{G} Q_{ij} a_i = f_j(t)
\]

(3.9)

where

\[
C_{ij} = \langle \varphi_i, \varphi_j \rangle, \quad Q_{ij} = \langle \mathcal{N}(\varphi_i, \varphi_j), \varphi_j \rangle, \quad f_j = \langle f, \varphi_j \rangle
\]

### 3.3 Weighted-Residuals

A slight generalization of Galerkin's method is obtained if the problem of determining weak solutions to (3.1) is viewed in the setting of dual or conjugate spaces as suggested by (2.6) and (3.1). Then \( \mathcal{U} \) and \( \mathcal{Y} \) (or \( \mathcal{H}^* \)) need not coincide and the function \( h \) in (2.9) need not belong to the same subspace as the approximation (3.1). Then an element \( h(x) \) in the conjugate space \( \mathcal{H}^* \) is assumed to be of the form

\[
h(x) = \sum_{i=1}^{G} b_i \varphi_i(x)
\]

(3.10)
where $\chi^1(x), \chi^2(x), \ldots, \chi^n(x)$ are $G$ linearly independent functions which provide a basis for $\mathcal{G}$ and which may be unrelated to the functions $\omega_i(x)$ of (3.1). Then, instead of (3.3), we arrive at the system

$$\left(\sum_i a_i \omega_i \right) \cdot f(x^1) = 0$$

by arguing that if $r(u) = 0$, then $(r, h) = 0$ for fixed $r(u)$ and all $h \in \mathcal{G}$ [(see (3.1)]. Simidiscretizations can, of course, be obtained in a like manner.

### 3.4 The Ritz Method

The Ritz method is obtained from the results presented thus far by simply requiring that $E$ be potential. Then, instead of (3.3), we determine the stationary value of the associated functional $K(u)$ of (2.20) restricted to the subspace $\mathcal{H}$. Further elaboration is not needed.

### 4. THE FINITE ELEMENT METHOD

#### 4.1 Introductory Remarks

Prior to the mid-1950's, approximate methods of the Ritz-Galerkin type found rather limited applications in the more difficult problem areas of mathematical physics because of the difficulty in generating appropriate basis functions $\omega_i(x)$ in (3.1). This was particularly true in problems involving complex domains and mixed boundary conditions. Moreover, the conditioning of the Ritz-Galerkin equations (3.3) is highly sensitive to the choice of the functions $\omega_i(x)$ and the considerable effort required to generate such equations for significant problems was, in the past, a serious disadvantage.

The finite-element method is a systematic technique for constructing the basis functions $\omega_i(x)$ for Ritz-Galerkin approximations (3.1) for irregular domains. Aside from a number of other advantages, the method overcomes all of the traditional disadvantages of Ritz-Galerkin procedures mentioned above:

$$R_{p+1}(u) = \frac{1}{(p+1)!} \sum_{i_1}^{n} \sum_{i_2}^{n} \cdots \sum_{i_{p+1}}^{n} \frac{\partial^{p+1} u(x + \theta y)}{\partial x_{i_1} \partial x_{i_2} \cdots \partial x_{i_{p+1}}} (y_1 - x_1) \cdots (y_{p+1} - x_{p+1}) (\theta_1 - 1)$$

The basis functions $\omega_i(x)$ are generated in a straightforward and systematic manner. Irregular domains and mixed boundary conditions are easily accommodated, the resulting equations describing the discrete model are generally well-conditioned, and the method is exceptionally well suited for implementation via electronic computers. Other advantages and properties of the method are discussed below.

#### 4.2 The Method

Consider again an open bounded domain $\Omega$ in $\mathbb{R}^n$ with closure $\partial \Omega$ and let $x = (x_1, x_2, \ldots, x_n)$ denote a point in $\Omega$. Let $C^1(\Omega)$ denote the space of functions $u(x)$ with continuous derivatives of order $k \leq m$ on $\Omega$. Taylor-type (mean-value) expansions of $u(x)$ can be written concisely if we use the standard multi-integer notation: let $\mathbb{Z}^n$ denote the set of all $n$-tuples of non-negative integers (i.e., if $\alpha \in \mathbb{Z}_+^n$, then $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n)$, $\alpha_1$ being integers $\geq 0$); then the multi-integer conventions are defined as follows:

$$\begin{align*}
|\alpha| &= \alpha_1 + \alpha_2 + \cdots + \alpha_n; \\
\alpha! &= \prod_{i=1}^{n} \alpha_i !; \\
C = \prod_{i=1}^{n} \alpha_i !; \\
\alpha' &= (\alpha_1, \alpha_2, \ldots, \alpha_n); \\
\beta' &= (\beta_1, \beta_2, \ldots, \beta_n); \\
\alpha'_{\beta'} &= \frac{\partial|\alpha|}{\partial \alpha_1 \partial \alpha_2 \cdots \partial \alpha_n} \cdot \frac{\partial|\beta|}{\partial \beta_1 \partial \beta_2 \cdots \partial \beta_n}; \\
\partial^{|\alpha|}{u(x)} &= \frac{\partial^{\alpha_1} u(x)}{\partial x_1^{\alpha_1}} \frac{\partial^{\alpha_2} u(x)}{\partial x_2^{\alpha_2}} \cdots \frac{\partial^{\alpha_n} u(x)}{\partial x_n^{\alpha_n}}.
\end{align*}$$

Then the expansion of $u(x) \in C^{p+1}(\Omega)$ about $x \in \Omega$ can be written

$$u(x + y) = \sum_{|\alpha| \leq p+1} \frac{\alpha!}{\alpha'_{\beta'} !} \partial^{|\alpha|}{u(x)} + R_{p+1}(u) \quad (4.2)$$

where $R_{p+1}(u)$ is the remainder.
We now consider a finite-element model \( \mathcal{R} \) of the region \( \mathbb{R} \) which is the union of \( e \) closed and bounded subregions \( \mathcal{R}_e \) of \( e \). The subregions \( \mathcal{R}_e \), where \( \mathcal{R}_e \) is the closure of an open region \( \mathcal{R}_e (\mathcal{R}_e = \mathcal{R}_e + \partial \mathcal{R}_e) \), are called finite elements, and the regions \( \mathcal{R}_e \) are disjoint:

\[
\mathcal{R} = \bigcup_{e=1}^{E} \mathcal{R}_e ; \quad \mathcal{R}_e \cap \mathcal{R}_f = \emptyset , \quad e \neq f \quad (4.4)
\]

Conceptually, the finite elements are considered to be connected together at a number \( G \) of nodal points (i.e., nodes, joints, knots, etc.) labeled \( \mathcal{A} \), \( \Delta = 1, 2, \ldots, G \). Locally, it is meaningful to label the nodal points belonging to element \( \mathcal{R}_e \) by \( \mathcal{A} \), \( N = 1, 2, \ldots, N_e \), \( N_e \) being the number of nodal points belonging to element \( \mathcal{R}_e \). (For simplicity, we shall henceforth assume that the global and local coordinate systems coincide, thereby avoiding the necessity of introducing a coordinate transformation for each element). Then, assuming the nodal compatibility conditions are satisfied (i.e., there exists a one-to-one correspondence between all nodal points \( \mathcal{A} \) in \( \mathcal{R} \) and points \( \mathcal{A} \) in the connected model \( \mathcal{R} \), see [1]), the connectivity and decomposition of the model are established by the respective mappings:

\[
\mathcal{A} = \sum_{N=1}^{N_e} \omega_{N}^{A} \mathcal{A}_N \quad \text{e fixed} \quad \mathcal{A} = \sum_{N=1}^{N_e} \omega_{N}^{A} \mathcal{A}_N \quad (4.5)
\]

where

\[
\omega_{N}^{A} = \begin{cases} 1 & \text{if node } \Delta \text{ of the connected model } \mathcal{R} \text{ is coincident with node } \mathcal{N} \text{ of element } \mathcal{R}_e \\ 0 & \text{if otherwise} \end{cases} \quad (4.6)
\]

and \( \omega_{N}^{A} \) is simply the transpose of \( \omega_{N}^{A} \). It is clear that the mappings (4.6) can be used to form identity mappings through the compositions:

\[
\sum_{\Delta=1}^{G} \sum_{N=1}^{N_e} \omega_{N}^{A} \delta^{\mathcal{A}_N} \mathcal{A} = \delta^{\mathcal{A}} \quad \sum_{N=1}^{N_e} \omega_{N}^{A} \delta^{\mathcal{A}_N} \mathcal{A} = \delta^{\mathcal{A}} \quad (4.7a,b)
\]

where \( \delta^{\mathcal{A}} \) are Kronecker deltas. The mappings (4.6) are called incidence or Boolean mappings. The incidence mapping \( \Lambda \) of (4.6), is said to establish the connectivity of the discrete model \( \mathcal{R} \), while \( \Omega \) of (4.5) established a decomposition of \( \mathcal{R} \) into finite elements. A geometrical interpretation is given in Fig. 1.

A function \( \mathcal{U}(x) \) with domain \( \mathcal{R} \) is called a finite-element representation of order \( q \) if and only if

\[
\mathcal{U}(x) = \sum_{e=1}^{E} \mathcal{U}_e(x) \quad \mathcal{U}_e(x) = \left\{ \begin{array}{ll} 0 & \text{if } x \notin \mathcal{R}_e \\ \sum_{N} \omega_{N}^{A} \mathcal{U}_N(x) & \end{array} \right. \quad (4.8)
\]

where \( \mathcal{U}_N(x) \) are local interpolation functions corresponding to element \( \mathcal{R}_e \) which are defined so as to have the properties:

\[
D_{\mathcal{R}} \mathcal{U}_N(x) = 0 , \quad x \notin \mathcal{R}_e ; \quad \mathcal{B} \in Z^G_e \quad (4.9)
\]

where \( \delta_{\mathcal{B}}^{\mathcal{A}} \) are Kronecker deltas. Globally, we write

\[
\mathcal{U}(x) = \sum_{e=1}^{E} \sum_{N=1}^{N_e} \mathcal{A}_{N}^e \mathcal{U}_N(x) \quad (4.10)
\]

where

\[
\mathcal{A}_{N}^e = \sum_{\Delta} \omega_{N}^{A} \delta^{\mathcal{A}_N} \mathcal{A} \quad (4.11)
\]

\[
\mathcal{U}_N(x) = \sum_{e=1}^{E} \sum_{N=1}^{N_e} \omega_{N}^{A} \mathcal{U}_N(x) \quad (4.12)
\]

If \( \mathcal{U}(x) \) is a first-order representation, we need only have \( \mathcal{U}(x) = \sum_{\Delta} \mathcal{A}_{N}^e \mathcal{U}_N(x) \)

where

\[
\mathcal{O}_{\mathcal{A}}(x) = \sum_{e=1}^{E} \sum_{N=1}^{N_e} \omega_{N}^{A} \mathcal{U}^e(x) ; \quad \mathcal{U}^e(x) = \left\{ \begin{array}{ll} 0 & \text{if } x \notin \mathcal{R}_e \\ \mathcal{U}_N(x) & \end{array} \right. \quad (4.13)
\]

\[
\mathcal{A}^e = \sum_{\Delta} \omega_{N}^{A} \mathcal{A} \quad (e \text{ fixed}) ; \quad \omega_{N}^{A} = \sum_{\Delta} \omega_{N}^{A} (\mathcal{A}^e) \quad (e \text{ fixed})
\]

Then
Remark 1. Finite-element representations (4.8) and (4.13) are linear combinations of functions which have compact and almost disjoint support. (Recall that the support of \( f(x) \) is the closure of the set of points \( x \) in the domain of \( f \) such that \( f(x) \neq 0 \). If \( f(x) = 0 \) for \( |x| > \sigma \), the support of \( f(x) \) is compact, and if the support (written supp.) functions \( \psi_1(x) \) and \( \psi_2(x) \) is such that supp \( \psi_1 \cap \text{supp} \psi_2 = \emptyset \), except possibly at a finite number of points, lines, or surfaces, then \( \psi_1 \) and \( \psi_2 \) have "almost" disjoint support).

Remark 2. The properties (4.9) of the local interpolation functions \( \phi_\alpha(x) \), are preserved under the Boolean mappings \( \Omega \), i.e., the global functions \( \phi_\alpha'(x) \) have the properties

\[
\phi_\alpha'(x) \cap \phi_\alpha''(x) = \delta_\alpha.
\]

or, for first-order representations,

\[
\phi_\alpha'(x) = \delta_\alpha.
\]

Remark 3. As a result of (4.9), (4.13), and (4.14), the coefficients \( \psi_\alpha \) and \( \psi_\alpha'' \) have a special interpretation:

\[
\psi_\alpha(x) = \sum_{|q|<q-1} \psi_\alpha(x) = \delta_\alpha.
\]

4.3 Finite-Element Approximation.

It is now a simple matter to combine the concepts of Ritz-Galerkin approximation discussed earlier and the notion of finite-element representations of functions to obtain approximations of weak solutions of a wide range of boundary-value problems. Considering (4.8) as a Galerkin approximation of the solution of the boundary-value problem (2.1), i.e., associating (4.8) with (3.1), we seek coefficients \( \psi_\alpha \) which satisfy (3.2), i.e.,

\[
\left\{ \sum_{|q|<|\Delta|} \psi_\alpha(x) - f, \psi_\alpha \right\} = 0,
\]

\( \alpha \in \mathbb{Z}_+, \Delta = 1, 2, \ldots, \mathbb{G} \). If \( \mathbb{L} \) is linear, we have

\[
\sum_{|q|<|\Delta|} \psi_\alpha(x) - f, \psi_\alpha \right\} = 0,
\]

where

\[
\sum_{|q|<|\Delta|} \psi_\alpha(x) = \left\{ \psi_\alpha(x), \psi_\alpha \right\}.
\]

If \( \mathbb{L} \) is nonlinear, (4.14), of course, represents a system of nonlinear equations in the coefficients \( \psi_\alpha \).

Remark. The Fundamental Property of Finite Element Approximation. The success and utility of the finite-element concept as a method of approximation is primarily due to the following fundamental property: the finite-element approximations can be completely formulated locally, one element at a time and each element independent of the others, and the global approximation can then be obtained by simple transformations of the local equations. Conditions (4.9) and (4.12) are responsible for this property of finite-element approximations. As a result of this local character of the approximation, it is possible to design large-scale computer programs in which local approximations of a given class of boundary-value problems are automatically generated for a typical element of a certain type; then, by appropriately
connecting elements together, global models are easily generated for whatever domain and boundary conditions the analyst wishes to consider.

To demonstrate this property mathematically, we introduce (4.10) into (4.19) to obtain

\[
\sum_{\alpha} \int \sum_{m=1}^{N} \left[ E \left( \sum_{f=1}^{M} \varepsilon_{\alpha f} \varepsilon_{\alpha f} \right) - f \right] U \left( \sum_{f=1}^{M} \varepsilon_{\alpha f} \varepsilon_{\alpha f} \right) = 0
\]

Introducing (4.12), noting that

\[
\left( \varepsilon_{\alpha f}^{B}, \varepsilon_{\alpha f}^{D} \right) = 0 \quad \text{if} \quad f = \alpha
\]

and recalling that the \( \varepsilon_{\alpha f}^{D}(x) \) have almost disjoint support, we have

\[
\sum_{\alpha} \sum_{m=1}^{N} \sum_{f=1}^{M} \left( \frac{\partial}{\partial \alpha} \varepsilon_{\alpha f}^{D} \right) \delta_{\alpha f}^{B} - \left( \frac{\partial}{\partial \alpha} \varepsilon_{\alpha f}^{D} \right) \delta_{\alpha f}^{B} = 0
\]

Thus, the local finite-element model of (4.24) corresponding to element \( \alpha \)

\[
\sum_{\alpha} \sum_{m=1}^{N} \sum_{f=1}^{M} \left( \frac{\partial}{\partial \alpha} \varepsilon_{\alpha f}^{D} \right) \delta_{\alpha f}^{B} - \left( \frac{\partial}{\partial \alpha} \varepsilon_{\alpha f}^{D} \right) \delta_{\alpha f}^{B} = 0
\]

where

\[
\frac{\partial}{\partial \alpha} \varepsilon_{\alpha f}^{D} = \left( \frac{\partial}{\partial \alpha} \varepsilon_{\alpha f}^{D} \right) ; \quad \frac{\partial}{\partial \alpha} \varepsilon_{\alpha f}^{B} = \left( \frac{\partial}{\partial \alpha} \varepsilon_{\alpha f}^{B} \right)
\]

Then the global equations (4.20) are obtained by computing

\[
\sum_{\alpha} \sum_{m=1}^{N} \sum_{f=1}^{M} \left( \frac{\partial}{\partial \alpha} \varepsilon_{\alpha f}^{D} \right) \delta_{\alpha f}^{B} = \sum_{\alpha} \sum_{m=1}^{N} \sum_{f=1}^{M} \left( \frac{\partial}{\partial \alpha} \varepsilon_{\alpha f}^{D} \right) \delta_{\alpha f}^{B}
\]

5. APPLICATIONS

We shall now investigate a number of applications of the finite-element method to problems in various branches of mathematical physics.

5.1 The Equations of Continuum Mechanics.

Consider a continuous body \( B \) in motion in three-dimensional euclidean space \( \mathbb{E}^3 \). At some reference time \( t = 0 \), \( B \) occupies a reference configuration \( C_0 \subset \mathbb{E}^3 \) in which we establish a fixed reference frame \( X = (x_1, x_2, x_3) \). The coordinates of a point in \( \mathbb{E}^3 \) relative to this frame are denoted \( x_1 \) and are called the spatial coordinates of the point. The body \( B \) is a compact continuum of material particles \( \mathbb{X} \) which can be set into one-to-one correspondence with order triples \( X = (x^1, x^2, x^3) \) of real numbers called intrinsic coordinates or material coordinates of the particle \( X \). It is convenient to select the spatial position of a particle \( X \) at time \( t \) relative to the fixed spatial reference frame, the motion of \( B \) relative to \( C_0 \) is described by the one-parameter family of configurations,

\[
X = X(x, t)
\]

or, in component form,

\[
x_1 = x_1 (x^1, x^2, x^3, t)
\]

The displacement \( u(X, t) \), velocity \( v(X, t) \), and acceleration \( a(X, t) \) of particle \( X \) at time \( t \) are then

\[
u = \frac{\partial}{\partial t} X(x, t) \quad v = \frac{\partial}{\partial t} X(x, t) \quad a = \frac{\partial}{\partial t} v(X, t) = \frac{\partial^2}{\partial t^2} X(x, t)
\]

A variety of different devices can be used as a measure of the deformation of \( B \) relative to \( C_0 \). We prefer to use the Green-Saint Venant deformation tensor \( G_{ij} \) and/or the Green-Saint-Venant strain tensor \( \varepsilon_{ij} \) which describe the deformation of a material line element of length \( ds_0 \) in \( C_0 \) and \( ds \) in \( C \) in the following sense:

\[
ds^2 - ds_0^2 = 2\varepsilon_{ij} dx^i dx^j = (G_{ij} - \delta_{ij} \sigma^2) dx^i dx^j
\]

The strain-displacement relations are then

\[
\varepsilon_{ij} = \frac{1}{2} (u_{i,j} + u_{j,i} + u_{k,k} \delta_{ij})
\]

wherein \( i, j, k = 1, 2, 3 \) and the commas denote partial differentiation with respect to the material coordinates \( (u_{i,j} = \partial u_i / \partial x^j) \). In (5.4) and (5.5) (and in the discussion to follow) we employ the standard index notations and
the summation convention.

While the body $B$ occupies a configuration $C(t = t)$, the stress vector $\sigma(x, t)$ at a particle $x$ developed on a material surface through $x$ with unit normal $\mathbf{n}$ is of the form

$$
\sigma = \sigma^{ij}(x, t)n_iG_j(x, t)
$$

(5.6)

where $n_i = n\cdot G_i$, $G_j(x, t)$, and $\sigma^{ij}$ are the contravariant components of the Piola-Kirchhoff stress tensor. Let $F^i(x, t)$ and $S^i(x, t)$ denote the components of body force per unit mass in $C$ and the contravariant components of surface traction per unit of material surface area on the exterior surface $\partial B$ of $B$ at time $t$. Then linear momentum and angular momentum are balanced at a particle $x$ at time $t$ if and only if

$$
\sigma^{ij}n_j + \rho F^i = \rho \dot{u}^i \quad x \in \mathcal{B} ;
\sigma^{ij}n_j = S^i \quad x \in \partial \mathcal{B} ;
\sigma^{ij} = \sigma^{ij}
$$

(5.7)

where $(\cdot ;)$ denotes covariant differentiation with respect to $x^i$ and $\rho$ is the mass density at $x$ at time $t$. If, instead of $\sigma(x, t)$ of (5.6), we use the stress vector $\tau(x, t) = \sigma(x, t)dA/dA_x$, where $dA$ and $dA_x$ is the element of material area on which $\sigma$ acts in $C$ and $C_x$, respectively, then (5.7) can be rewritten in the form

$$
\left[\sigma^{ij}(t)u_{ij} + \rho \frac{\partial}{\partial t} \right] = \rho \dot{u}^i \quad x \in \mathcal{B} ;
\tau^{ij} = \tau^{ij}
$$

(5.8)

Here $\rho = \rho(x, 0)$, $\dot{u}^i = \partial u^i(x, t)/\partial t$, $\mathbf{f}$, and $\mathbf{S}$ are the components of body force and surface force relative to orthonormal basis vectors $\mathbf{e}_i = \nabla(x, 0)$, $\dot{\mathbf{e}}_i = \nabla(x, 0)$, and $\tau^{ij} = \sqrt{G} \sigma^{ij}$, where $G = \det G_{ij}$.

In addition to the purely mechanical equations (5.1) - (5.7), we impose the principle of conservation of energy and the Clausius-Duhem inequality. Locally, at a particle $x$ at time $t$, these take the respective forms

$$
\rho \dot{\varepsilon} = \sigma^{ij} \dot{y}_{ij} + q^i - \omega + \frac{1}{3} \mathbf{q}^2 \mathbf{\varepsilon} \geq 0
$$

(5.10)

where $\varepsilon$ is the internal energy density, $q^i$ are the contravariant components of heat flux per unit of material surface area, $\omega$ is the heat supplied from internal sources per unit mass, $\theta$ is the absolute temperature, and $\mathbf{\varepsilon}$ is the entropy density. Alternately, we can introduce the Helmholtz free energy

$$
\varphi = \varepsilon - \theta \mathbf{\varepsilon}
$$

(5.11)

and the internal dissipation

$$
\sigma = \rho \dot{\varepsilon} - q^i - \omega
$$

(5.12)

and rewrite (5.9) and (5.10) in the form

$$
\rho \dot{\varphi} = \sigma^{ij} \dot{y}_{ij} - \rho \theta \dot{\mathbf{\varepsilon}} + \frac{1}{3} \mathbf{q}^2 \mathbf{\varepsilon} \geq 0
$$

(5.13)

To complete the description of the behavior of the body $B$, we need only add to the basic kinematical and physical laws the constitutive equations which characterize the material of which the body is composed. In most of the applications we consider in the following sections, the materials fall into a general class which can be described by a collection of four constitutive equations which give the free energy $\varphi$, the stress tensor $\sigma^{ij}$, the entropy density $\mathbf{\varepsilon}$ and the heat flux $q^i$ at particle $x$ at time $t$ as functions of the histories of the deformation and temperature at $x$ and current values of the temperature gradient at $x$. For example,

$$
\varphi = \int\left[\mathbf{\varepsilon}(s), \theta(s); \mathbf{g}(s)\right] \quad \sigma^{ij} = \mathbf{G}^{ij}\left[\mathbf{y}^i(s), \theta^i(s); \mathbf{g}(s)\right]
$$

(5.14)

where $\mathbf{G}^{ij} = \mathbf{G}(x, t)$ and $\mathbf{G}(x)$, $\theta^i(s)$ and $\mathbf{g}(s)$, are the total histories of the strain tensor $\gamma$ and the absolute temperature at $x$ at time $t$ and $\mathbf{G}^{ij} = \mathbf{G}(x, t)$ is the temperature gradient at $x$ at time $t$. Special forms of these and other constitutive equations are to be introduced in special applications.
5.3 General Equations of Motion and Heat Conduction of Finite Elements.

The general equations of motion and heat conduction of finite elements of continuous bodies have been presented in a number of places (see [1]). In the interest of completeness, we record them here with only brief details as to their physical interpretation.

Following the usual procedure, we represent the body $B$ as a collection $E \cup \Theta_\ast$ of material finite elements $\Theta_\ast$. Making use of the fundamental property $e^1$ of finite-element approximations (see [1]), we isolate a typical element $\Theta_\ast$ and represent the local displacement and temperature fields $u_\ast(\xi, t)$ and $\theta_\ast(\xi, t)$ by semidiscrete finite-element models:

$$u_\ast(\xi, t) = \bar{u}_\ast^n(t)N_\ast^\xi(\xi) \quad \theta_\ast(\xi, t) = \bar{\theta}_\ast^n(t)N_\ast^\xi(\xi)$$

(5.15)

Here the local interpolation functions $N_\ast^\xi(\xi)$ and $N_\ast^\xi(\xi)$ satisfy (4.9) so that if $\xi^j$ is a nodal point of element $\Theta_\ast$,

$$u_\ast^n(t) = D_{\ast}u_\ast^n(\xi^j, t) \quad \bar{\theta}_\ast^n(t) = D_{\ast}\theta_\ast^n(\xi^j, t) - T_\ast$$

(5.16)

In (5.15) and (5.16), we have used the following conventions: the summation signs $\sum_z \sum_N$ are dropped, repeated multi-indices $\sigma < z^\xi$ and nodal indices $N$ shall be summed throughout their ranges, ($|z| \leq q$ for a representation of order $q - 1$, $N = 1, 2, \ldots, N_\ast$); $N_\ast^\xi(\xi)$ and $N_\ast^\xi(\xi)$ may be different types of local interpolation functions, $T_\ast$ is a uniform reference temperature in $\Theta_\ast$, and $T(\xi, t) = \theta(\xi, t) - T_\ast$ is the change in temperature experienced at particle $\xi$ during the motion of the body from configuration $\Theta_\ast$ to $C$.

Linear momentum is balanced in a mean sense over the body $B$ if

$$\frac{\partial}{\partial t}(\bar{u}^\ast - \bar{u}^\ast) - \sum_{\ast} (\bar{f}_\ast - \bar{p}_\ast - \bar{s}_\ast) = 0$$

(5.17)

and energy is conserved if $\langle \rho \bar{q} \eta \rangle - \eta^\prime - 2\sigma - \sigma_h \eta = 0$ where

$$\langle u, v \rangle = \int_{\Omega} u v d\Omega$$

In (5.17), the first member in the pair $(\cdot, \cdot)$ is the net force acting at particle $\xi$ at time $t$. If $\mathcal{F}$ is the space of all such forces, $h$ is a member of the conjugate space $\mathcal{F}^\ast$. Physically, $h = h(\xi, t)$ can be interpreted as the velocity field at $\xi$ at $t = t$; then $\langle \rho \bar{u}_\ast^\xi(t), h \rangle$ is the time-rate of change of kinetic energy, $-\langle \bar{f}^\ast(t), \bar{u}_\ast^\xi(t) \rangle$ is the time-rate of change of the internal energy (minus the heat developed), and $\langle \rho \bar{p}_\ast^\xi(t), h \rangle$ can be associated with the mechanical power of the external forces $\rho \bar{p}_\ast^\xi$. Then (5.17) appears as a statement of conservation of mechanical energy for $B$. Locally, for a typical element $\Theta_\ast$, we introduce (5.14) into (5.17) to obtain the Galerkin approximation:

$$\langle u, v \rangle_{\Theta_\ast} = \langle \bar{u}^\ast(t), \bar{v}^\ast(t) \rangle - \langle \rho \bar{p}_\ast^\xi(t), \bar{v}^\ast(t) \rangle - \langle \bar{f}^\ast(t), \bar{u}_\ast^\xi(t) \rangle - \langle \rho \bar{p}_\ast^\xi(t), \bar{u}_\ast^\xi(t) \rangle$$

(5.18)

where

$$\langle u, v \rangle = \int_{\Omega} u v d\Omega \quad \langle u, v \rangle_{\Theta_\ast} = \int_{\Omega} u v d\Omega_{\Theta_\ast}$$

(5.19)

Introducing (5.16) into (5.19) and (5.18) into the boundary integral, we arrive at the general equations of motion of a finite element of a continuous body:

$$\rho^\ast \frac{\partial^2}{\partial t^2} \bar{u}_\ast^\xi \frac{\partial}{\partial t} \bar{u}_\ast^\xi - \sum_{\ast} (\bar{f}_\ast - \bar{p}_\ast - \bar{s}_\ast) = 0$$

(5.20)

where $\bar{M}_\ast$ is the consistent mass matrix, $\bar{p}_\ast^\xi(t)$ are the consistent generalized forces:

$$\rho^\ast \frac{\partial^2}{\partial t^2} \bar{u}_\ast^\xi \frac{\partial}{\partial t} \bar{u}_\ast^\xi = \langle \rho \bar{p}_\ast^\xi, \bar{u}_\ast^\xi(t) \rangle$$

(5.21)

$$\bar{p}_\ast^\xi(t) = \langle \rho \bar{p}_\ast^\xi, \bar{u}_\ast^\xi(t) \rangle + \langle \bar{s}_\ast^\xi, \bar{u}_\ast^\xi(t) \rangle$$

(5.22)

A similar procedure, starting with (5.12) and using the temperatures $T = T^\ast$, as weight functions, yields the general equation of heat conduction for finite elements:

$$\rho^\ast \frac{\partial}{\partial t} \bar{q}_\ast^\xi(t) - \sum_{\ast} (\bar{f}_\ast^\xi + \bar{u}_\ast^\xi(t) - \bar{\theta}_\ast^\xi(t) - \bar{p}_\ast^\xi(t)) = 0$$

(5.23)

Here $\bar{q}_\ast^\xi$ is the generalized normal heat flux at node $N$ and $\bar{q}_\ast^\xi$ are the
5.4 General Mixed Finite Element Models.

In many applications of the finite-element method to problems in continuum mechanics, it is convenient to select as the basic dependent variables to be approximated quantities other than (or in addition to) the usual variables dictated by classical notions of determinism (e.g., displacements, velocities, temperatures). For example, in problems of fracture mechanics or plasticity, greater accuracy in stresses than in displacements may be desirable. Then mixed or hybrid finite element models can often prove to be useful. To illustrate the character of such approximations, we present in this section the general mixed finite-element model for nonlinear continua.

Let $u * v$ denote the convolution operator

$$u * v = \int_0^T u(x, t-\tau)v(x, \tau) d\tau$$

and define the convolution-scalar product

$$(u, v)_* = \int_0^T u * v d\tau$$

Then the functional

$$K(A) = \frac{1}{2} \left[ \sum_{s=0}^{n} \left[ 2g_u * \delta_t^{-1} u + 2g_u * \delta_t^{-1} v + 2g_u * \delta_t^{-1} \right] - g_u * \delta_t^{-1} + 2g_u * \delta_t^{-1} \right]$$

assumes a stationary value whenever the linear momentum equation, the constitutive equation for stress, the strain-displacement equations, the "constitutive" equations for internal dissipation, the constitutive equations for heat flux and entropy, and the temperature gradient identity $g_{\eta} = \theta_{\eta}$ are satisfied by the field $A = [u_\alpha, v_\alpha, t^{\alpha\beta}, \sigma, q_\alpha, \epsilon, \theta]$. In (5.26), $g = t^\alpha \gamma_\alpha, g_\alpha = 1, g_{\eta} = \delta_\eta \gamma_\eta$, and $\gamma_\alpha$ are functionals of $[-] = [t^{\alpha\beta}, \theta, \delta_\eta \gamma_\eta]$, $\delta_\eta \gamma_\eta$ being an appropriate functional for the internal dissipation. Also, for convenience, we have defined $\dot{T}_\eta = 0$ and $\hat{T} = g*[(h + \sigma)(x, t) + \rho T_\eta (x, t)]$.

The euler equations for the function $K(A)$ are then the entire collection of equations governing the thermomechanical behavior of a continuous body $B$. Note that among the euler equations of $K(A)$ we obtain, instead of (5.7) and (5.12), the convolution $g* (\cdot)$ of these equations; also note that, for simplicity, boundary integrals are omitted; i.e., homogeneous boundary conditions are assumed.

Considering now a typical element $\Theta$ of a finite-element model of the body $B$, we introduce the independent local approximations

$$u_i(t), v_i(t), t^{ij}(t), \sigma_i(t), q_i(t), \epsilon_i(t), \theta_i(t)$$

Here we have omitted the multi-indices $\alpha$ for simplicity, the local interpolation functions $t_i(t), t^{ij}(t), \sigma_i(t), q_i(t), \epsilon_i(t), \theta_i(t)$ satisfy (4.9), the repeated nodal indices $N$ are summed from 1 to $N$, and the coefficients $u_i, v_i, t^{ij}, \sigma_i, q_i, \epsilon_i, \theta_i$ are understood to be functions of time.

Now let $K_\alpha(\Lambda)$ denote the value of $K(A)$ at time $t = T$. We partition the interval $[0, T]$ into partitions $\tau_i = [t_i, t_{i+1}]$ and represent each nodal function $u_i^n(t), \ldots, \eta^n(t)$ by some appropriate difference scheme in the time domain, e.g.,

$$u_i^n(t) = \sum_{s=0}^{n} (u_i^n(t_s) h_i + \dot{u}_i^n(t_s) h_i')$$

etc. where $t \in [t_i, t_{i+1}]$ and $h_i(t), h_i'(t)$ are Hermite interpolation functions.
Introducing (5.27) with the time-approximations (5.28) into (5.26) and performing the indicated integrations over $B \times [0,T]$ yields a function $\tilde{K}(\lambda)$ of independent coefficients $u^i_t, \gamma_i, \gamma_{ij}, \ldots, \gamma_{i\ldots} \ldots \eta$ etc. Direct application of the Ritz method yields finite-element analogs of the entire collection of field equations. We shall reserve a detailed discussion of such models for a future paper.

The static (thermoelastic) case is also interesting. Then $u_t = \dot{\gamma} = \dot{\eta} = 0$ and $K(\lambda)$ reduces to

$$
\tilde{K}(\lambda) = \frac{1}{2} \int_B \left[ 2 \gamma_{ij} \dot{u}^i u^j + \sigma^2 + 2\sigma \delta \sigma \right] - u_x \left( 2 \gamma_{ij} \dot{u}^i - (\tau^i \dot{u}_t)^j \right) + 2 \eta \eta' + 2 \eta \gamma_i \gamma^j + q^l (g_i - \theta^l) + q^l \gamma_i - 2 \tau_t \gamma^l \right] du
$$

wherein $\gamma_{ij} \ldots \gamma_{i\ldots} \ldots \eta$ are now functions of the current values of $\gamma_{ij}, \gamma$, and $\eta$. Introducing (5.27) into (5.29) and noting that now the nodal values $u^i_t, \ldots, \eta_t$ are independent of time, we obtain a function of the type

$$
\tilde{K}(u_t, \gamma_{ij}, \gamma^l, \sigma, q^l, \theta, \gamma_i, \gamma, \eta)
$$

We then generate the systems of equations

$$
\frac{\partial \sigma}{\partial u_t} = 0, \quad \frac{\partial \sigma}{\partial \gamma_{ij}} = 0, \quad \frac{\partial \sigma}{\partial \tau_t} = 0, \quad \frac{\partial \sigma}{\partial \gamma^l} = 0, \quad \frac{\partial \sigma}{\partial q^l} = 0, \quad \frac{\partial \sigma}{\partial \gamma_i} = 0
$$

$$
\frac{\partial \gamma_{ij}}{\partial u_t} = 0, \quad \frac{\partial \gamma_{ij}}{\partial \gamma_{ij}} = 0, \quad \frac{\partial \gamma_{ij}}{\partial \tau_t} = 0, \quad \frac{\partial \gamma_{ij}}{\partial \gamma^l} = 0, \quad \frac{\partial \gamma_{ij}}{\partial q^l} = 0, \quad \frac{\partial \gamma_{ij}}{\partial \gamma_i} = 0
$$

Equation (5.30a) is the finite-element analogue of the equilibrium equations, (5.30b) that of the constitutive equations for stress, (5.30c) the compatibility equation (5.5), (5.30d) the constitutive equation for the internal dissipation, (5.30e) the temperature gradient-temperature identity, (5.30f) the heat conduction equation, (5.30g) the constitutive equation for heat flux, and (5.30h) the constitutive equation for specific entropy. Space does not permit us to discuss the numerous variants and special cases available in (5.21) - (5.30).

Acknowledgement. The support of this work by the U.S. Air Force Office of Scientific Research under Contract F44620-69-C-0124 is gratefully acknowledged.

APPENDIX I

REFERENCES