QUARTERLY OF APPLIED MATHEMATICS
APRIL, 1971

CONJUGATE APPROXIMATION FUNCTIONS IN FINITE-ELEMENT ANALYSIS*

BY

H. J. BRAUCHLI AND J. T. ODEN

Research Institute, University of Alabama in Huntsville

Summary. In this paper a theory of conjugate approximations is developed which provides a fundamental basis for most methods of continuous piecewise approximation. It is shown that for a given finite set of base functions used in an approximation there corresponds another set of conjugate functions which play a significant role in approximate methods of analysis. In the case of finite-element approximations, it is shown that the domain of the conjugate functions includes the entire assembly of elements, and, consequently, the established method of computing stresses locally in elements based on displacement approximations is not strictly valid. Indeed, the domain of such "local" stress fields is the entire connected system of elements. Procedures for computing derivatives and discrete analogues of linear operators which are consistent with the theory of conjugate functions are also discussed. For a given linear operator equation, the significance of the conjugate approximations in connection with the adjoint problem is also discussed.

1. Introduction. In a recent paper [1] the idea of "conjugate fields" in connection with finite-element approximations of a given function was introduced as a generalization of the concept of generalized forces and displacements familiar in Lagrangian mechanics. Basically, the idea follows from the observation that in applications of the finite-element method we generally encounter, together with the primary function $f(x)$ which is to be approximated by a discrete model $F(x)$, another function $g(x)$ and a linear functional $q[f(x), g(x)]$. We say that the function $g(x)$ is conjugate to $f(x)$ with respect to $q[\ ]$. It is assumed that the function $F(x)$ can be defined in terms of its values $F^k = F(x^k)$ at a finite number G of points. The usual procedure is to construct a corresponding approximation $G(x)$ of $g(x)$ so that $\sum_k^G F(x^k) \cdot G(x_k)$ and $q[F(x), g(x)]$ coincide. Archer [2] first used this procedure to derive "consistent" mass and force approximations in certain structural dynamics problems. We shall also refer to the function values $G(x_k)$ so computed as *consistent* with the approximations $F(x)$ with respect to the functional $q[\ ]$.

In this paper, we explore the concept of conjugate approximation functions in detail, and show that these functions possess certain properties which are fundamental to approximation methods in general, and to the finite-element method in particular. Our investigation is based on the fact that the functions $g(x)$ are elements in a space $\mathfrak{F}^*$ which is the conjugate space (dual space) of the space $\mathfrak{F}$ to which $f(x)$ belongs. We show that while the primary function $F(x)$ may be defined only locally, i.e., only over specific finite elements, their corresponding conjugate approximations are defined globally. This means that the often-used procedure of calculating stresses locally in finite elements

* Received October 30, 1969; revised version received February 23, 1970.

using locally-defined displacement fields is not strictly correct, and that only a global description of the stress field may be consistent with the displacement approximation with respect to (say) the potential-energy functional. We cite a simple example to demonstrate this point.

Following this introduction are three sections devoted to developing the general concepts of conjugate approximations. We then show that such approximations provide the best approximation of a given function $f(x)$ in the sense of a metric defined on $\mathfrak{F}$. Sec. 6 of the paper deals with basic properties of conjugate approximations which pertain to their relation with the Dirac delta function when it is admitted into $\mathfrak{F}$. This is followed by a discussion of discrete models of derivatives of functions. Secs. 9 and 10 of the paper demonstrate the application of the general theory to finite-element representations and include an example to emphasize the importance of conjugate approximations in stress calculations. We also address ourselves to the problem of determining consistent approximations of linear operators defined on linear function spaces. We show that for certain classes of equations the conjugate functions appear in approximate solutions to the adjoint problem.

2. **Finite projections of function spaces.** Consider a complete, normed, linear space $\mathfrak{F}$ (a Banach space) the elements of which are functions defined on a bounded domain $\mathfrak{R}$ of an n-dimensional Euclidean space. Elements of $\mathfrak{F}$ are denoted $f(x)$, $g(x)$, $\cdots$ where x is a point in $\mathfrak{R}$. We also assume that there is an inner product p defined on $\mathfrak{F} \times \mathfrak{F}$ denoted (f, g), and we shall use the natural norm $||f|| = (f, f)^{1/2}$ with respect to p.

In most applications we are concerned with the special case in which $\mathfrak{F}$ is the space of square integrable functions on $\mathfrak{R}$. Then the inner product of two elements $f(x)$, $g(x) \in \mathfrak{F}$ is given by

$$(f, g) = \int_{\mathfrak{R}} f(x)g(x)\, d\mathfrak{R}, \tag{2.1}$$

where integration in the Lebesgue sense is implied and $f(x)$ and $g(x)$ are orthogonal if $(f, g) = 0$. In this case

$$(f, f) < + \infty \tag{2.2}$$

for every $f(x) \in \mathfrak{F}$.

Let Φ denote a linear subspace of $\mathfrak{F}$ of dimension G which is spanned by a set of G linearly independent functions $\varphi_k(x)$; $k = 1, 2, \cdots, G$. In the following we confine our attention to the case in which Φ contains only continuous functions; the interesting case in which Φ is expanded to include functions with finite discontinuities is to be the subject of a later paper. Thus $\mathfrak{F}$ is regarded as the direct sum of two subspaces Φ and Γ: $\mathfrak{F} = \Phi \oplus \Gamma$ with $\Phi \cap \Gamma = \theta$, θ being the null space, such that each element $f(x) \in \mathfrak{F}$ can be represented in a unique way as the sum of a function $F(x) \in \Phi$ and a function $\bar{f}(x) \in \Gamma$:

$$f(x) = F(x) + \bar{f}(x). \tag{2.3}$$

By construction,

$$(\bar{f}, \varphi_k) = 0 \qquad k = 1, 2, \cdots, G, \tag{2.4}$$

so that Γ is the orthogonal complement of Φ.

It follows that (2.3) and (2.4) define a projection operator Π: $\mathfrak{F} \rightarrow \Phi$ such that

$$F(x) = \Pi f(x). \tag{2.5}$$

Since the functions $\varphi_k(x)$ are linearly independent, every element $F(x) \in \Phi$ can be represented as a unique linear combination of $\varphi_1(x)$, $\varphi_2(x)$, $\cdots$, $\varphi_G(x)$:

$$F(x) = F^k \varphi_k(x), \tag{2.6}$$

where the repeated index k is to be summed from 1 to G. The multipliers F^k are referred to as components of $F(x)$.

It is important to realize that the functions $\varphi_k(x)$ are not necessarily orthogonal. Indeed, we shall introduce the symmetric $G \times G$ matrix

$$C_{ij} \equiv (\varphi_i , \varphi_j) \qquad i, j = 1, 2, \cdots, G, \tag{2.7}$$

which is referred to as the fundamental matrix of Φ. Then the $\varphi_i(x)$ are orthogonal if and only if $C_{ij} = c_{(i)}\, \delta_{ij}$, where δ_{ij} is the Kronecker delta; that is, if C_{ij} is diagonal. Since the set $\{\varphi_k\}$ is linearly independent and is prescribed in the definition of Φ, the matrix C_{ij} is regular and can be generated directly by means of (2.7). We denote the inverse of C_{ij} by C^{ij}:

$$(C_{ij})^{-1} = C^{ij} \tag{2.8}$$

or

$$C_{ij} C^{jk} = \delta_i^k , \tag{2.9}$$

where $i, j, k = 1, 2, \cdots, G$. The matrices C_{ij} and C^{ij} play an important role in the theory of conjugate approximations.

3. Conjugate subspaces. Let $q(f)$ denote a linear functional defined on $\mathfrak{F}$; that is, the operator q defines a linear mapping of the elements $f \in \mathfrak{F}$ into the real (or complex) number field. The collection of all such linear functionals on $\mathfrak{F}$ is itself a linear space $\mathfrak{F}^*$ called the conjugate space (or dual space) of $\mathfrak{F}$. Since q: $\mathfrak{F} \rightarrow R$ is linear,

$$q(f) = q(F) + q(\bar{f}), \tag{3.1}$$

where

$$q(F) = q(\Pi f) = F^k q(\varphi_k). \tag{3.2}$$

Thus the conjugate space $\mathfrak{F}^*$ can also be decomposed into the direct sum of two subspaces, $\mathfrak{F}^* = \Phi^* \oplus \Gamma^*$, where Φ^* is the G-dimensional subspace conjugate to Φ. Indeed, if $f^*(x)$ is an element of $\mathfrak{F}^*$, the subspace Φ^* is obtained as a projection Π^*: $\mathfrak{F}^* \rightarrow \Phi^*$ and every element $F^*(x) \in \Phi^*$ is the image of the projection; $F^*(x) = \Pi^* f^*(x)$. To describe $F^*(x)$ further, we must identify a basis for the subspace Φ^*.

In order to generate the subspace Φ^* which is conjugate to Φ, we introduce a system of G linearly independent functions $\varphi^k(x)$, $k = 1, 2, \cdots, G$ defined by

$$(\varphi^k, \varphi_m) = \delta_m^k . \tag{3.3}$$

In order that the set $\varphi^k(x)$ be uniquely defined, we impose the condition

$$(\varphi^k, \bar{f}) = 0. \tag{3.4}$$

Then each $\varphi^k(x)$ can be represented as a linear combination of the functions $\varphi_k(x)$:

$$\varphi^k(x) = G^{km}\varphi_m(x). \tag{3.5}$$

Thus $\varphi^k(x) \in \Phi$, which means that the subspaces Φ and Φ^* are identical. The set of functions $\varphi^k(x)$ is referred to as the conjugate basis of the set $\varphi_k(x)$. The two sets of functions $\varphi_k(x)$ and $\varphi^k(x)$ are said to form a biorthogonal basis of Φ.

We shall now show that since (3.3) and (3.4) define a unique conjugate basis, the matrix G^{km} is equal to C^{km}, the inverse of the fundamental matrix defined in (2.8). Simply take the inner product of (3.5) with $\varphi_r(x)$:

$$(\varphi^k, \varphi_r) = (G^{km}\varphi_m , \varphi_r) = G^{km}(\varphi_m , \varphi_r) = G^{km}C_{mr} = \delta^k_r . \tag{3.6}$$

Therefore, instead of (3.5) we have

$$\varphi^k(x) = C^{km}\varphi_m(x). \tag{3.7}$$

By virtue of the positive definiteness of C_{mr} , which will be demonstrated subsequently, it can be shown to follow that (3.5) defines a unique set of functions $\varphi^k(x)$ satisfying (3.3).

Another important property follows from (3.7). Observe that

$$(\varphi^k, \varphi^r) = (C^{km}\varphi_m , \varphi_r) = C^{km}(\varphi_m , \varphi_r). \tag{3.8}$$

Thus

$$(\varphi^k, \varphi^r) = C^{kr} = (C_{kr})^{-1}. \tag{3.9}$$

Therefore, the inverse of the fundamental matrix C_{kr} of the subspace Φ is the fundamental matrix of the conjugate space Φ^*.

Multiplying (3.7) by C_{rk} and taking into account (2.9), we also find that

$$\varphi_k(x) = C_{km}\varphi^m(x). \tag{3.10}$$

Additional properties of the functions $\varphi_k(x)$ and $\varphi^k(x)$ are to be explored in later sections.

4. Conjugate approximations. If the functions $\varphi_k(x)$ are suitably chosen, it is understood that the functions $F(x)$ defined by (2.6) can be used as approximations to the original functions $f(x) \in \mathfrak{F}$. Explicit equations for the components F^k of $F(x)$ and for the projection $F(x) = \Pi f(x)$ can now be obtained using the biorthogonal bases $\{\varphi_k(x)\}$ and $\{\varphi^k(x)\}$. Taking the inner product of $F = F^k\varphi_k(x)$ with $\varphi^m(x)$ and introducing (3.3), we get

$$F^k = (F, \varphi^k). \tag{4.1}$$

Furthermore, this result suggests that we also compute a set of conjugate components using the functions $\varphi_k(x)$

$$F_k = (F, \varphi_k). \tag{4.2}$$

It follows from (3.4) and (3.8) that

$$F^k = (F, C^{km}\varphi_m(x)) = C^{km}F_m , \tag{4.3}$$

$$F_k = (F, C_{km}\varphi^m(x)) = C_{km}F^m. \tag{4.4}$$

Thus the fundamental matrices C_{km} and C^{km} have the property of "raising and lowering" indices on the components F_k , F^k. Furthermore, the projection $F(x)$ can now be written

$$F(x) = F^k\varphi_k(x) = F_k\varphi^k(x) \tag{4.5}$$

wherein, as usual, the repeated index k is summed from 1 to G. The form of (4.5) suggests that F^k and F_k be called contravariant and covariant components of $F(x)$, respectively.

Let $F(x)$ and $G(x)$ denote two elements of the subspace Φ:

$$F(x) = F^k\varphi_k(x) = F_k\varphi^k(x) \quad \text{and} \quad G(x) = G^k\varphi_k(x) = G_k\varphi^k(x). \tag{4.6}$$

Then the inner product of F and G is given by

$$(F, G) = (F^k\varphi_k , G^m\varphi_m) = (F^k\varphi_k , G_m\varphi^m) = (F_k\varphi^k, G_m\varphi^m) \tag{4.7}$$

or

$$(F, G) = C_{km}F^kG^m = F^kG_k = C^{km}F_kF_m . \tag{4.8}$$

In particular,

$$||F||^2 \equiv (F, F) = C_{km}F^kF^m = F^kF_k = C^{km}F_kF_m \geq 0, \tag{4.9}$$

where $||F|| = (F, F)$ is the natural norm of F on Φ. Eq. (4.9) demonstrates the positive-definite character of C_{km} mentioned previously.

In certain applications when $\mathfrak{F}$ is a Hilbert space with an inner product given by (2.1), it is convenient to also introduce quantities M_k and M^k defined by

$$M_k = \int_{\mathfrak{R}} \varphi_k(x)\, d\mathfrak{R}, \tag{4.10}$$

$$M^k = \int_{\mathfrak{R}} \varphi^k(x)\, d\mathfrak{R} = C^{km}M_m . \tag{4.11}$$

Then

$$\int_{\mathfrak{R}} F(x)\, d\mathfrak{R} = F^kM_k = F_kM^k. \tag{4.12}$$

5. Best approximation. We shall now show that for a given element $f(x) \in \mathfrak{F}$ and a given projection defined by a prescribed set of functions φ_k , the "best" approximation to $f(x)$ in the subspace Φ is that function for which the components are given by (4.3) (or by (4.4)). By the "best approximation," we shall mean the element in Φ closest to $f(x)$ in the sense of the natural metric $d[f(x), g(x)] = ||f(x) - g(x)||$ defined on $\mathfrak{F}$.

Let Λ^k denote an arbitrary collection of G quantities such that $\Lambda^k\varphi_k(x) \in \Phi$. We are concerned with determining the Λ^k so as to minimize the functional

$$J(\Lambda^k) = (f - \Lambda^k\varphi_k , f - \Lambda^k\varphi_k), \tag{5.1}$$

which, since $J(\Lambda^k) = (d[f(x), \Lambda^k\varphi_k(x)])^2$, is a measure of the distance between $f(x)$ and $\Lambda^k\varphi_k(x)$. Noting that

$$(f, \Lambda^k\varphi_k) = \Lambda^k(f, \varphi_k) = \Lambda^kF_k \tag{5.2}$$

and introducing (4.9), we find that (5.2) can be recast in the form

$$J(\Lambda^k) = ||f||^2 - ||F||^2 + C_{km}(\Lambda^k - F^k)(\Lambda^m - F^m). \tag{5.3}$$

It is easily seen that the functional $J(\Lambda^k)$ assumes minimal value when

$$F^k = \Lambda^k. \tag{5.4}$$

Thus, the components F^k computed using (4.1) provide the best approximation of f in Φ in the sense of (5.1). We observe that the mean-square error, in the sense of $J(f)$, induced by representing $f(x)$ by its projection $F(x) = \Pi f(x)$ is

$$E(F^k) = ||f||^2 - ||F||^2. \tag{5.5}$$

In view of (2.3),

$$||f||^2 - ||F||^2 = ||\bar{f}||^2 + 2(F, \bar{f}) = ||\bar{f}||^2, \tag{5.6}$$

since F and $\bar{f}$ are orthogonal. Thus

$$E(F^k) = ||\bar{f}||^2. \tag{5.7}$$

Since in general $||\bar{f}||$ may be unbounded, $E(F^k)$ is small only for f close to the subspace Φ.

6. Some properties of conjugate approximations. Suppose that we enlarge the space $\mathfrak{F}$ so that it includes distributions such as the Dirac delta function[1] $\delta(x - a)$; $x, a \in \mathfrak{R}$, defined by

$$\begin{aligned} \delta(x - a) &= 0 \quad x \neq a \\ &= \infty \quad x = a \end{aligned}; \qquad \int_{\mathfrak{R}} \delta(x - a)\, d\mathfrak{R} = 1. \tag{6.1}$$

Let $\Delta(x - a)$ denote the projection of $\delta(x - a)$ on the subspace Φ. Then

$$\Delta(x - a) = \Pi\delta(x - a) = \Delta^k\varphi_k(x) = \Delta_k\varphi^k(x), \tag{6.2}$$

where, according to (4.1) and (4.2),

$$\Delta^k = (\delta(x - a), \varphi^k(x)) = \varphi^k(a), \tag{6.3}$$

$$\Delta_k = (\delta(x - a), \varphi_k(x)) = \varphi_k(a). \tag{6.4}$$

Hence

$$\Delta(x - a) = \varphi^k(a)\varphi_k(x) = \varphi_k(a)\varphi^k(x). \tag{6.5}$$

Thus the values of the functions φ^k (or φ_k) at an arbitrary point $a \in \mathfrak{R}$ are the components Δ^k (or Δ_k) of the projection of the delta function $\delta(x - a)$.

Observe that while the function $\delta(x - a)$ assumes a nonzero value only at a, the projection $\Delta(x - a)$ may take on nonzero values almost everywhere in the domain of functions in Φ. However, the essential properties of $\delta(x - a)$ are preserved under the projection $\Pi\colon \mathfrak{F} \to \Phi$. For example, note that

$$\begin{aligned} (\Delta(x - a), F(x)) &= (\varphi_k(a)\varphi^k(x), F^m\varphi_m(x)) \\ &= \varphi_k(a)F^m(\varphi^k, \varphi_m) \\ &= \varphi_k(a)F^k \\ &= F(a). \end{aligned} \tag{6.6}$$

Also if $F(x) = 1$ is an element of Φ, then

$$(\Delta(x - a), 1) = 1. \tag{6.7}$$

To demonstrate another property of the functions $\varphi_k(x)$, suppose that we identify

[1] We recognize that $\delta(x - a)$ is not square integrable. Thus our conclusions concerning the best approximation of elements of $\mathfrak{F}$ do not hold. However, the projection $\Pi\delta(x - a)$ is well defined.

in $\mathfrak{R}$ a finite number G of points a^m; $m = 1, 2, \cdots, G$. We shall refer to such points as nodal points or simply nodes. Further, suppose the function $\varphi_k(x)$ has a value of unity at node a^k but is zero at all other nodes:

$$\varphi_k(a^m) = \delta_k^m . \tag{6.8}$$

Then the projection of the delta function in (6.5) becomes

$$\Delta(x - a^m) = \varphi_k(a^m)\varphi^k(x) = \varphi^m(x). \tag{6.9}$$

We shall refer to base functions $\varphi_k(x)$ with property (6.8) as being normalized with respect to the G nodes a^m. In Sec. 9 we discuss the procedure for computing such normalized base functions for finite element approximations. Eq. (6.9) shows that the conjugate functions $\varphi^k(x)$ represent the projection of the delta function $\delta(x - a^k)$ at node a^k.

Assuming that $F(x) = 1 \in \Phi$, note that in addition to (6.8) the functions $\varphi_k(x)$ satisfy the stronger condition

$$\sum_{k=1}^{G} \varphi_k(x) = 1. \tag{6.10}$$

Then the quantities M^k of (4.11) become

$$M^k = C^{km}M_m = C^{km}\int_{\mathfrak{R}} \varphi_m \sum_r \varphi_r \, d\mathfrak{R} = C^{km}\sum_r (\varphi_m , \varphi_r), \tag{6.11}$$

but $(\varphi_m , \varphi_r) = C_{mr}$, by (2.7). Therefore

$$M^k = \sum_{r=1}^{G} C^{km}C_{mr} = \sum_{r=1}^{G} \delta_r^k = 1. \tag{6.12}$$

Similarly,

$$M_k = \sum_{r=1}^{G} (\varphi_k , \varphi_r) = \sum_{r=1}^{G} C_{kr} . \tag{6.13}$$

Observe that

$$\sum_{k=1}^{G} M_k = \int_{\mathfrak{R}} \sum_{k=1}^{G} \varphi_k(x) \, d\mathfrak{R} = \int_{\mathfrak{R}} d\mathfrak{R} = \sum_{k=1}^{G}\sum_{r=1}^{G} C_{kr} = \mathcal{V}, \tag{6.14}$$

where $\mathcal{V}$ is the volume of $\mathfrak{R}$.

In view of (6.12) and (4.12),

$$\int_{\mathfrak{R}} F(x) \, d\mathfrak{R} = \sum_{k=1}^{G} F_k . \tag{6.15}$$

We observe that if the $\varphi_k(x)$ satisfy (6.8) and (6.10) then

$$F(a^m) = F^m = C^{mk}F_k . \tag{6.16}$$

Thus the average value of $F(x)$ over $\mathfrak{R}$ as given by the integral in (6.15) *is not the sum of the values of $F(x)$ at the nodes a^m; rather it is the sum of the conjugate values $F_k = C_{km}F(a^m)$.* We see that F^k is the value of $F(x)$ at node a^k, but F_k represents an average value of $F(x)$ in the neighborhood of a^k.

7. Derivatives of conjugate approximations. We now examine properties of derivatives of the conjugate approximation functions. Let $\partial f(x)$ denote the partial derivative

of a function $f(x) \in \mathfrak{F}$. We shall assume that $\partial_\mu f(x)$ exists and that the derivatives of the base functions $\varphi_k(x)$ also belong to the subspace Φ: $\partial_\mu \varphi_k \in \Phi$. (The latter assumption, of course, is not always valid.) We begin by introducing the array

$$D_\mu^{km} = (\varphi^k, \partial_\mu \varphi^m). \tag{7.1}$$

Then, according to (2.5) and (4.1),

$$\Pi(\partial_\mu \varphi^k(x)) = D_\mu^{km} \varphi_m(x), \tag{7.2}$$

A fundamental question now arises: under what conditions can the projection (7.2) be used as an approximation of the derivatives of the conjugate functions $\varphi^k(x)$? In other words, when is it legitimate to set

$$\partial_\mu \varphi^k(x) = D_\mu^{km} \varphi_m(x)? \tag{7.3}$$

The answer to this question is provided by the following theorem.

THEOREM 7.1. *The following conditions are equivalent to* (7.3):

(a) $\partial_\mu \Pi = \Pi \partial_\mu$ *on the space* Φ,

(b) $\partial_\mu \Phi \subset \Phi$,

(c) Φ *is the solution space of a system of linear differential equations in* X_r *with constant coefficients.*

Proof. Briefly, the proof follows by showing that (7.3) implies conditions (a) and (c), that (a) implies (b), and that (b) implies (7.3). To wit,

(7.3) $\Rightarrow$ (a); $\Pi \partial_\mu \varphi^k(x) = D_\mu^{km} \varphi_m(x) = \partial_\mu \varphi^k(x) = \partial_\mu(\Pi \varphi^k(x))$ for each $\varphi^k(x) \in \Phi$.

(a) $\Rightarrow$ (b); for $\partial_\mu(\Phi) = \partial_\mu \Pi \Phi = \Pi(\partial_\mu \Phi)$ and $\Pi(\partial_\mu \Phi) \subset \Phi$, by definition.

(b) $\Rightarrow$ (7.3); the fact that $\partial_\mu(\Phi) \subset \Phi$ implies that $\partial_\mu \varphi^k(x)$ is a linear combination of the functions $\varphi_k(x)$; as indicated by (7.3).

(7.3) $\Rightarrow$ (c); this follows immediately from the fact that the set $\{\varphi_k(x)\}$ is a complete solution of the equation $\partial_\mu X_k = C_{km} D_\mu^{mr} X_r$.

Finally, (c) $\Rightarrow$ (7.3) because a system of linear differential equations of order n can always be transformed into a system of first-order equations. Hence $\partial_\mu \varphi_k(x)$ is contained in Φ. This completes the proof.

Assuming that the conditions of the theorem are met, we can use (7.3) to obtain the derivative of an arbitrary element $F(x) = F^k \varphi_k(x) \in \Phi$:

$$\partial_\mu F(x) = D_\mu^{km} F_k \varphi_m(x). \tag{7.4}$$

However, the relation

$$\Pi \partial_\mu f(x) = D_\mu^{km} F_k \varphi_m(x) \tag{7.5}$$

need not hold unless the following stronger conditions are imposed on Φ:

($\bar{\text{a}}$) $\partial_\mu \Pi = \Pi \partial_\mu$ on $\mathfrak{F}$,

($\bar{\text{b}}$) $\partial_\mu(\Phi) = \Phi$,

($\bar{\text{c}}$) D_μ^{km} is regular.

It can be shown that conditions (7.4), ($\bar{\text{a}}$), ($\bar{\text{b}}$), and ($\bar{\text{c}}$) are equivalent.

Under the assumption that these stronger conditions are also met, the action of a differentiation ∂_μ in $\mathfrak{F}$ can be completely described in our approximation by the array D_μ^{km} of (7.1), and higher derivatives may be represented as well using powers of D_μ^{km} .

Indeed, introducing the discrete operators

$$D_{\mu\lambda}^{km} = D_{\mu}^{kr} C_{rs} D_{\lambda}^{sm} \tag{7.6}$$

and

$$B_{\mu\lambda}^{km} = D_{\mu}^{kr} D_{\lambda}^{ms} C_{rs} , \tag{7.7}$$

it is easily shown, for example, that

$$\partial_\mu \partial_\lambda f(x) = D_{\mu\lambda}^{km} F_k \varphi_m(x), \tag{7.8}$$

$$\text{div a} = \partial_\mu a^\mu = D_{\mu}^{km} A_k^\mu \varphi_m(x), \tag{7.9}$$

$$\int_{\mathcal{R}} \text{div a} \, d\mathcal{R} = D_{\mu}^{km} A_k^\mu M_m , \tag{7.10}$$

$$(f, \partial_\mu g) = D_{\mu}^{mk} F_k G_m , \tag{7.11}$$

$$(\partial_\mu f, \partial_\lambda g) = B_{\mu\lambda}^{km} F_k G_m , \tag{7.12}$$

$$\int_{\mathcal{R}} \text{grad } f \cdot \text{grad } g \, d\mathcal{R} = B_{\mu\mu}^{mk} F_m G_k . \tag{7.13}$$

We also note that it is possible to use other types of discrete operators such as

$$D_{\mu,km} = (\partial_\mu \varphi_k , \varphi_m) = C_{kr} C_{ms} D_{\mu}^{sr}, \tag{7.14}$$

$$D_{\mu,m}^{\cdot k} = C_{mr} D_{\mu}^{rk} = C^{kr} D_{\mu,rm} , \tag{7.15}$$

$$D_{\mu,m}^{\cdot k} = C_{mr} D_{\mu}^{kr} = C^{kr} D_{\mu,mr} . \tag{7.16}$$

Then

$$\partial_\mu \varphi_k(x) = D_{\mu,km} \varphi^m(x) = D_{\mu,k}^{\cdot m} \varphi_m(x), \tag{7.17}$$

$$\partial_\mu \varphi^k(x) = D_{\mu,m}^{k\cdot} \varphi^m(x) = D_{\mu}^{km} \varphi_m(x), \tag{7.18}$$

$$\partial_\mu F(x) = F^k D_{\mu,km} \varphi^m(x) = D_{\mu}^{km} F_k \varphi_m(x). \tag{7.19}$$

Clearly (7.14)–(7.16) can be used to obtain a variety of alternate forms of the examples (7.8)–(7.13).

Derivatives of higher order than the second can be computed, in general, by the formula

$$\partial_\alpha \partial_\beta \cdots \partial_\mu \partial_\lambda F(x) = C_{ij} C_{km} \cdots C_{pq} C_{rs} D_{\alpha}^{it} D_{\beta}^{kj} \cdots D_{\mu}^{rq} D_{\lambda}^{su} F_t \varphi_u(x). \tag{7.20}$$

It should be noted that the stronger conditions ($\bar{\text{a}}$), ($\bar{\text{b}}$), and ($\bar{\text{c}}$) (and even the weaker conditions (a), (b), (c)) are rarely satisfied in applications. However, we hope that the dimension G can be taken sufficiently large so that these conditions are satisfied in some approximate sense. If these conditions are not met, it is interesting to note that the derivative of the projection of a constant need not be zero!

8. Linear operators. As an extension of the ideas covered in Sec. 7 on derivatives, we now consider the properties of approximations to a linear operator $\mathfrak{L}$: $\mathfrak{F} \to \mathfrak{F}$. In general, the action of $\mathfrak{L}$ on Φ will be described by the matrix

$$L^{km} = (\mathfrak{L}\varphi^k, \varphi^m). \tag{8.1}$$

Then

$$\Pi\mathfrak{L}\varphi^k(x) = L^{km}\varphi_m(x). \tag{8.2}$$

In general

$$\mathfrak{L}f(x) = \mathfrak{L}F(x) + \mathfrak{L}\bar{f}(x) \tag{8.3}$$

and

$$\Pi\mathfrak{L}f(x) = \Pi\mathfrak{L}F(x) + \Pi\mathfrak{L}\bar{f}(x) = L^{km}F_k\varphi_m(x) + \Pi\mathfrak{L}\bar{f}(x). \tag{8.4}$$

Therefore, to simplify the analysis we once again assume commutativity of $\mathfrak{L}$ and Π

$$\Pi\mathfrak{L} = \mathfrak{L}\Pi. \tag{8.5}$$

It is understood, of course, that (8.5) may hold only approximately in many cases, the difference $\mathfrak{L}\Pi - \mathfrak{L}\Pi$ being a measure of how well the subspace Φ corresponds to the problem at hand. With the commutativity (8.5) in effect, we can set

$$\Pi\mathfrak{L}\bar{f} = \mathfrak{L}\Pi\bar{f} = 0 \tag{8.6}$$

and

$$\Pi\mathfrak{L}f(x) = L^{mk}F_m\varphi_k(x). \tag{8.7}$$

The adjoint operator $\hat{\mathfrak{L}}$ of an operator $\mathfrak{L}$ is defined by

$$(\mathfrak{L}f, g) = (\hat{\mathfrak{L}}g, f). \tag{8.8}$$

In our approximation,

$$L^{km}F_kG_m = (\Pi\mathfrak{L}F, G) = (\Pi\hat{\mathfrak{L}}G, F) = \hat{L}^{mk}F_kG_m \tag{8.9}$$

or

$$L^{km} = \hat{L}^{mk}. \tag{8.10}$$

Thus, as expected, the matrix of the adjoint operator is the transpose of the matrix of $\mathfrak{L}$.

Consider now the two eigenvalue problems

$$\mathfrak{L}f = \lambda f \quad \text{and} \quad \hat{\mathfrak{L}}g = \lambda g. \tag{8.11}$$

For our approximation, these equations can be represented by

$$(L^{km} - \lambda C^{km})F_m = 0 \tag{8.12}$$

and

$$(\hat{L}^{km} - \lambda C^{km})G_m = 0. \tag{8.13}$$

But it is well known that a matrix and its transpose have the same sets of eigenvalues. It is further possible to choose biorthogonal sets of eigenfunctions, these being uniquely defined when all eigenvalues are distinct. Thus, if $F_m^{(\alpha)}$ and $G_m^{(\beta)}$ denote eigenvectors corresponding to eigenvalues λ_α and λ_β of L^{km} and $\hat{L}^{km}$, then

$$L^{km}F_m^{(\alpha)} = \lambda_\alpha C^{kr}F_r^{(\alpha)}, \tag{8.14}$$

$$\hat{L}^{km}G_m^{(\beta)} = \lambda_\beta C^{kr}G_r^{(\beta)} \tag{8.15}$$

and we have

$$F_k^{(\alpha)}G_{(\beta)}^k = C^{mk}F_m^{(\alpha)}G_k^{(\beta)} = \delta_{\alpha\beta}\,. \tag{8.16}$$

In particular, *if the functions* $\varphi_k(x)$ *are eigenfunctions of a linear operator* $\mathfrak{L}$ *on* $\mathfrak{F}$, *the conjugate functions* $\varphi^k(x)$ *are eigenfunctions of the adjoint operator* $\hat{\mathfrak{L}}$. That is,

$$\mathfrak{L}\varphi_k(x) = \lambda_{(k)}\varphi_k(x) \Rightarrow \hat{\mathfrak{L}}\varphi^k(x) = \lambda_{(k)}\varphi^k(x). \tag{8.17}$$

If, then, $\mathfrak{L}$ is self-adjoint, the eigenfunctions coincide: $\varphi_k(x) = \varphi^k(x)$ (within a constant).

9. **Applications to finite-element approximations.** In order to apply the theory of conjugate approximations developed thus far to finite-element approximations, it is necessary that we identify the character of the base functions $\varphi_k(x)$ for general finite-element representations of the function $f(x)$. This requires some modifications in notation and the incorporation of additional structure in our approximations of $f(x)$.

Toward this end, we begin by representing the domain $\mathfrak{R}$ by a domain $\bar{\mathfrak{R}}$ which consists of a collection of E subdomains r_e, each usually being of relatively simple shape, connected appropriately together so as to approximate $\mathfrak{R}$. The subdomains r_e are called finite elements. The union of the disconnected subdomains is denoted $\mathfrak{R}^*$; i.e. $\mathfrak{R}^* = \bigcup_{e=1}^{E} r_e$.

The description of relationships between $\bar{\mathfrak{R}}$ and $\mathfrak{R}^*$ involves certain concepts which are fundamental to finite-element approximations. To proceed further, it is convenient at this point to introduce the idea of nodal points discussed previously. In finite-element approximations, however, we make a distinction between local nodal points identified in the finite elements and global nodal points in the connected domain $\bar{\mathfrak{R}}$. To distinguish further between local and global approximations we follow the notation of [1] and use upper-case Greek indices for discrete global quantities and upper-case Latin indices for discrete local quantities. For example, $a^N_{(e)}$ denotes the local coordinates of node N of element e while A^Δ denotes the global coordinates of node Δ of the connected model.

The *connectivity* of the finite-element model is established by E embeddings ${}^{(e)}\Lambda$ of local nodes in the disconnected element r_e into appropriate global nodes in $\bar{\mathfrak{R}}$ defined by the Boolean transformation

$${}^{(e)}\Lambda^\Delta_N a^N_{(e)} = A^\Delta, \tag{9.1}$$

where

$$\begin{aligned} {}^{(e)}\Lambda^\Delta_N &= 1 \quad \text{if node } \Delta \text{ of the connected model is incident with node } N \\ &\qquad\quad \text{of element } e \\ &= 0 \quad \text{if otherwise.} \end{aligned} \tag{9.2}$$

In (9.1) the repeated index N is to be summed from 1 to N_e, where N_e is the total number of nodes of element e, $\Delta = 1, 2, \cdots, G$, and, for simplicity, the local coordinates $\xi_{(e)}$ and global coordinates x are assumed to coincide. The mapping ${}^{(e)}\Lambda$ describes an embedding of r_e in $\bar{\mathfrak{R}}$. Note that $\Lambda = \sum^{(e)} \Lambda$ can be used to describe, alternatively, a mapping of $\mathfrak{R}^*$ into $\bar{\mathfrak{R}}$.

The transpose of ${}^{(e)}\Lambda^\Delta_N$ is denoted ${}^{(e)}\Omega^N_\Delta$ and defined a mapping of nodes in $\bar{\mathfrak{R}}$ into nodes in r_e:

$$a^N_{(e)} = {}^{(e)}\Omega^N_\Delta A^\Delta. \tag{9.3}$$

The definition of ${}^{(e)}\Omega^N_\Delta$ is thus the same as ${}^{(e)}\Lambda^\Delta_N$:

$$\begin{aligned} {}^{(e)}\Omega^N_\Delta &= 1 \quad \text{is node } N \text{ of element } e \text{ is coincident with node } \Delta \text{ of} \\ &\qquad\quad \text{the connected model} \\ &= 0 \quad \text{if otherwise.} \end{aligned} \tag{9.4}$$

Clearly, the arrays ${}^{(e)}\Omega_{\Delta}^{N}$ and ${}^{(e)}\Lambda_{N}^{\Delta}$, can be used to form the composition ${}^{e}\Lambda^{(e)}\Omega = I_{r_e}$, I_{r_e} being the identity matrix:

$$ {}^{(e)}\Lambda_{M}^{\Delta}\,{}^{(e)}\Omega_{\Delta}^{N} = \delta_{M}^{N} . \tag{9.5} $$

We now turn to the finite-element approximation of an element $f(x) \in \mathfrak{F}$. We recall that the projection Π: $\mathfrak{F} \rightarrow \Phi$ describes a G dimensional subspace of $\mathfrak{F}$, spanned by the functions $\varphi_{\Delta}(x)$ (denoted previously $\varphi_k(x)$), in which $f(x)$ is approximated by

$$ f(x) \doteq F(x) = F^{\Delta}\varphi_{\Delta}(x). \tag{9.6} $$

Let $P^{(e)}$ denote a projection of $\mathfrak{F}$ into an N_e-dimensional subspace $\Psi^{(e)}$ of functions with domain r_e which is defined by introducing a system of N_e linearly independent *local* base functions $\psi_{N}^{(e)}(\xi)$, ξ being local coordinates, with the properties

$$ \psi_{N}^{(e)}(a^{M}) = \delta_{N}^{M} \qquad \sum_{N=1}^{N_e} \psi_{N}^{(e)}(\xi) = 1. \tag{9.7} $$

Then locally

$$ f(x) \doteq \tilde{f}^{(e)}(\xi) = f_{(e)}^{N}\psi_{N}^{(e)}(\xi) \tag{9.8} $$

where

$$ f_{(e)}^{N} = \tilde{f}^{(e)}(\xi^{N}). \tag{9.9} $$

To relate the local functions $\psi_{N}^{(e)}(\xi)$ to the global functions $\varphi_{\Delta}(x)$, we require that $\psi_{N}^{(e)}(\xi)$ be the restriction of $\varphi_{\Delta}(x)$ to element r_e described by

$$ \psi_{N}^{(e)}(x) = {}^{(e)}\Lambda_{N}^{\Delta}\varphi_{\Delta}(x). \tag{9.10} $$

In (9.10) and in results to follow, we do not distinguish between ξ and x unless confusion is likely. This relation defines a mapping ${}^{(e)}\Lambda$: $\Phi \rightarrow \Psi^{e}$ and also follows from the fact that the values of $\psi_{N}^{(e)}(x)$ and $\varphi_{\Delta}(x)$ are in one-to-one correspondence with points in their domains r_e and $\mathfrak{R}$, which are related by (9.1) and (9.3). It follows that for $x \in r_e$, we also have the relation defined by

$$ \varphi_{\Delta}(x) = {}^{(e)}\Omega_{\Delta}^{N}\psi_{N}^{(e)}(x). \tag{9.11} $$

and, almost[2] everywhere in $\overline{\mathfrak{R}}$,

$$ \varphi_{\Delta}(x) = \sum_{e=1}^{E} {}^{(e)}\Omega_{\Delta}^{N}\psi_{N}^{(e)}(x). \tag{9.12} $$

Eqs. (9.10) and (9.11) define fundamental relationships between the local and global base functions which are needed in finite-element approximations. From the definitions of P^{e} and Π it is clear that

$$ P^{(e)} = {}^{(e)}\Lambda\Pi. \tag{9.13} $$

With the same provisions that apply to (9.12), we see that (9.6) can now be written

[2] This relation holds everywhere except on a set of Lebesgue measure zero, since at an inter-element boundary point common to m adjacent elements the value of (9.12) is m instead of unity. We shall nevertheless use formulas of this type in further developments, with the provision that a factor $1/m$ be introduced when boundary points are considered, because such sets of zero measure do not contribute to Lebesgue integrations over $\overline{\mathfrak{R}}$.

$$F(x) = F^{\Delta}\varphi_{\Delta}(x) = \sum_{e=1}^{E} F^{\Delta\ (e)}\Omega_{\Delta}^{N}\psi_{N}^{(e)}(x) = \sum_{e=1}^{E} f_{(e)}^{N}\psi_{N}^{(e)}(x) = \sum_{e=1}^{E} \tilde{f}^{\ (e)}(x), \tag{9.14}$$

wherein we have used the transformation

$$f_{(e)}^{N} = {}^{(e)}\Omega_{\Delta}^{N}F^{\Delta}. \tag{9.15}$$

With the basis $\varphi_{\Delta}(x)$ now described for the finite-element approximation, we can proceed to apply the theory of conjugate functions developed previously. Introducing (9.12) into (2.7), we see that the fundamental matrix of Φ for finite-element models is given by

$$C_{\Gamma\Delta} = (\varphi_{\Gamma}\ , \varphi_{\Delta}) = \sum_{e}^{E}\sum_{f}^{E} {}^{(e)}\Omega_{\Gamma}^{N}\ {}^{(e)}\Omega_{\Delta}^{M}(\psi_{N}^{(e)}, \psi_{M}^{(f)}). \tag{9.16}$$

Since, for the connected model,

$$(\psi_{N}^{(e)}, \psi_{M}^{(f)}) = 0 \qquad e \neq f \tag{9.17}$$

(9.16) can be written

$$C_{\Gamma\Delta} = \sum_{e}^{E} {}^{(e)}\Omega_{\Gamma}^{N}\ {}^{(e)}\Omega_{\Delta}^{M}c_{NM}^{(e)}\ , \tag{9.18}$$

where $c_{NM}^{(e)}$ is the local component of $C_{\Gamma\Delta}$ relative to element e:

$$c_{NM}^{(e)} = (\psi_{N}^{(e)}, \psi_{M}^{(e)}). \tag{9.19}$$

An important observation is that the local matrices $c_{NM}^{(e)}$ cannot be used to generate local conjugate functions in a manner analogous to that used earlier to compute $\varphi^{k}(x)$ since we have limited Φ (and Φ^{*}) to continuous functions. We have designed the functions $\psi_{N}^{(e)}(x)$ so as to give continuous base functions $\varphi_{\Delta}(x)$ on (9.5), but we have no reason to expect that $\sum_{e}(c_{NM}^{(e)})^{-1}\psi_{M}^{(e)}(x)$ will also be continuous on $\bar{\mathcal{R}}$.

We now direct our attention to the important problem of determining conjugate approximations for finite-element representations. We begin by considering a linear functional obtained by forming the inner product of the function

$$f(x) = F^{\Delta}\varphi_{\Delta}(x) = \sum_{e} {}^{(e)}\Omega_{\Delta}^{N}\psi_{N}^{(e)}(x)F^{\Delta} = F_{\Delta}\varphi^{\Delta}(x), \tag{9.20}$$

and an arbitrary function $G(x) \in \Phi$ which is also the sum of E local approximations $\tilde{g}^{(e)}(x)$:

$$G(x) \sum_{e} \tilde{g}^{(e)}(x) = G_{\Delta}\varphi^{\Delta}(x). \tag{9.21}$$

We have

$$(F, G) = F^{\Delta}G_{\Delta} = F_{\Delta}G^{\Delta} = F^{\Delta}(\sum_{e} {}^{(e)}\Omega_{\Delta}^{N}\psi_{N}^{(e)}(x), G), \tag{9.22}$$

wherein

$$G_{\Delta} = (G, \varphi_{\Delta}) = \sum_{e} {}^{(e)}\Omega_{\Delta}^{N}g_{N}^{(e)}, \tag{9.23}$$

and we have defined

$$g_{N}^{(e)} = (G, \psi_{N}^{(e)}). \tag{9.24}$$

We see that, unlike (9.15), the global values G_{Δ} at a global node Δ are obtained by summing all of the local values $g_{N}^{(e)}$ at local nodes incident on Δ.

Returning to (9.21), we can now define systems of local conjugate base functions, for

$$G(x) = \sum_{e} {}^{(e)}\Omega_{\Delta}^{N} g_{N}^{(e)} \varphi^{\Delta}(x) = \sum_{e} \tilde{g}^{(e)}(x). \tag{9.25}$$

Thus

$$\tilde{g}^{(e)}(x) = \psi_{(e)}^{N}(x) g_{N}^{(e)}, \qquad N = 1, 2, \cdots, N_e \tag{9.26}$$

where, from (9.25),

$$\psi_{(e)}^{N}(x) = {}^{(e)}\Omega_{\Delta}^{N} \varphi^{\Delta}(x) = {}^{(e)}\Omega_{\Delta}^{N} C^{\Delta\Gamma} \varphi_{\Gamma}(x) \tag{9.27}$$

or, in view of (9.12),

$$\psi_{(e)}^{N}(x) = {}^{(e)}\Omega_{\Delta}^{N} C^{\Delta\Gamma} \sum_{f=1}^{E} {}^{(f)}\Omega_{\Gamma}^{M} \psi_{M}^{(f)}(x). \tag{9.28}$$

The form of (9.28) is significant; it shows that *the "local" conjugate base functions for element e are linear combinations of the base functions* $\psi_{N}^{(f)}(x)$ *of all E finite elements. Thus, the functions* $\psi_{(e)}^{N}(x)$ *need not have local support; indeed, the support of each local function* $\psi_{(e)}^{N}(x)$ *is the entire connected domain* $\bar{\mathcal{R}}$. *This means that the usual procedure of calculating local values of conjugate approximations by taking local averages of the nodal values* $g_{N}^{(e)}$ (e.g., computing element stresses from a displacement approximation) *is not strictly correct. In order that the local conjugate approximation be consistent with the linear functional defined on* Φ (e.g. energy), *it is necessary that it be referred to a basis which has as its domain the entire collection of finite elements.* We shall demonstrate these properties of local conjugate approximations by means of an example in the following section.

Observe that

$$\begin{aligned} (\psi_{(e)}^{N}, \psi_{R}^{(e)}) &= {}^{(e)}\Omega_{\Delta}^{N} C^{\Delta\Gamma} \sum_{f=1}^{E} {}^{(f)}\Omega_{\Gamma}^{M} (\psi_{M}^{(f)}, \psi_{R}^{(e)}) \\ &= {}^{(e)}\Omega_{\Delta}^{N} C^{\Delta\Gamma} {}^{(e)}\Omega_{\Gamma}^{M} c_{MR}^{(e)} . \end{aligned} \tag{9.29}$$

Thus

$$\sum_{e=1}^{E} (\psi_{(e)}^{N}, \psi_{M}^{(e)}) {}^{(e)}\Omega_{\Delta}^{M} {}^{(e)}\Lambda_{N}^{\Gamma} = \delta_{\Delta}^{\Gamma} . \tag{9.30}$$

Also note that

$$\varphi^{\Delta}(x) = C^{\Delta\Gamma} \sum_{e=1}^{E} {}^{(e)}\Omega_{\Gamma}^{N} \psi_{N}^{(e)}(x). \tag{9.31}$$

With the base functions $\varphi_{\Delta}(x)$ and $\varphi^{\Delta}(x)$ determined for finite-element representations by (9.11) (or (9.12) and (9.31), it is now a simple matter to generate other quantities needed in finite-element approximations. For example, from (4.10) and (4.11),

$$M_{\Delta} = \int_{\mathcal{R}} \varphi_{\Delta}(x)\, d\mathcal{R} = \sum_{e=1}^{E} {}^{(e)}\Omega_{\Delta}^{N} m_{N}^{(e)} = \sum_{r=1}^{E} C_{\Delta\Gamma} , \tag{9.32}$$

where

$$m_{N}^{(e)} = \int_{\tau_e} \psi_{N}^{(e)}(x)\, d\tau_e , \tag{9.33}$$

$$M^{\Delta} = C^{\Delta\Gamma} \sum_{e=1}^{E} {}^{(e)}\Omega_{r}^{N} m_{N}^{(e)} = 1. \tag{9.34}$$

Also, if we define a local discrete operator by

$$d^{(e)}_{\mu,NM} = (\partial_\mu \psi^{(e)}_N, \psi^{(e)}_M), \tag{9.35}$$

then, according to (7.14) and (7.1),

$$D_{\mu,\Delta\Gamma} = \sum_{e=1}^{E} {}^{(e)}\Omega^N_\Delta \, {}^{(e)}\Omega^M_\Gamma \, d^{(e)}_{\mu,NM}, \tag{9.36}$$

$$D^{\Delta\Gamma}_\mu = C^{\Delta\Lambda} C^{\Gamma\Omega} \sum_{e=1}^{E} {}^{(e)}\Omega^N_\Lambda \, {}^{(e)}\Omega^M_\Omega \, d^{(e)}_{\mu,NM}. \tag{9.37}$$

Thus, for example,

$$\partial_\mu F(x) = \sum_e {}^{(e)}\Omega^N_\Delta \, {}^{(e)}\Omega^M_\Gamma \, d^{(e)}_{\mu,NM} F^\Delta \varphi^\Gamma(x). \tag{9.38}$$

Since $D^{\Delta\Gamma}_{\mu\Lambda}$ and $B^{\Delta\Gamma}_{\mu\Lambda}$ can be obtained immediately for finite-element models by introducing (9.37) into (7.6) and (7.7), it is a simple matter to write relations such as (7.8)–(7.13), (7.17)–(7.19), and (8.1) in forms appropriate for finite-element approximations. We omit these details here, but consider an example in Sec. 10.

10. Examples.

10.1 *Stress calculations.* To demonstrate the significance of the conjugate functions $\psi^N_{(e)}(x)$ described in the previous section, we present in this section a simple example involving the computation of stresses in a model based on approximate displacement fields.

Consider a nonhomogeneous bar for which the stress $\sigma(x)$ is given by the formula

$$\sigma(x) = k(x)(du(x)/dx). \tag{10.1}$$

Here $u(x)$ is the displacement field, and the modulus $k(x)$ is assumed to vary linearly according to

$$k(x) = k_0(1 + x), \tag{10.2}$$

k_0 being a material constant.

For simplicity, we shall employ a rather crude finite-element representation consisting of only three one-dimensional elements, each of unit length. To simplify matters further, we take for the local base functions $\psi^{(e)}_N(x)$, corresponding to a typical element e the linear forms

$$\psi^{(e)}_1(\xi) = 1 - \xi, \qquad \psi^{(e)}_2(\xi) = \xi, \tag{10.3}$$

ξ being a local coordinate, so that the local fundamental matrices are

$$c^{(e)}_{NM} = (\psi^{(e)}_N, \psi^{(e)}_M) = \frac{1}{6}\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}. \tag{10.4}$$

Determining, by inspection, the incidence operators ${}^{(e)}\Omega^N_\Delta$ and introducing (10.4) into (9.14), we get

$$C_{\Delta\Gamma} = \frac{1}{6}\begin{bmatrix} 2 & 1 & 0 & 0 \\ 1 & 4 & 1 & 0 \\ 0 & 1 & 4 & 1 \\ 0 & 0 & 1 & 2 \end{bmatrix} \quad \text{and} \quad C^{\Delta\Gamma} = \frac{6}{45}\begin{bmatrix} 26 & -7 & 2 & -1 \\ -7 & 14 & -4 & 2 \\ 2 & -4 & 14 & -7 \\ -1 & 2 & -7 & 26 \end{bmatrix}. \tag{10.5}$$

The conjugate base functions can now be computed with the aid of (8.25):

$$\begin{aligned}
\psi^1_{(1)}(x) &= \tfrac{6}{45}[26\psi_1^{(1)}(x) - 7(\psi_2^{(1)}(x) + \psi_1^{(2)}(x)) + 2(\psi_2^{(2)}(x) + \psi_1^{(3)}(x)) - \psi_2^{(3)}(x)],\\
\psi^2_{(1)}(x) &= \psi^1_{(2)}(x)\\
&= \tfrac{6}{45}[-7\psi_1^{(1)}(x) + 14(\psi_2^{(1)}(x) + \psi_1^{(2)}(x)) - 4(\psi_2^{(2)}(x) + \psi_1^{(3)}(x)) + 2\psi_2^{(3)}(x)],\\
\psi^2_{(2)}(x) &= \psi^1_{(3)}(x)\\
&= \tfrac{6}{45}[2\psi_1^{(1)}(x) - 4(\psi_2^{(2)}(x) + \psi_1^{(2)}(x)) + 14(\psi_2^{(2)}(x) + \psi_1^{(3)}(x)) - 7\psi_2^{(3)}(x)],\\
\psi^2_{(3)}(x) &= \tfrac{6}{45}[-\psi_1^{(1)}(x) + 2(\psi_2^{(1)}(x) + \psi_1^{(2)}(x)) - 7(\psi_2^{(2)}(x) + \psi_1^{(3)}(x)) + 26\psi_2^{(3)}(x)].
\end{aligned} \tag{10.6}$$

The forms of the functions $\psi_N^{(e)}(x)$ and $\psi^N_{(e)}(x)$ are shown in Fig. 1.

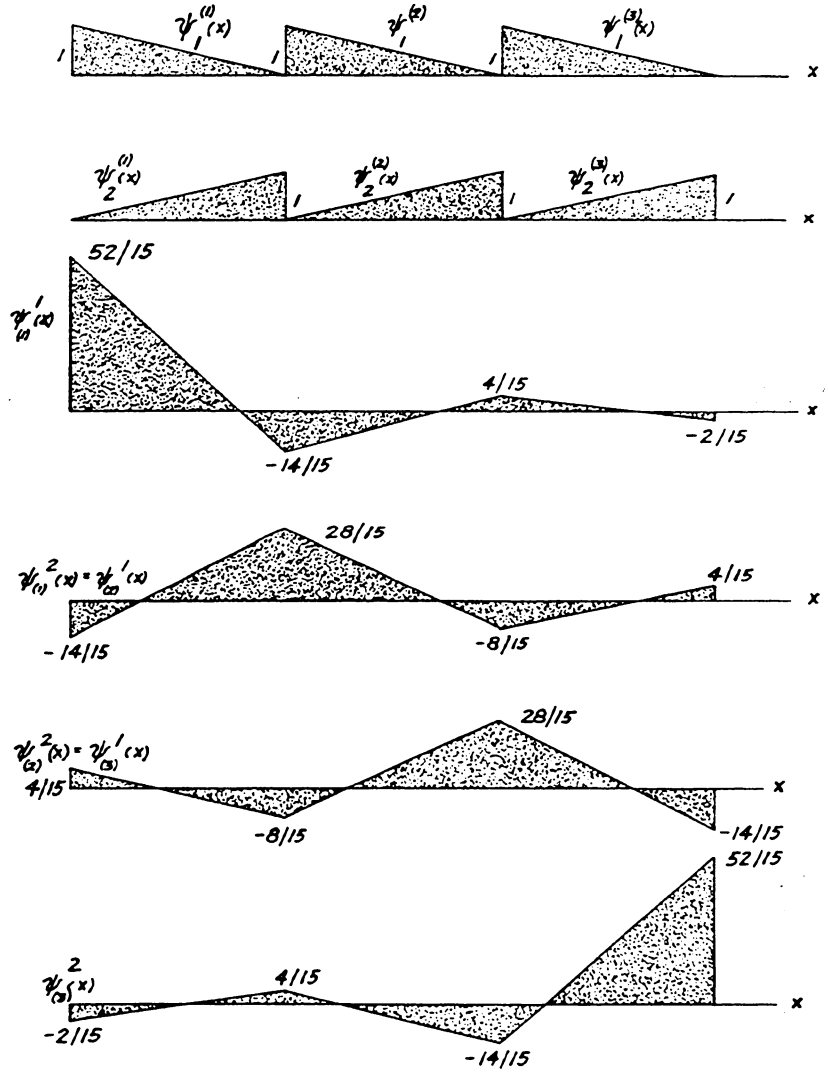


FIG. 1. Local base functions $\psi_N^{(e)}(x)$ and their conjugate base functions $\psi^N_{(e)}(x)$.

We shall now assume that the bar is given a prescribed quadratic displacement field of the form

$$u(x) = \alpha(1 - (x/3)^2), \tag{10.7}$$

where α is a small constant. Clearly, the exact stress distribution is

$$\sigma(x) = (-2\alpha k_0/9)x(1 + x). \tag{10.8}$$

However, the displacement field, as represented by the finite-element model, is piecewise linear:

$$U(x) = (\alpha/9)[9\varphi_1(x) + 8\varphi_2(x) + 5\varphi_3(x)], \tag{10.9}$$

where

$$\varphi_1(x) = \psi_1^{(1)}(x), \quad \varphi_2(x) = \psi_2^{(1)}(x) + \psi_1^{(2)}(x), \quad \varphi_3(x) = \psi_2^{(2)}(x) + \psi_1^{(3)}(x). \tag{10.10}$$

If the usual procedure for computing stresses in finite elements is used, we simply introduce (10.9) into the constitutive equation (10.1) for each element. This results in a discontinuous stress distribution which exhibits a finite discontinuity at the juncture of each element (Fig. 2). Further, the maximum stress computed in this manner is 16.7 percent in error.

A quite different profile is obtained if the proper conjugate approximations are used. Introducing (10.9) into (10.1) gives, as before, a local stress field $\sigma_N^{(e)}(x)$ for each element. The conjugate (nodal) components $\sigma_N^{(e)}$ are then obtained with the aid of (8.21):

$$\sigma_N^{(e)} = (\sigma^{(e)}(x), \psi_N^{(e)}(x)). \tag{10.11}$$

Therefore, the conjugate-function representation of stress is given by

$$\sigma(x) = \sum_{e=1}^{3} \sigma_N^{(e)} \psi_{(e)}^N(x), \qquad N = 1, 2. \tag{10.12}$$

This conjugate stress profile is shown in Fig. 2 along with the exact solution and the discontinuous distribution obtained using common procedures. We see that the distribution obtained using conjugate approximation functions is continuous at the junction of adjacent elements and that it indicates a maximum stress which is less than 6.5 percent in error.

10.2 *Piecewise linear approximation functions of one variable.* As an explicit but simple example to the foregoing discussion, consider the piecewise linear approximation functions arising by dividing an interval I in N equal subintervals of length h, and requiring $\varphi_k(x)$ to vanish at all nodes except node $x_k = hk$, as shown in Fig. 3. Then $M_0 = M_N = \int_{\mathcal{R}} \varphi_0 \, d\mathcal{R} = h/2$, $M_k = \int_{\mathcal{R}} \varphi_k \, d\mathcal{R} = h$ for $k = 1, \cdots, N - 1$ and the fundamental matrix is

$$C_{km} = \int_{\mathcal{R}} \varphi_k \varphi_m \, d\mathcal{R} = \frac{h}{6} \begin{bmatrix} 2 & 1 & 0 & \cdots & 0 \\ 1 & 4 & 1 & \cdots & 0 \\ 0 & 1 & 4 & \cdots & 0 \\ & & \cdots & & \\ 0 & \cdots & 4 & 1 & 0 \\ 0 & \cdots & 1 & 4 & 1 \\ 0 & \cdots & 0 & 1 & 2 \end{bmatrix} \tag{10.13}$$

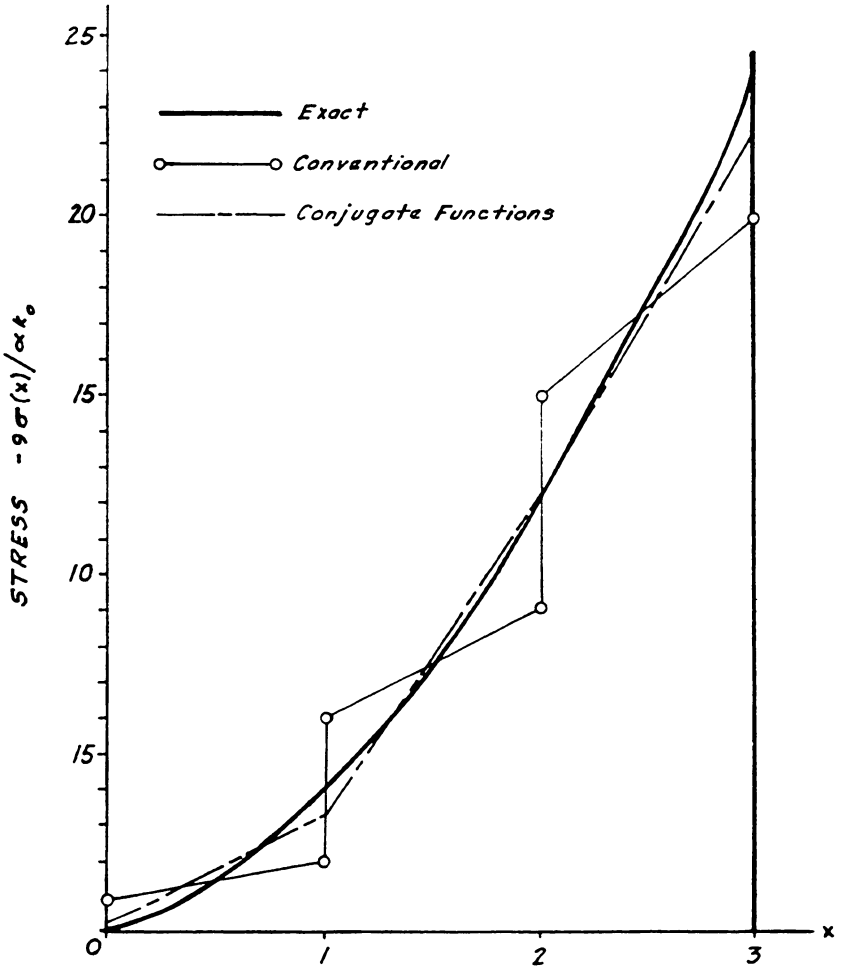


FIG. 2. Comparison of stress distributions computed conventionally with conjugate function approximation.

Define the n by n determinants

$$A_n = \begin{vmatrix} 4 & 1 & & & & & \\ 1 & 4 & 1 & & \mathbf{0} & & \\ & 1 & \cdot & & & & \\ & & & \cdot & & 1 & \\ & & \mathbf{0} & & 1 & 4 & 1 \\ & & & & & 1 & 4 \end{vmatrix}$$

satisfying the recurrence relation

$$A_{n+1} - 4A_n + A_{n-1} = 0. \tag{10.15}$$

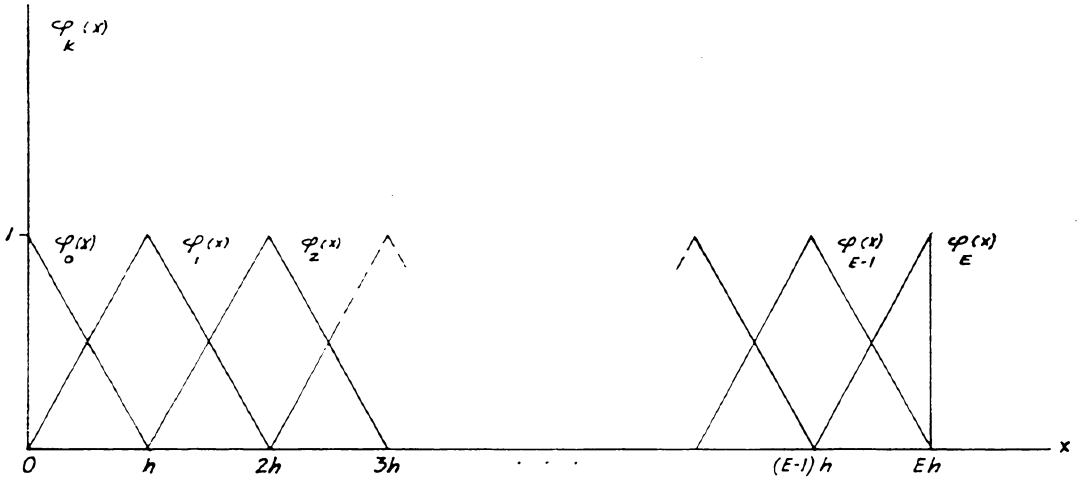


FIG. 3. Base functions $\varphi_k(x)$ for finite-element representation of interval on real axis.

Since

$$B_n = \begin{vmatrix} 2 & 1 & 0 & & & & \\ 1 & 4 & 1 & & O & & \\ 0 & 1 & 4 & & & & \\ & & & \cdot & & & \\ & & & & \cdot & & \\ & & & & & \cdot & \\ & & O & & & 4 & 1 \\ & & & & & 1 & 4 \end{vmatrix}, \quad C_n = \begin{vmatrix} 2 & 1 & 0 & & & & \\ 1 & 4 & 1 & & O & & \\ 0 & 1 & 4 & & & & \\ & & & \cdot & & & \\ & & & & \cdot & & \\ & & & & & \cdot & \\ & & O & & & 4 & 1 \\ & & & & & 1 & 2 \end{vmatrix} \tag{10.16}$$

are linear combinations of A_k

$$B_n = 2A_{n-1} - A_{n-2} = A_n - 2A_{n-1}\,, \tag{10.17}$$

$$C_n = B_n - 2B_{n-1} = 3A_{n-2}\,. \tag{10.18}$$

They also satisfy (10.15). Now (10.15) admits two independent solutions in geometric series $a_k = \alpha^k$, $b_k = \alpha^{-k}$ where $\alpha = 2+\sqrt{3} = 3.732050808$ and $(\alpha^{-1} = 2-\sqrt{3} = 0.267949191)$ satisfies the quadratic equation

$$\alpha^2 - 4\alpha + 1 = 0. \tag{10.19}$$

A_k and B_k are then linear combinations of α^k and α^{-k}:

$$A_k = \gamma(\alpha^k - \alpha^{-k-2}) \qquad B_k = \tfrac{1}{2}(\alpha^k + \alpha^{-k}), \tag{10.20}$$

where $\gamma = \alpha/(\alpha - \alpha^{-1}) = 1.077350269$. The beginnings of the series A_k, B_k are represented in Table 1 below in which the relation

$$\Delta^2 A_k = A_{k+1} - 2A_k + A_{k-1} = 2A_k \tag{10.21}$$

has been used.

TABLE 1

k	A_k	ΔA_k	$\Delta^2 A_k$	B_k	ΔB_k	$\Delta^2 B_k$
−2	−1			7		
		1			−5	
−1	0		0	2		4
		1			−1	
0	1		2	1		2
		3			1	
1	4		8	2		4
		11			5	
2	15		30	7		14
		41			15	
3	56		112	26		52
		153			71	
4	209			97		

Now the inverse of the fundamental matrix is

$$C^{km} = (-1)^{k+m}(2/h)(B_k B_{N-m}/A_{N-1}), \qquad k \leq m. \tag{10.22}$$

The elements for $k > m$ being given by symmetry, $C^{km} = C^{mk}$. As an example, for $N = 6$

$$C^{km} = \frac{1}{390h}\begin{bmatrix} 1352 & -362 & 97 & -26 & 7 & -2 & 1 \\ -362 & 724 & -194 & 52 & -14 & 4 & -2 \\ 97 & -194 & 679 & -182 & 49 & -14 & 7 \\ -26 & 52 & -182 & 676 & -182 & 52 & -26 \\ 7 & -14 & 49 & -182 & 679 & -194 & 97 \\ -2 & 4 & -14 & 52 & -194 & 724 & -362 \\ 1 & -2 & 7 & -26 & 97 & -362 & 1351 \end{bmatrix} \tag{10.23}$$

The corresponding conjugate functions are shown in Fig. 4. As a check, we can verify $M^m = M_k C^{km} = 1 = \int_{\mathcal{R}} \varphi^m(x)^2 \, d\mathcal{R}$. Using (10.20), (10.21) becomes

$$C^{km} = (-1)^{k+m}\frac{\sqrt{3}}{h}\frac{(\alpha^k + \alpha^{-k})(\alpha^{G-m} + \alpha^{-G+m})}{\alpha^G - \alpha^{-G-2}}, \qquad k \leq m. \tag{10.24}$$

or in the limit $G \to \infty$, $k \simeq m \simeq N/2$

$$C^{km} \simeq (-1)^{|k-m|}(\sqrt{3}/h)\alpha^{-|k-m|}, \qquad k, m \text{ arbitrary}, \tag{10.25}$$

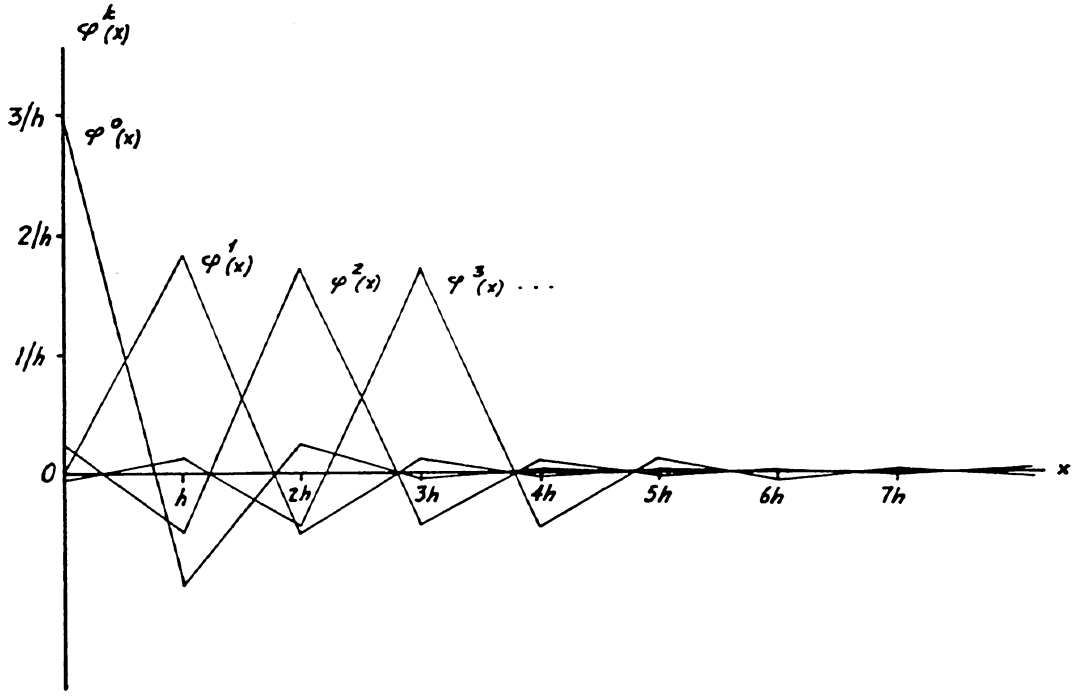


FIG. 4. Conjugate functions corresponding to base functions in Fig. 3.

or

$$C^{km} \approx \frac{\sqrt{3}}{h} \begin{bmatrix} \cdot & & & & & & \\ & \cdot & & & & & \\ & & \cdot & & & & \\ & & & 1 & -\alpha^{-1} & \alpha^{-2}-\alpha^{-3} & \cdot \\ & & & -\alpha^{-1} & 1 & -\alpha^{-1} & \alpha^{-2}-\alpha^{-3} \\ & & & \cdot & -\alpha^{-1} & 1 & -\alpha^{-1} \\ & & & \cdot & \cdot & -\alpha^{-1} & 1 \end{bmatrix} \tag{10.26}$$

For an analytic function $f(x) = \sum_0^\infty a_n x^n$ we may obtain explicit formulas for the components F_k. Since for $0 < k < N$

$$\int_{\mathfrak{R}} x^n \varphi_k \, d\mathfrak{R} = (h^{n+1}/(n+1)(n+2))\, \Delta^2 k^{n+2}, \tag{10.27}$$

we obtain

$$F_k = (1/h)\, \Delta^2 f^{(-2)}(kh), \tag{10.28}$$

where $f^{(-2)} = \sum_0^\infty (a_n x^{n+2}/(n+1)(n+2))$ is a second antiderivative of $f(x)$ and Δ^2

denotes a second difference operator. The corresponding formulas at the boundaries are

$$F_0 = (1/h)f^{(-2)}(h), \qquad F_N = f^{(-1)}(Nh) - (1/h)\,\Delta f^{(-2)}(Nh), \tag{10.29}$$

$f^{(-1)}$ being a first antiderivative, Δ the first backward difference. In the limit $N \to \infty$ (10.28) may be replaced by

$$F_k = hf(kh). \tag{10.30}$$

Since the derivatives $\partial\varphi_k(x)$ are discontinuous, the commutativity condition $\partial\Pi = \Pi\partial$ (see Theorem 7.1) is clearly violated. Yet it makes sense to introduce matrices D_{km} and B_{km} in analogy to (7.1) and (7.7),

$$D_{km} = \int_{\mathcal{R}} \varphi_k\, \partial\varphi_m\, d\mathcal{R}, \qquad B_{km} = \int_{\mathcal{R}} \partial\varphi_k\, \partial\varphi_m\, d\mathcal{R}. \tag{10.31}$$

While B_{mk} is symmetric, D_{km} is almost antisymmetric, because

$$\int_{\mathcal{R}} \varphi_k\, \partial\varphi_m\, d\mathcal{R} = -\int_{\mathcal{R}} \varphi_m\, \partial\varphi_k\, d\mathcal{R} \quad \text{for} \quad 0 < k, m < N.$$

Numerically

$$D_{km} = \begin{bmatrix} -1 & 1 & 0 & 0 & & & \\ -1 & 0 & 1 & 0 & & & \\ 0 & -1 & 0 & 1 & & & \\ & & & \cdot & & & \\ & & & & \cdot & & \\ & & & & & & \\ & & & & -1 & 0 & 1 \\ & & & & 0 & -1 & 1 \end{bmatrix}$$

$$B_{km} = \frac{1}{h}\begin{bmatrix} 1 & -1 & 0 & 0 & & & \\ -1 & 2 & -1 & 0 & & & \\ 0 & -1 & 2 & -1 & & & \\ & & & \cdot & & & \\ & & & & \cdot & & \\ & & & & & & \\ & & & & -1 & 2 & -1 \\ & & & & 0 & -1 & 1 \end{bmatrix}$$

Now (7.13), i.e.,

$$\int_{\mathcal{R}} \partial f\, \partial g\, d\mathcal{R} \doteq B_{km}F^kG^m \tag{10.33}$$

is a reasonable approximation, while (7.5), i.e.,

$$\partial f \doteq D_{km}F^k\varphi^m(x) \tag{10.34}$$

will be rather crude. This is illustrated by Fig. 5, where D_{km} is applied to a base func-

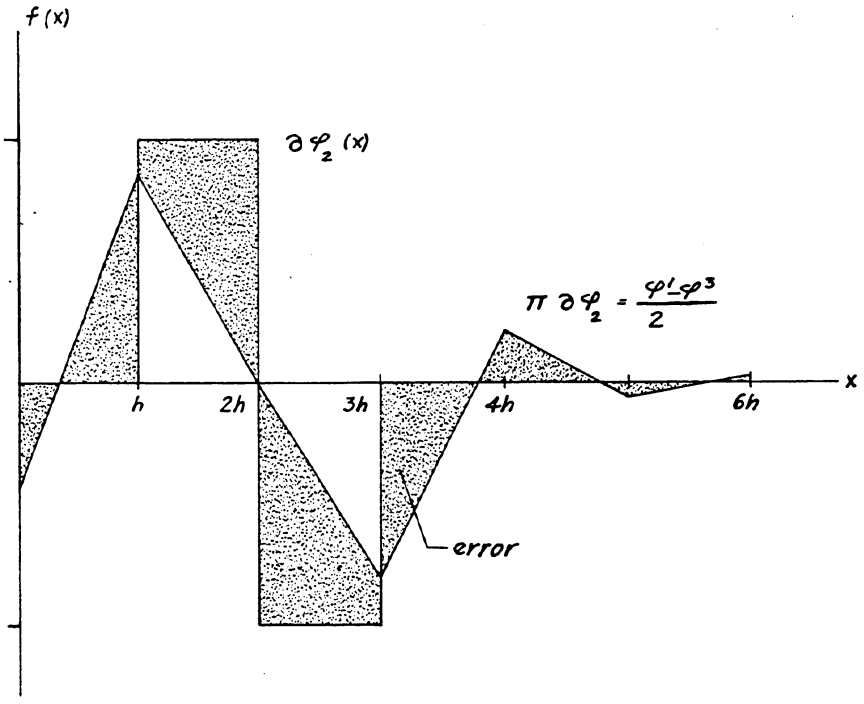


FIG. 5. Projection of derivative.

tion $\varphi_2(x)$. This explains why computing B_{km} out of D_{km} by (7.7) would give a false result. On the other hand,

$$\int_{\mathfrak{R}} \varphi_k \, \partial^2 \varphi_m \, d\mathfrak{R} = -\int_{\mathfrak{R}} \partial\varphi_k \, \partial\varphi_m \, d\mathfrak{R}, \qquad 0 < k, m < N, \tag{10.35}$$

suggests that $-B_{km}$ may be used to compute second derivatives

$$\partial^2 f(x) \doteq -B_{km} F^k \varphi^m(x) \tag{10.36}$$

instead of (7.6), (7.8). In fact,

$$\begin{aligned} \Pi \, \partial^2 \varphi_k &= \Pi[(1/h)\, \delta(x - kh - h) - (2/h)\, \delta(x - kh) + (1/h)\, \delta(kh + h)] \\ &= (1/h)(\varphi^{k-1} - 2\varphi^k + \varphi^{k+1}) \end{aligned}$$

yields

$$\int_{\mathfrak{R}} \varphi_k \, \partial^2 \varphi_m \, d\mathfrak{R} = -B_{km} \, . \tag{10.37}$$

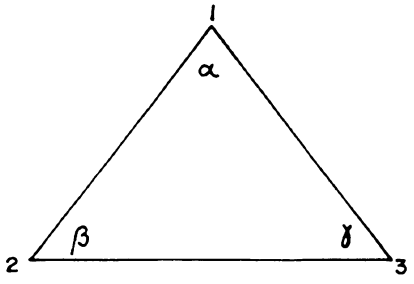
10.3 *A two-dimensional example.* Essentially the same procedure outlined previously can be used for two- and three-dimensional finite elements. As a final example, we outline briefly the construction of the conjugate approximation functions corresponding to a two-dimensional network of triangular elements.

Consider a triangular element in the x_1 , x_2-plane, the vertices of which are the local nodal points. The local interpolation functions $\psi_N^{(e)}(x)$, wherein $x = (x_1 , x_2)$, are linear functions of x_1 and x_2 and satisfy $\psi_N^{(e)}(a^M) = \delta_N^M$; $M, N = 1, 2, 3$. Introducing these

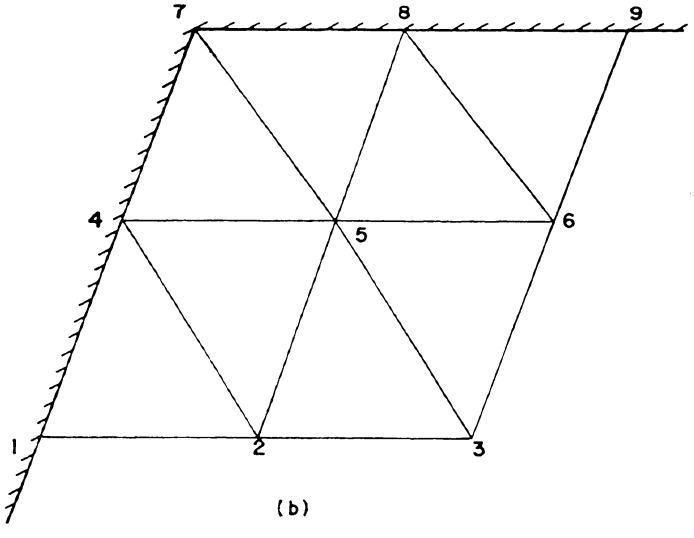
functions into (9.19), we obtain for the local component of the fundamental matrix $C_{\Gamma\Delta}$,

$$c_{NM}^{(e)} = \int_A \psi_N^{(e)}(x)\psi_M^{(e)}(x)\, dA = \frac{A}{12}\begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix} \tag{10.38}$$

wherein A is the area of the triangle. We observe that (10.38) is independent of the included angles α, β, γ formed by sides of the triangle. However, discrete models of various differential operators may depend on these angles; for example, for the triangle shown in Fig. 6a,



(a)



(b)

FIG. 6. Two-dimensional network of triangular elements.

$$d_{NM}^{(e)} = \int_A \text{grad } \psi_N^{(e)}(x) \text{ grad } \psi_M^{(e)}(x)\, dA$$

$$= \frac{1}{2}\begin{bmatrix} \cot\beta + \cot\gamma & -\cot\gamma & -\cot\beta \\ -\cot\gamma & \cot\gamma + \cot\alpha & -\cot\alpha \\ -\cot\beta & -\cot\alpha & \cot\alpha + \cot\beta \end{bmatrix}. \tag{10.39}$$

To demonstrate the character of the conjugate approximation functions for a specific finite-element representation, consider the network shown in Fig. 6b. In this case, we

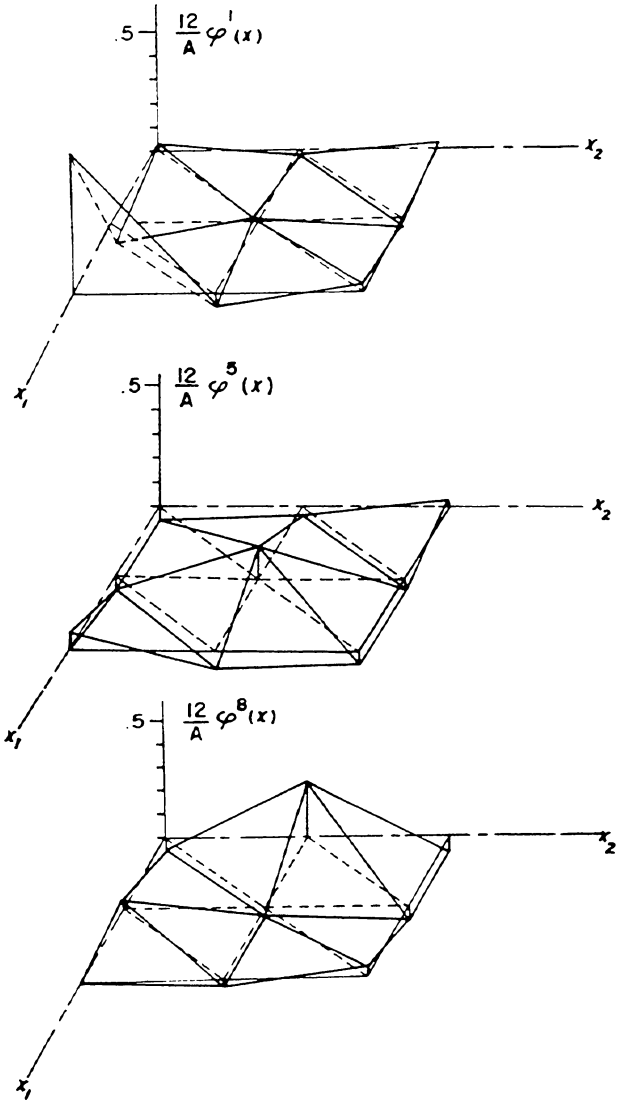


FIG. 7. Representative conjugate functions.

have from (9.16) or (9.18),

$$C_{\Delta\Gamma} = (\varphi_\Delta , \varphi_\Gamma) = \frac{A}{12}\begin{bmatrix} 2 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ & 6 & 1 & 2 & 2 & 0 & 0 & 0 & 0 \\ & & 4 & 0 & 2 & 1 & 0 & 0 & 0 \\ & & & 6 & 2 & 0 & 1 & 0 & 0 \\ & & & & 12 & 2 & 2 & 2 & 0 \\ & & & & & 6 & 0 & 2 & 1 \\ & \text{Symmetric} & & & & & 4 & 1 & 0 \\ & & & & & & & 6 & 1 \\ & & & & & & & & 2 \end{bmatrix} \tag{10.39}$$

A being the area of an element. Inverting this matrix and making use of (3.7), we obtain the conjugate approximation functions $\varphi^\Delta(x)$. Since, in the present example, the functions $\varphi_\Delta(x)$ are linear, $\varphi^\Delta(x)$ are also piecewise linear and it is sufficient to merely calculate the values of the conjugate functions at each node. Rather than to write out the entire collection of functions, we cite as representative examples the nodal values

$$\begin{aligned} \varphi^1(A^\Delta) &= \frac{12}{A}\,[.580, -.080, .009, -.080, .027, -.009, .009, -.009, .009] \\ \varphi^5(A^\Delta) &= \frac{12}{A}\,[.027, -.027, -.045, -.027, .116, -.027, -.045, -.027, .027] \\ \varphi^8(A^\Delta) &= \frac{12}{A}\,[-.009, .000, .027, .018, -.027, -.054, -.045, .214, -.080] \end{aligned} \tag{10.40}$$

These functions are illustrated in Fig. 7.

Acknowledgment. The research reported in this paper was supported through Contract F44620-69-C-0124 under Project Themis at the University of Alabama Research Institute.

References

[1] J. T. Oden, *A general theory of finite elements.* I: *Topological considerations,* Internat. J. Numerical Methods Engrg. 2, 205–221 (1969)

[2] J. S. Archer, *Consistent matrix formulation for structural analysis using finite element techniques,* AIAA J. 3, 1910–1918 (1965)

[3] A. E. Taylor, *Introduction to functional analysis,* Wiley, New York, 1958