Hierarchal Vector Basis Functions of Arbitrary Order for Triangular and Tetrahedral Finite Elements

Jon P. Webb, *Member, IEEE*

***Abstract*—New vector finite elements are proposed for electromagnetics. The new elements are triangular or tetrahedral edge elements (tangential vector elements) of arbitrary polynomial order. They are hierarchal, so that different orders can be used together in the same mesh and p-adaption is possible. They provide separate representation of the gradient and rotational parts of the vector field. Explicit formulas are presented for generating the basis functions to arbitrary order. The basis functions can be used directly or after a further stage of partial orthogonalization to improve the matrix conditioning. Matrix assembly for the frequency-domain curl–curl equation is conveniently carried out by means of universal matrices. Application of the new elements to the solution of a parallel-plate waveguide problem demonstrates the expected convergence rate of the phase of the reflection coefficient, for tetrahedral elements to order 4. In particular, the full-order elements have only the same asymptotic convergence rate as elements with a reduced gradient space (such as the Whitney element). However, further tests reveal that the optimum balance of the gradient and rotational components is problem-dependent.**

***Index Terms*—Basis functions, finite-element methods.**

I. Introduction

In solving vector electromagnetic problems by the finite-element method, it is generally recognized that the simple approach of treating each Cartesian component of the vector field as a scalar function does not work well. There are difficulties in handling the interface conditions at boundaries between materials; modeling of field singularities at sharp edges and corners is very poor; and spurious (nonphysical) modes appear among the computed eigenvalues. These problems can be avoided by the use of *edge elements* (also called tangential vector finite elements), which interpolate the field in such a way that tangential continuity between adjacent elements is enforced, while the normal component is allowed to be discontinuous. This relaxation of continuity is usually sufficient to eliminate spurious modes (though further precautions are needed for brick elements [1], [2]); it also greatly facilitates the imposition of correct boundary and interface conditions, and improves substantially the field modeling around singularities [3].

The most widely used edge element is the Whitney element [4]. The tetrahedral version has six degrees of freedom, one per edge. The Whitney element is simple and cheap to assemble, and has degrees of freedom which have a ready physical interpretation (the line integral of the tangential field along the edges). However, it provides a field interpolation that is not even first-order: a linear field such as $\boldsymbol{H} = x\hat{\boldsymbol{y}}$ cannot be represented exactly. Consequently, convergence of the field solution to the correct values as the mesh is refined is comparatively slow, and large numbers of elements may be needed. Recognition of that fact has led to the invention of a variety of edge elements that are at least fully first-order and usually involve second-order polynomials [5]–[10]. Though more complicated, these are more accurate, and some have found application in widely used commercial software.

Then the question arises, as it did in the scalar case, can we develop a general, pth-order edge element, i.e., formulas that would allow us, in principle, to build an edge element of any order? Since convergence rates increase with p, it should be computationally advantageous to use the highest p possible in a given problem. Rather than inventing each successive order of element in an *ad hoc* way, it would be better to have general expressions from which any order of element could be obtained. Furthermore, once a range of high-order elements is available, *p-adaption* is possible, i.e., iterative increase of the element orders in different regions of the problem until convergence of the field to a specified accuracy is achieved [11]. Experience with the scalar wave equation suggests that p-adaption is computationally very efficient [12]; in combination with the more conventional h-adaption (mesh refinement), it can lead to exceptional performance [13].

The first problem one faces when considering the pth-order element is, what function space should it provide? The obvious answer is the space of vector functions *complete to order* p, i.e., all functions that are polynomials of degree no higher than p in the space coordinates x, y, and z. However, there is a good reason to believe that this choice is not necessarily optimal. In vector electromagnetics, the curl of the field is often as important as the field itself. If the field is represented as a polynomial of order p, its curl will be a polynomial of order $p-1$, and the overall convergence of the solution will be dominated by this lower order. Why not, therefore, remove those degrees of freedom that do not affect the curl—the *gradient* degrees of freedom—while keeping the field order complete to order $p-1$? The result is an element with fewer degrees of freedom, but with a better balance in the accuracy of representation of the field and its curl. The Whitney element mentioned above is the result of applying this idea to an element complete to first order.

Manuscript received October 30, 1998; revised June 29, 1999. This work was supported by the Natural Sciences and Engineering Research Council of Canada.

The author is with the Department of Electrical and Computer Engineering, McGill University, Montreal, PQ, H3A 2A7, Canada.

Publisher Item Identifier S 0018-926X(99)07953-3.

This approach was introduced by Nedelec [14] in the earliest paper on high-order edge elements. However, a few years later he published another paper [15] in which the function spaces are now complete to order p. The debate between proponents of complete-order elements and those who favor a removal of some gradient functions has continued. The argument will be made in this paper that the optimum function space is, in fact, problem-dependent. Consequently, two series of function spaces are introduced: one representing gradient (irrotational) functions, with zero curl; and one representing *rotational* functions, i.e., functions with a nonzero curl. The general high-order element has, then, *two* orders, for the gradient and rotational parts separately. When these two orders are the same, the result is an element which is complete to pth order; when the gradient order is one less than the rotational order, the earlier Nedelec spaces are obtained. Other combinations are also possible.

In fact, there are other advantages to maintaining the separation of gradient and rotational spaces in higher order elements. Efficient low-frequency formulations, such as those used to compute the induced eddy currents in solid conductors [16], depend on the separation of the two spaces. So does a promising new technique for solving the linear equation system generated by the application of a second-order edge element to the vector wave equation in the frequency domain [17].

Having determined the spaces (Section II, below), the next problem is to decide on basis functions for the spaces. Nedelec's answer is incomplete. He defines degrees of freedom in terms of projections of the field onto complete-order polynomial spaces, but does not give bases for these spaces. Graglia *et al*. [18] describe a general pth-order element of the reduced-gradient kind, and provide an interpolatory basis for it, i.e., a set of basis functions and an associated set of points in the element, such that each basis function vanishes at all the points except one. Yioultsis and Tsiboukis [19] give a procedure for generating essentially the same kind of element, and earlier Wang and Ida [20] took the same approach with complete-order elements. The reduced-gradient element of Sun *et al*. [21] is also partly interpolatory. Interpolatory bases have certain advantages (good linear independence, ease of physical interpretation of the unknowns) and have been widely used for scalar elements [22], but have the decided drawback of being nonhierarchal. The basis functions of the pth-order element are all different from those of the lower order elements, and so mixing different orders within the same mesh while preserving tangential continuity is next to impossible. That rules out p-adaption. Anderson and Volakis [23] recently described a hierarchal basis for the reduced-gradient series, with explicit expressions up to order 3. In Section III, general expressions for pth-order hierarchal basis functions are given. Though these could be used directly, in Section IV it is argued that a partial orthogonalization will improve the matrix conditioning. Section V shows how the finite-element matrices for the frequency-domain curl–curl equation can be assembled efficiently for the new elements, using universal matrices. Results for two test problems are presented in Section VI.

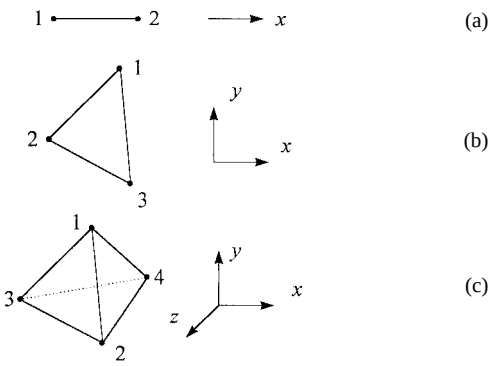


Fig. 1. (a) One-dimensional (1-D) (line). (b) Two-dimensional (2-D) (triangular). (c) Three-dimensional (3-D) (tetrahedral) finite elements.

II. Function Spaces

A. One Dimension: Line Element

The line element is shown in Fig. 1. The normalized coordinates for this simplex are ζ_1 and ζ_2, satisfying the relationship $\zeta_1+\zeta_2=1$. At node 1, $\zeta_1=0$; at node 2, $\zeta_1=1$.

Let V_p be the space of all one–dimensional (1-D) vector functions on this element, that are polynomials in ζ_1 and ζ_2 of degree less than or equal to p. Since there are $p+1$ linearly independent scalar polynomials of degree p, the dimension of V_p is $p+1$. Note that, because of the dependence of ζ_2 on ζ_1, a function in V_p can be written in any number of different ways; in particular, each can be written in the form

$$P_1(\zeta_1,\zeta_2)\zeta_2\nabla\zeta_1+P_2(\zeta_1,\zeta_2)\zeta_1\nabla\zeta_2 \tag{1}$$

where P_1 and P_2 are polynomials of degree $p-1$. The gradient operator here is just $\hat{\boldsymbol{x}}\frac{\partial}{\partial x}$.

Let G_p be the subspace of V_p containing those vectors that are gradients of scalars vanishing at nodes 1 and 2. Since there are p linearly independent scalar polynomials of degree $p+1$, vanishing at $\zeta_1=0$ and $\zeta_1=1$, the dimension of G_p is p. Again, functions in G_p can be written in a number of different ways, including

$$\nabla\big(\zeta_1\zeta_2P(\zeta_1,\zeta_2)\big) \tag{2}$$

where P is a polynomial of degree $p-1$.

To make up the difference between G_p and V_p, a one-dimensional space W_p is needed, having no functions in common with G_p (other than zero). A convenient basis for this space is a constant function, $-\nabla\zeta_1$ say, which when written in form (1) becomes

$$\zeta_1\nabla\zeta_2-\zeta_2\nabla\zeta_1. \tag{3}$$

B Two Dimensions: Triangular Element

The triangular element is shown in Fig. 1. The normalized coordinates for this simplex are ζ_1, ζ_2, and ζ_3, satisfying the relationship $\zeta_1+\zeta_2+\zeta_3=1$. At node 1, $\zeta_1=1$; on the opposite edge (connecting nodes 2 and 3), $\zeta_1=0$. The linear relationship between simplex coordinates and Cartesian coordinates (x,y) is given by (for example) Silvester and Ferrari [24].

Let V_p be the space of all 2-D vector functions on this element, which are polynomials in ζ_1, ζ_2, and ζ_3 of degree less than or equal to p. Since there are $(p+1)(p+2)/2$ linearly independent scalar polynomials of degree p in two dimensions, the dimension of V_p is $(p+1)(p+2)$.

Consider a basis for the 1-D V_p space. If each basis function is written in the form (1), it may be regarded as a 2-D vector function in the 2-D V_p space, with these properties:

a) its tangential component vanishes on the edge $\zeta_1 = 0$ and on the edge $\zeta_2 = 0$ (remember that $\nabla\zeta_i$ is perpendicular to edge $\zeta_i = 0$);
b) its tangential component on the edge $\zeta_3 = 0$ is identical to the 1-D basis function.

As a consequence of b), the 2-D functions generated in this way must be linearly independent, and constitute a basis for a subspace of dimension $p+1$. This is a space $V_p^{(e)}$ of *edge* functions, i.e., functions associated with the edge $\zeta_3 = 0$ and vanishing tangentially on the other two edges. By cyclically permuting the indices in (1) $(1 \to 2 \to 3 \to 1)$, two similar spaces can be obtained, one for edge $\zeta_1 = 0$ and one for edge $\zeta_2 = 0$.

$V_p^{(e)}$ can be divided into two parts. Consider a basis for the 1-D G_p space. Each basis function, written in the form (2), may be regarded as the 2-D gradient of a 2-D scalar function. These p linearly independent 2-D gradient functions span a subspace $G_p^{(e)}$ of $V_p^{(e)}$. Similarly, the 1-D space W_p gives rise to a 2-D space $W_p^{(e)}$. So we have

$$V_p^{(e)} = G_p^{(e)} \oplus W_p^{(e)}. \tag{4}$$

Let $V_p^{(f)}$ be the subspace of V_p consisting of functions with vanishing tangential component on all three edges of the triangle. Since these functions are not associated with edges, but with the face of the element, they are called *face* functions. It is easy to see that $V_p^{(f)}$ and the three $V_p^{(e)}$ spaces have no function in common (except zero), and that any function in V_p can be expressed as the sum of functions from the four spaces, i.e.,

$$V_p = V_p^{(f)} \oplus 3V_p^{(e)} \tag{5}$$

where $3V_p^{(e)}$ is just the sum of the three edge spaces (one per edge). From this it follows that the dimension of $V_p^{(f)}$ is $(p+1)(p+2) - 3(p+1) = (p-1)(p+1)$. Functions in $V_p^{(f)}$ can be written in a variety of ways, including

$$\zeta_1\zeta_2 P_3(\zeta_1,\zeta_2,\zeta_3)\nabla\zeta_3 + \zeta_2\zeta_3 P_1(\zeta_1,\zeta_2,\zeta_3)\nabla\zeta_1 + \zeta_3\zeta_1 P_2(\zeta_1,\zeta_2,\zeta_3)\nabla\zeta_2 \tag{6}$$

where P_1, P_2, and P_3 are polynomials of degree $p-2$. Just like $V_p^{(e)}$, the face space $V_p^{(f)}$ can be partitioned. Let $G_p^{(f)}$ be the subspace of $V_p^{(f)}$ consisting of functions that are gradients of scalars. Since there are $p(p-1)/2$ linearly independent scalar polynomials of degree $p+1$, vanishing on the edges $\zeta_1 = 0$, $\zeta_2 = 0$, and $\zeta_3 = 0$, the dimension of $G_p^{(f)}$ is $p(p-1)/2$. Functions in $G_p^{(f)}$ can be written in the form

$$\nabla(\zeta_1\zeta_2\zeta_3 P(\zeta_1,\zeta_2,\zeta_3)) \tag{7}$$

TABLE I
SPACES AND THEIR DIMENSIONS FOR THE TRIANGULAR ELEMENT (NOTE THAT R_p INCLUDES TWO CONSTANT FUNCTIONS (GRADIENTS) FROM $W_p^{(e)}$)

	Gradient		Rotational/Whitney		Complete to order p	
Edge	$G_p^{(e)}$	$3 \times p$	$W_p^{(e)}$	3×1	$V_p^{(e)}$	$3 \times (p+1)$
Face	$G_p^{(f)}$	$p(p-1)/2$	$R_p^{(f)}$	$(p-1)(p+2)/2$	$V_p^{(f)}$	$(p-1)(p+1)$
Total	G_p	$(p+2)(p+3)/2 - 3$	R_p	$p(p+1)/2 + 2$	V_p	$(p+1)(p+2)$

where P is a polynomial of degree $p-2$. Let $R_p^{(f)}$ be any other subspace of $V_p^{(f)}$ that has no functions in common with $G_p^{(f)}$ (except zero) and has dimension equal to

$$\dim V_p^{(f)} - \dim G_p^{(f)} = (p-1)(p+2)/2.$$

Then

$$V_p^{(f)} = G_p^{(f)} \oplus R_p^{(f)}. \tag{8}$$

Since functions in $R_p^{(f)}$ cannot also belong to $G_p^{(f)}$, they cannot be expressed as a gradient, and must therefore have a nonzero curl. $R_p^{(f)}$ is the space of *rotational* face functions, and (8) represents a (nonunique) split of the face functions into gradient and rotational subspaces.

Note that the split (4) is not quite the same. The space $W_p^{(e)}$ contains the function (3), which is the Whitney edge function. The three Whitney edge functions for the element span a space that includes a rotational function, but also includes the two constant functions, which are gradients. A full separation of $V_p^{(e)}$ into gradient and rotational parts can be achieved using the tree and cotree of the graph of edges of the mesh [25], but that approach will not be pursued here.

The various spaces and their dimensions for the triangular element are summarized in Table I.

C. Three Dimensions: Tetrahedral Element

The tetrahedral element is shown in Fig. 1. The normalized coordinates for this simplex are ζ_1, ζ_2, ζ_3, and ζ_4, satisfying the relationship $\zeta_1 + \zeta_2 + \zeta_3 + \zeta_4 = 1$. At node 1, $\zeta_1 = 1$; on the opposite face (connecting nodes 2, 3, and 4), $\zeta_1 = 0$. The linear relationship between simplex coordinates and Cartesian coordinates (x, y, z) is given by (for example) Silvester and Ferrari [24].

Let V_p be the space of all 3-D vector functions on this element, which are polynomials in ζ_1, ζ_2, ζ_3, and ζ_4 of degree less than or equal to p. Since there are $(p+1)(p+2)(p+3)/6$ linearly independent scalar polynomials of degree p in three dimensions, the dimension of V_p is $(p+1)(p+2)(p+3)/2$.

Each basis function of the 1-D V_p space, written in the form (1), may regarded as a 3-D vector function in the 3-D V_p space, with these properties:

a) its tangential component vanishes on the face $\zeta_1 = 0$ and on the face $\zeta_2 = 0$;
b) its tangential component on the edge $\zeta_3 = \zeta_4 = 0$ (connecting nodes 1 and 2) is identical to the 1-D basis function.

As a consequence of b), the 3-D functions generated in this way must be linearly independent, and constitute a basis for a subspace of dimension $p+1$. This is a space $V_p^{(e)}$ of edge

functions associated with the edge $\zeta_3 = \zeta_4 = 0$. By replacing the indices $(1,2)$ in (1) by $(1,3)$, $(1,4)$, $(2,3)$, $(2,4)$, $(3,4)$, in turn, five similar spaces can be obtained, one for each edge.

As before, $V_p^{(e)}$ can be subdivided. Consider a basis for the 1-D G_p space. Each basis function, written in the form (2) may be regarded as the 3-D gradient of a 3-D scalar function. These p linearly independent 3-D gradient functions span a subspace $G_p^{(e)}$ of $V_p^{(e)}$. Similarly, the 1-D space W_p gives rise to a 3-D space $W_p^{(e)}$. So we have (4) again.

Each basis function of the 2-D $V_p^{(f)}$ space, written in the form (6), may regarded as a 3-D vector function in the 3-D V_p space, with the following properties:

a) its tangential component vanishes on the faces $\zeta_1 = 0$, $\zeta_2 = 0$, and $\zeta_3 = 0$ (and, therefore, on all six edges);
b) its tangential component on the face $\zeta_4 = 0$ (connecting nodes 1, 2, and 3) is identical to the 2-D basis function.

As a consequence of b), the 3-D functions generated in this way must be linearly independent, and constitute a basis for a subspace of dimension equal to that of the 2-D $V_p^{(f)}$, i.e., $(p-1)(p+1)$. This is a space $V_p^{(f)}$ of face functions associated with the face $\zeta_4 = 0$ and vanishing tangentially on the other three faces. By replacing the indices $(1,2,3)$ in (6) by $(1,2,4)$, $(1,4,3)$, $(2,3,4)$, in turn, three similar spaces can be obtained, one for each face.

$V_p^{(f)}$ can be divided into two parts. Consider a basis for the 2-D $G_p^{(f)}$ space. Each basis function, written in the form (7), may be regarded as the 3-D gradient of a 3-D scalar function. These $p(p-1)/2$ linearly independent 3-D gradient functions span a subspace $G_p^{(f)}$ of $V_p^{(f)}$. Similarly, the 2-D space $R_p^{(f)}$ gives rise to a 3-D subspace of $V_p^{(f)}$. So we have (8) again.

Functions $\boldsymbol{r}$ in the 3-D space $R_p^{(f)}$ satisfy the following relation:

$$\hat{\boldsymbol{n}} \cdot (\nabla \times \boldsymbol{r})|_{\zeta_4=0} = \hat{\boldsymbol{n}} \cdot \nabla_{2\text{-D}} \times \boldsymbol{r}_{2\text{-D}} \tag{9}$$

where $\boldsymbol{r}_{2\text{-D}}$ is the equivalent function in the 2-D space, and $\hat{\boldsymbol{n}}$ is a unit vector perpendicular to the face. Since the 2-D space is rotational, the right-hand side cannot vanish; therefore, $\boldsymbol{r}$ must have a nonzero curl, i.e., the 3-D space $R_p^{(f)}$ is also rotational.

Let $V_p^{(v)}$ be the subspace of V_p consisting of functions with vanishing tangential component on all four faces of the tetrahedron. These are the *volume* functions. It is easy to see that $V_p^{(v)}$, the four $V_p^{(f)}$ spaces, and the six $V_p^{(e)}$ spaces have no overlap (except zero), and that any function in V_p can be expressed as the sum of functions from the 11 spaces, i.e.,

$$V_p = V_p^{(v)} \oplus 4V_p^{(f)} \oplus 6V_p^{(e)}. \tag{10}$$

From this it follows that the dimension of $V_p^{(v)}$ is

$$(p+1)(p+2)(p+3)/2 - (p-1)(p+1) - 6(p+1) = (p-2)(p-1)(p+1)/2.$$

Just like $V_p^{(f)}$ and $V_p^{(e)}$, the interior space $V_p^{(v)}$ can be subdivided. Let $G_p^{(v)}$ be the subspace of $V_p^{(v)}$ consisting of functions that are gradients of scalars. Since there are $p(p-1)(p-2)/6$ linearly independent scalar polynomials of degree $p+1$, vanishing on the faces $\zeta_1 = 0$, $\zeta_2 = 0$, $\zeta_3 = 0$, and $\zeta_4 = 0$, the dimension of $G_p^{(v)}$ is $p(p-1)(p-2)/6$. Let $R_p^{(v)}$ be any other subspace of $V_p^{(v)}$ that has no function in common with $G_p^{(v)}$ (except zero) and has dimension equal to

$$\dim V_p^{(v)} - \dim G_p^{(v)} = (p-2)(p-1)(2p+3)/6.$$

Then

$$V_p^{(v)} = G_p^{(v)} \oplus R_p^{(v)}. \tag{11}$$

Since functions in $R_p^{(v)}$ cannot also belong to $G_p^{(v)}$, they cannot be expressed as a gradient, and must, therefore, have a nonzero curl. $R_p^{(v)}$ is a rotational space.

The various spaces and their dimensions for the tetrahedral element are summarized in Table II. Note that the dimensions presented here for the edge, face, and volume functions complete to order p coincide with those of Nedelec [15].

TABLE II
SPACES AND THEIR DIMENSIONS FOR THE TETRAHEDRAL ELEMENT (NOTE THAT R_p INCLUDES THREE CONSTANT FUNCTIONS (GRADIENTS) FROM $W_p^{(e)}$)

	Gradient		Rotational/Whitney		Complete to order p	
Edge	$G_p^{(e)}$	6 x p	$W_p^{(e)}$	6 x 1	$V_p^{(e)}$	6 x (p+1)
Face	$G_p^{(f)}$	4 x p(p–1)/2	$R_p^{(f)}$	4 x (p–1)(p+2) /2	$V_p^{(f)}$	4 x (p–1)(p+1)
Volume	$G_p^{(v)}$	p(p–1)(p–2)/6	$R_p^{(v)}$	(p–2)(p–1) ($2p$+3) /6	$V_p^{(v)}$	(p–2)(p–1)(p+1)/2
Total	G_p	(p+2)(p+3)(p+4)/6 – 4	R_p	p($2p$+7)(p+1)/6 + 3	V_p	(p+1)(p+2)(p+3)/2

III. BASIS FUNCTIONS

A. Edge Basis Functions

The 1-D gradient space G_p needs p basis functions, i.e., one additional function is needed when the order increases from $p-1$ to p. This must be the gradient of a scalar which vanishes at $\zeta_1 = 0$ and $\zeta_2 = 0$. A suitable function, of form (2), is

$$\boldsymbol{G}_p^{(e)} = \nabla(\zeta_1\zeta_2(\zeta_1 - \zeta_2)^{p-1}), \qquad p \geq 1. \tag{12}$$

Note that this function is either symmetric in ζ_1, ζ_2 (for p odd), or antisymmetric (for p even). This is important. A symmetric basis function will take the same form in every element that shares the edge, so continuity is imposed simply by choosing the same coefficient for the function in each element. An antisymmetric basis function, on the other hand, comes in two varieties which differ by a sign, depending on which end of the edge is taken as local node 1 and which is local node 2 within an element. Because of this, choosing the same coefficient for the function for all the elements that share the edge will not, on its own, impose continuity. However, the fix is relatively simple: arbitrarily designate the 1 and 2 ends of the edge, and for elements whose local numbering does not match this choice, introduce a minus sign, i.e., multiply the row and column of the local matrix corresponding to this basis function by minus one.

Without symmetry or antisymmetry of the basis functions, enforcing tangential continuity would be more difficult. Provision would have to be made for assembling two varieties of each asymmetric edge function, with the appropriate variety chosen at runtime in accordance with the relative orientation of the edge.

The single basis function for W_p is $\boldsymbol{W}_1^{(e)}$, given in (3). It is antisymmetric.

B. Face Basis Functions

A basis for the 2-D gradient space $G_p^{(f)}$ may conveniently be found by first defining a basis for the scalar polynomials that vanish on the three edges of the face, and then taking their gradients.

A 2-D scalar space of order p, vanishing on $\zeta_1 = 0$, $\zeta_2 = 0$, and $\zeta_3 = 0$, has dimension $(p-1)(p-2)/2$, so when the order increases from $p-1$ to p, the number of extra scalar basis functions required is $p-2$. In choosing these $p-2$ functions, consideration must be given to the imposition of tangential continuity between two tetrahedra sharing a face. To ensure that a basis function $f(\zeta_1, \zeta_2, \zeta_3)$ of one tetrahedron has a matching basis function in the neighboring tetrahedron, two requirements are introduced:

a) Each basis function must be symmetric (or antisymmetric) in ζ_1, ζ_2, or ζ_2, ζ_3, or ζ_3, ζ_1.
b) The basis functions must be defined in *triplets*: f, ρf, $\rho^2 f$, where ρ is an operator that "rotates" a function in the direction $1 \to 2 \to 3 \to 1$

$$\begin{aligned} \rho f(\zeta_1, \zeta_2, \zeta_3) &= f(\zeta_2, \zeta_3, \zeta_1) \\ \rho^2 f(\zeta_1, \zeta_2, \zeta_3) &= f(\zeta_3, \zeta_1, \zeta_2). \end{aligned} \tag{13}$$

Then, for example, depending on the relative orientation of the two tetrahedra, function f of one tetrahedron might match function ρf (or $-\rho f$) of the neighboring tetrahedron.

Without loss of generality, we will assume henceforth that the first function f of the triplet is symmetric (antisymmetric) in ζ_1, ζ_2; it follows that ρf and $\rho^2 f$ are symmetric (antisymmetric) in ζ_2, ζ_3 and ζ_3, ζ_1, respectively.

The following $p-2$ polynomials of order p satisfy a). They are divided into τ triplets, where $\tau = \text{int}\,(p/3)$. All except the last triplet satisfies b). The first function of the ith triplet is

$$F_{pi} = (\zeta_1\zeta_2\zeta_3)^i \zeta_1\zeta_2(\zeta_1 - \zeta_2)^{p-3i-2}, \qquad i = 1, \cdots, \tau - 1 \tag{14}$$

and

$$\begin{aligned} F_{p\tau} &= (\zeta_1\zeta_2\zeta_3)^\tau, && \text{if } p \bmod 3 = 0 \\ &= (\zeta_1\zeta_2\zeta_3)^\tau(\zeta_1 - \zeta_2), && \text{if } p \bmod 3 = 1 \\ &= (\zeta_1\zeta_2\zeta_3)^\tau \zeta_1\zeta_2, && \text{if } p \bmod 3 = 2. \end{aligned} \tag{15}$$

When $p \bmod 3 = 0$, all three functions of the τth triplet are the same, i.e., the triplet degenerates to a single function. Since the single function is itself unaffected by the rotation operator ρ, it always has a matching function in the neighboring tetrahedron.

When $p \bmod 3 = 1$, there are apparently three distinct functions in the τth triplet, but in fact $\rho^2 F_{p\tau} + \rho F_{p\tau} + F_{p\tau} = 0$, so there are really only two. The simplest way to handle this is to provide code for the assembly of all three functions, but assemble just two of them, making sure that the one omitted in one tetrahedron matches the one omitted in the neighboring tetrahedron.

When $p \bmod 3 = 2$, there are three distinct functions in the τth triplet, just as there are for all triplets $i < \tau$. Overall, the functions defined by (14) and (15) lead to $p-2$ linearly independent scalar functions, as intended.

The gradient space $G_p^{(f)}$ needs $p(p-1)/2$ basis functions, so in going from order $p-1$ to order p an additional $p-1$ basis functions of order p are needed. These can be generated from the scalar triplets of order $p+1$. The first functions of the $\text{int}\,((p+1)/3)$ triplets of $G_p^{(f)}$ are

$$\boldsymbol{G}_{pi}^{(f)} = \nabla F_{p+1,i+1}, \qquad i = 0, \cdots, \text{int}\left(\frac{p+1}{3}\right) - 1. \tag{16}$$

The rotational space $R_p^{(f)}$ is handled in a similar way. In going from order $p-1$ to order p an additional p basis functions of order p are needed, each of the form (6). The symmetry requirements a) and b), above, apply to these vector basis functions too. A suitable basis has $\sigma = \text{int}\,\big((p-1)/3\big)$ triplets. The first function of the ith triplet is

$$\boldsymbol{R}_{pi}^{(f)} = F_{pi}\nabla\zeta_3, \qquad i = 0, \cdots, \sigma - 1 \tag{17}$$

(equation (14) is extended to include the case $i = 0$), and

$$\begin{aligned} \boldsymbol{R}_{p\sigma}^{(f)} &= F_{p\sigma}\nabla\zeta_3 \quad \text{if } p \bmod 3 = 0 \\ &= (\zeta_1\zeta_2\zeta_3)^\sigma[(\zeta_1 - \zeta_2)\nabla\zeta_3 \\ &\quad + (\zeta_2 - \zeta_3)\nabla\zeta_1 + (\zeta_3 - \zeta_1)\nabla\zeta_2], \qquad \text{if } p \bmod 3 = 1 \\ &= (\zeta_1\zeta_2\zeta_3)^\sigma[-2\zeta_1\zeta_2\nabla\zeta_3 + \zeta_2\zeta_3\nabla\zeta_1 + \zeta_3\zeta_1\nabla\zeta_2], \\ &\qquad\qquad \text{if } p \bmod 3 = 2. \end{aligned} \tag{18}$$

When $p \bmod 3 = 1$, the triplet σ degenerates to a single function. When $p \bmod 3 = 2$, there are apparently three distinct functions in the triplet σ, but in fact $\rho^2\boldsymbol{R}_{p\sigma} + \rho\boldsymbol{R}_{p\sigma} + \boldsymbol{R}_{p\sigma} = 0$, so there are really only two; one of the three must be omitted when building the local stiffness matrix, as before. With the special cases taken into account, the overall number of basis functions obtained from (17) and (18) is p, as required.

As pointed out earlier, the rotational space $R_p^{(f)}$ is not unique. By selecting the above basis, one possible $R_p^{(f)}$ has been defined. In fact, it is possible to choose a basis such that $R_p^{(f)}$ is spatially biased: that is, a permutation of the numbering of the three nodes of the face would result in a different space. This kind of anisotropy is exhibited by the earlier $R_2^{(f)}$ functions of the author [8] and of Lee [6], and is intuitively undesirable. The question then arises, is the new basis given above isotropic? For triplets that are nondegenerate, the answer is clearly yes. The single function that arises in the $p \bmod 3 = 1$ case is also isotropic, by inspection of (18). In the $p \bmod 3 = 2$ case, it might be thought that omitting one of the three functions would induce anisotropy, but this is not true. The function omitted is linearly dependent on the other two, and omitting it does not change the space spanned by the original triplet, which is isotropic. So in all cases, $R_p^{(f)}$ is isotropic.

C. Volume Basis Functions

A 3-D scalar space of order p, vanishing on $\zeta_1 = 0$, $\zeta_2 = 0$, $\zeta_3 = 0$, and $\zeta_4 = 0$, has dimension $(p-1)(p-2)(p-3)/6$,

so when the order increases from $p-1$ to p, the number of extra basis functions required is $(p-2)(p-3)/2$. Fortunately, these functions can be chosen without regard to the need for imposing continuity between elements. A suitable basis is

$$i = 0, \cdots, p-4: \quad V_{pij} = (\zeta_1\zeta_2\zeta_3\zeta_4)\zeta_1^{p-4-i}\zeta_2^{i-j}\zeta_3^j, \qquad j = 0, \cdots, i. \tag{19}$$

A basis for the gradient space $G_p^{(v)}$ is obtained by taking the gradient of the scalar basis of degree $p+1$. The additional basis functions needed when the order increases from $p-1$ to p are

$$i = 0, \cdots, p-3: \quad \boldsymbol{G}_{pij}^{(v)} = \nabla V_{p+1,i,j}, \qquad j = 0, \cdots, i. \tag{20}$$

The rotational space $R_p^{(v)}$ has dimension $(p-2)(p-1)(2p+3)/6$, so when the order increases from $p-1$ to p, the number of extra basis functions required is $p(p-2)$. A suitable basis is

$$\begin{aligned} i = 0, \cdots, p-3: \quad & \boldsymbol{R}_{pij}^{(v)} = (\zeta_1\zeta_2\zeta_3\nabla\zeta_4)\zeta_1^{p-3-i}\zeta_2^{i-j}\zeta_3^j, && j = 0, \cdots, i \\ & \boldsymbol{R}_{p,i,i+1+j}^{(v)} = (\zeta_2\zeta_3\zeta_4\nabla\zeta_1)\zeta_2^{p-3-i}\zeta_3^{i-j}\zeta_4^j, && j = 0, \cdots, i \\ & \boldsymbol{R}_{p,i,2i+2}^{(v)} = (\zeta_3\zeta_4\zeta_1\nabla\zeta_2)\zeta_3^{p-3-i}\zeta_4^i. \end{aligned} \tag{21}$$

IV. Orthogonality

The basis functions given above could be used directly to solve problems in vector electromagnetics. However, as the polynomial order increases, the issue of matrix ill-conditioning becomes important. The basis functions become increasingly similar to one another, and consequently the condition number of the stiffness matrix built from them deteriorates, affecting accuracy and, for iterative solution methods, overall computation time. Notice that this is not a particular problem for interpolatory bases, where a high degree of linear independence seems to follow from the interpolation property. However, with hierarchal bases it is an issue that must be addressed.

A solution is to find a new set of basis functions which are at least partially orthogonal to one another in an appropriate inner product. If complete orthogonality could be achieved in the natural inner product of the finite element method, the condition number of the stiffness matrix would be one. This is not possible. However, a partial approach does lead to substantial improvements.

The inner product chosen is based on the usual function-space inner product

$$(\boldsymbol{f}, \boldsymbol{g}) = \int \boldsymbol{f} \cdot \boldsymbol{g}\, dV. \tag{22}$$

Since we are looking for a new basis that is independent of the particular problem being solved, the domain of integration in (22) must be a single tetrahedron. In the scalar case [26], the integral turns out to be independent of the shape of the tetrahedron, scaling simply with its volume. Unfortunately, this is not the case with vector basis functions, so a particular tetrahedron must be chosen as the basis for the orthogonality. The equilateral tetrahedron of unit side is the natural choice. For this element, it can be shown that

$$(\boldsymbol{f}, \boldsymbol{g}) = \frac{3\sqrt{2}}{4} \sum_{p=1}^{4}\sum_{q=1}^{4} e_{pq} \times \int_{\zeta_3=0}^{1}\int_{\zeta_2=0}^{1-\zeta_3}\int_{\zeta_1=0}^{1-\zeta_2-\zeta_3} f_p g_q \, d\zeta_1 \, d\zeta_2 \, d\zeta_3 \tag{23}$$

where $e_{pq} = 1$ if $p = q$ and $-\frac{1}{3}$ otherwise; and f_p, g_q are defined by the following expansions of the vectors $\boldsymbol{f}, \boldsymbol{g}$, respectively:

$$\boldsymbol{f} = \sum_{p=1}^{4} f_p \nabla\zeta_p \qquad \boldsymbol{g} = \sum_{q=1}^{4} g_q \nabla\zeta_q. \tag{24}$$

Using (23) and a symbolic mathematics program, the inner product of any two basis functions can be found, and new basis functions derived.

Consider first the volume gradient functions. Straightforward Gram–Schmidt orthogonalization can be used to make these all mutually orthogonal. The lowest order function $\boldsymbol{G}_{300}^{(v)}$ is left alone; $\boldsymbol{G}_{400}^{(v)}$ is replaced by a linear combination of $\boldsymbol{G}_{400}^{(v)}$ and $\boldsymbol{G}_{300}^{(v)}$ that is orthogonal to $\boldsymbol{G}_{300}^{(v)}$; then $\boldsymbol{G}_{410}^{(v)}$ is replaced by a linear combination of $\boldsymbol{G}_{410}^{(v)}$, $\boldsymbol{G}_{400}^{(v)}$, and $\boldsymbol{G}_{300}^{(v)}$ that is orthogonal to $\boldsymbol{G}_{400}^{(v)}$ and $\boldsymbol{G}_{300}^{(v)}$; and so on.

The volume rotational functions can be orthogonalized in the same way, but in addition they can be orthogonalized with respect to the gradient functions of the same or lower order, e.g., $\boldsymbol{R}_{300}^{(v)}$ is orthogonalized with respect to $\boldsymbol{G}_{300}^{(v)}$ by replacing it by a linear combination of $\boldsymbol{R}_{300}^{(v)}$ and $\boldsymbol{G}_{300}^{(v)}$. The converse is not possible because it would destroy the irrotationality of the gradient functions.

Edge functions are orthogonalized only with respect to other edge functions on the same edge. Since symmetric and antisymmetric functions are automatically orthogonal, it is only necessary to orthogonalize a symmetric function against lower order symmetric functions, and similarly for antisymmetric functions.

Face functions are a little more involved, because of the need to maintain the triplet structure. The orthogonalization now must be *between triplets*, in the following sense. A triplet based on a function $\boldsymbol{f}$ (i.e., consisting of the functions $\boldsymbol{f}$, $\rho\boldsymbol{f}$, $\rho^2\boldsymbol{f}$) is said to be orthogonal to a triplet based on a function $\boldsymbol{g}$ when each function of the $\boldsymbol{f}$ triplet is orthogonal to all three functions of the $\boldsymbol{g}$ triplet. This orthogonalization can be achieved by replacing $\boldsymbol{f}$ by a linear combination of $\boldsymbol{f}$, $\boldsymbol{g}$, $\rho\boldsymbol{g}$, and $\rho^2\boldsymbol{g}$ that is orthogonal to $\boldsymbol{g}$, $\rho\boldsymbol{g}$, and $\rho^2\boldsymbol{g}$.

Once again, however, symmetry plays a part in simplifying the orthogonalization. Each triplet can be said to be symmetric or antisymmetric, depending on whether the first function $\boldsymbol{f}$ is symmetric or antisymmetric in ζ_1, ζ_2. A symmetric triplet has an alternative basis of three functions

$$\begin{aligned} & \boldsymbol{f} && (\text{symmetric in } \zeta_1, \zeta_2) \\ & (\rho+\rho^2)\boldsymbol{f} && (\text{symmetric in } \zeta_1, \zeta_2) \\ & (\rho-\rho^2)\boldsymbol{f} && (\text{antisymmetric in } \zeta_1, \zeta_2). \end{aligned}$$

TABLE III
FUNCTIONS TO USE TO ORTHOGONALIZE $\boldsymbol{f}$ WITH RESPECT TO TRIPLET $\boldsymbol{g}$

	g symmetric	g antisymmetric	g symmetric, single	g antisymmetric, single
f symmetric	g, $(\rho+\rho^2)$g	$(\rho-\rho^2)$g	g	*none*
f antisymmetric	$(\rho-\rho^2)$g	g, $(\rho+\rho^2)$g	*none*	g
f sym., single	g	*none*	g	*none*
f antisym., single	*none*	g	*none*	g

TABLE IV
EDGE FUNCTIONS TO ORDER 3
s = SYMMETRIC, a = ANTISYMMETRIC

		Before Orthogonalizing		After Orthogonalizing	
		Coeff. of $\nabla\zeta_1$	Coeff. of $\nabla\zeta_2$	Coeff. of $\nabla\zeta_1$	Coeff. of $\nabla\zeta_2$
$\mathbf{W}_1^{(e)}$	a	$-\zeta_2$	ζ_1	Unchanged	
$\mathbf{G}_1^{(e)}$	s	ζ_2	ζ_1	Unchanged	
$\mathbf{G}_2^{(e)}$	a	$\zeta_2(2\zeta_1-\zeta_2)$	$-\zeta_1(2\zeta_2-\zeta_1)$	Unchanged	
$\mathbf{G}_3^{(e)}$	s	$\zeta_2(\zeta_1-\zeta_2)$ $(3\zeta_1-\zeta_2)$	$\zeta_1(\zeta_2-\zeta_1)$ $(3\zeta_2-\zeta_1)$	$\zeta_2(45{\zeta_1}^2-60\zeta_1\zeta_2+15{\zeta_2}^2-2)$	$\zeta_1(45{\zeta_2}^2-60\zeta_2\zeta_1+15{\zeta_1}^2-2)$

TABLE V
FACE FUNCTIONS TO ORDER 3
JUST THE FIRST FUNCTION OF THE TRIPLET IS GIVEN;
s = SYMMETRIC, a = ANTISYMMETRIC, t = NUMBER OF INDEPENDENT FUNCTIONS IN THE TRIPLET

		t	Before Orthogonalizing			After Orthogonalizing		
			Coeff. $\nabla\zeta_1$	Coeff. $\nabla\zeta_2$	Coeff. $\nabla\zeta_3$	Coeff. $\nabla\zeta_1$	Coeff. $\nabla\zeta_2$	Coeff. $\nabla\zeta_3$
$\mathbf{G}_{20}^{(f)}$	s	1	$\zeta_2\zeta_3$	$\zeta_1\zeta_3$	$\zeta_1\zeta_2$		Unchanged	
$\mathbf{R}_{20}^{(f)}$	s	2	$\zeta_2\zeta_3$	$\zeta_1\zeta_3$	$-2\zeta_1\zeta_2$		Unchanged	
$\mathbf{G}_{30}^{(f)}$	a	2	$\zeta_2\zeta_3(2\zeta_1-\zeta_2)$	$-\zeta_1\zeta_3(2\zeta_2-\zeta_1)$	$\zeta_1\zeta_2(\zeta_1-\zeta_2)$		Unchanged	
$\mathbf{R}_{30}^{(f)}$	a	3	0	0	$\zeta_1\zeta_2(\zeta_1-\zeta_2)$	$-\zeta_2\zeta_3$ $(1398\zeta_1-699\zeta_2+86)$	$\zeta_1\zeta_3$ $(1398\zeta_2-699\zeta_1+86)$	$2020\zeta_1\zeta_2$ $(\zeta_1-\zeta_2)$

An antisymmetric triplet has this alternative basis

$$\begin{array}{ll} \boldsymbol{f} & (\text{antisymmetric in } \zeta_1,\zeta_2)\\ (\rho+\rho^2)\boldsymbol{f} & (\text{antisymmetric in } \zeta_1,\zeta_2)\\ (\rho-\rho^2)\boldsymbol{f} & (\text{symmetric in } \zeta_1,\zeta_2). \end{array}$$

To orthogonalize the symmetric triplet $\boldsymbol{f}$ with respect to the symmetric triplet $\boldsymbol{g}$, then, it is only necessary to orthogonalize $\boldsymbol{f}$ with respect to $\boldsymbol{g}$ and $(\rho+\rho^2)\boldsymbol{g}$; orthogonality between $\boldsymbol{f}$ and $(\rho-\rho^2)\boldsymbol{g}$ is automatic, since a face function symmetric in ζ_1,ζ_2 is necessarily orthogonal to a face function antisymmetric in ζ_1,ζ_2. Not only does this cut down on the work, but it ensures that the new, orthogonalized $\boldsymbol{f}$ is still symmetric in ζ_1,ζ_2, which is important for imposing continuity across faces (as explained above).

Similar simplification is possible for other combinations of symmetry and antisymmetry. The results are summarized in Table III. The "single" cases are degenerate triplets consisting of just one function, which is assumed to be unchanged by the ρ operator.

Tables IV–VI give the basis functions to order 3, both before and after orthogonalization. To explain how these tables can be used to construct an element providing gradient and rotational spaces of given orders, consider the example of a tetrahedron which is to have a gradient order of 2 and a rotational order of 3. Begin with Table IV. Every element requires $\boldsymbol{W}_1^{(e)}$, so that is the first set of basis functions, six in all, one per edge. Since the gradient order is 2, both $\boldsymbol{G}_1^{(e)}$ and $\boldsymbol{G}_2^{(e)}$ are needed. Each adds one function per edge, so there is a total of 18 edge functions for this element. Turning to Table V, $\boldsymbol{G}_{20}^{(f)}$ is needed. As explained in the table caption, $\boldsymbol{G}_{20}^{(f)}$ is just the first function in a triplet. However, in this case the triplet is degenerate, and

TABLE VI
VOLUME FUNCTIONS TO ORDER 3

	Before Orthogonalizing			After Orthogonalizing		
	Coeff. $\nabla\zeta_1$	Coeff. $\nabla\zeta_2$	Coeff. $\nabla\zeta_3$	Coeff. $\nabla\zeta_1$	Coeff. $\nabla\zeta_2$	Coeff. $\nabla\zeta_3$
$\mathbf{G}_{300}^{(v)}$	$\zeta_2\zeta_3(\zeta_4-\zeta_1)$	$\zeta_1\zeta_3(\zeta_4-\zeta_2)$	$\zeta_1\zeta_2(\zeta_4-\zeta_3)$		Unchanged	
$\mathbf{R}_{300}^{(v)}$	$\zeta_1\zeta_2\zeta_3$	$\zeta_1\zeta_2\zeta_3$	$\zeta_1\zeta_2\zeta_3$	$\zeta_2\zeta_3(\zeta_4+3\zeta_1)$	$\zeta_1\zeta_3(\zeta_4+3\zeta_2)$	$\zeta_1\zeta_2(\zeta_4+3\zeta_3)$
$\mathbf{R}_{301}^{(v)}$	$\zeta_2\zeta_3\zeta_4$	0	0	$2\zeta_2\zeta_3\zeta_4$	$-\zeta_3\zeta_1\zeta_4$	$-\zeta_1\zeta_2\zeta_4$
$\mathbf{R}_{302}^{(v)}$	0	$\zeta_3\zeta_1\zeta_4$	0	0	$\zeta_3\zeta_1\zeta_4$	$-\zeta_1\zeta_2\zeta_4$

contains only one member, so including $\boldsymbol{G}_{20}^{(f)}$ adds just one function per face, or a total of four for the tetrahedron. The next gradient function $\boldsymbol{G}_{30}^{(f)}$ is not needed, but both $\boldsymbol{R}_{20}^{(f)}$ and $\boldsymbol{R}_{30}^{(f)}$ are. The triplet $\boldsymbol{R}_{20}^{(f)}$ is again degenerate, with only two members

$$\zeta_2\zeta_3\nabla\zeta_1+\zeta_1\zeta_3\nabla\zeta_2-2\zeta_1\zeta_2\nabla\zeta_3$$

and

$$\zeta_3\zeta_1\nabla\zeta_2+\zeta_2\zeta_1\nabla\zeta_3-2\zeta_2\zeta_3\nabla\zeta_1 \tag{25}$$

(or any other two out of the triplet). $\boldsymbol{R}_{30}^{(f)}$, on the other hand, is not degenerate, and adds three functions per face, bringing the total number of face functions for the tetrahedron to 24. From Table VI just the rotational functions are needed, and there are three of these. The tetrahedron, then, has $18+24+3=45$ basis functions. (Cross check with Table II: $\dim G_2+\dim R_3 = 16+29=45$.)

V. MATRIX ASSEMBLY

Tetrahedral elements of the kind described above have been used to solve the equation governing the time-harmonic electric field $\boldsymbol{E}$ in a closed region Ω

$$\nabla\times\frac{1}{\mu_r}\nabla\times\boldsymbol{E}-k_o^2\varepsilon_r\boldsymbol{E}=0 \tag{26}$$

where μ_r and ε_r are the relative permeability and permittivity, respectively, given functions of position, and k_o is the free-space wavenumber. The boundary condition

$$\boldsymbol{E}_t=\boldsymbol{E}_o \quad \text{on } \partial\Omega \tag{27}$$

is applied, where $\boldsymbol{E}_t$ is the tangential part of $\boldsymbol{E}$, and $\boldsymbol{E}_o$ is a specified tangential field.

Application of the finite-element method approximates these equations by

$$\left([S]-k_o^2[T]\right)\{x\}=\{b\} \tag{28}$$

where $[S]$ and $[T]$ are square matrices, $\{x\}$ is a column vector of unknown coefficients, and $\{b\}$ is a known column vector arising from the nonzero boundary condition. The unknown coefficients are related to the electric field in each tetrahedral element as follows:

$$\boldsymbol{E}=\sum_i x_i\boldsymbol{N}_i \tag{29}$$

where the $\boldsymbol{N}_i$ are the vector basis functions of the element. The matrices $[S]$ and $[T]$ are built by the usual process of matrix assembly, starting from the following expressions for the entries of the local matrices for tetrahedron t:

$$\begin{aligned} S_{ij}&=\frac{1}{\mu_r}\int_t\nabla\times\boldsymbol{N}_i\cdot\nabla\times\boldsymbol{N}_j\,d\Omega\\ T_{ij}&=\varepsilon_r\int_t\boldsymbol{N}_i\cdot\boldsymbol{N}_j\,d\Omega. \end{aligned} \tag{30}$$

The assumption is made here that the material properties μ_r and ε_r are constant throughout the tetrahedron.

The integration in (30) could be carried out numerically, using for example a Gauss procedure. However, a universal-matrix approach is more efficient, particularly for higher order elements [27]. Let

$$\boldsymbol{N}_i = \sum_{m=1}^{3} N_{im}(\zeta_1, \zeta_2, \zeta_3)\nabla\zeta_m. \tag{31}$$

Any basis function can be written in this form by eliminating ζ_4 and $\nabla\zeta_4$ using $\zeta_1+\zeta_2+\zeta_3+\zeta_4 = 1$. Then by substituting (31) into (30) it can be shown that

$$[T] = 6V \sum_{m=1}^{3}\sum_{n=1}^{m} \nabla\zeta_m \cdot \nabla\zeta_n [T^{mn}] \tag{32}$$

where V is the volume of the tetrahedron; and the $[T^{mn}]$ are six symmetric universal matrices whose entries are given by

$$T_{ij}^{mn} = \theta_{mn} \int_{\zeta_3=0}^{1}\int_{\zeta_2=0}^{1-\zeta_3}\int_{\zeta_1=0}^{1-\zeta_2-\zeta_3} (N_{im}N_{jn} + N_{in}N_{jm})\, d\zeta_1\, d\zeta_2\, d\zeta_3. \tag{33}$$

θ_{mn} is 1 if m and n are unequal, $\frac{1}{2}$ if they are equal.

Taking the curl of (31) leads to

$$\nabla \times \boldsymbol{N}_i = \sum_{m=1}^{3} C_{im}(\zeta_1, \zeta_2, \zeta_3)\boldsymbol{e}_m \tag{34}$$

where

$$\begin{aligned} \boldsymbol{e}_1 &= \nabla\zeta_2 \times \nabla\zeta_3 \\ C_{i1} &= \frac{\partial N_{i3}}{\partial \zeta_2} - \frac{\partial N_{i2}}{\partial \zeta_3} \end{aligned} \tag{35}$$

and, similarly, by cyclic permutation of the indices, for $\boldsymbol{e}_2$, C_{i2}, $\boldsymbol{e}_3$, C_{i3}. Substituting (34) into (30) gives

$$[S] = 6V \sum_{m=1}^{3}\sum_{n=1}^{m} \boldsymbol{e}_m \cdot \boldsymbol{e}_n [S^{mn}] \tag{36}$$

where the $[S^{mn}]$ are another set of six symmetric universal matrices

$$S_{ij}^{mn} = \theta_{mn} \int_{\zeta_3=0}^{1}\int_{\zeta_2=0}^{1-\zeta_3}\int_{\zeta_1=0}^{1-\zeta_2-\zeta_3} (C_{im}C_{jn} + C_{in}C_{jm})\, d\zeta_1\, d\zeta_2\, d\zeta_3. \tag{37}$$

The $[T^{mn}]$ and $[S^{mn}]$ may be computed exactly with symbolic mathematics software (e.g., Maple) and stored in double precision with the finite-element program, to be used at run time to build the local matrices.

VI. Results

The following results are obtained with tetrahedral elements. The order of the element is designated by a pair of indices, (g, r), where g is the order of the gradient space (G_g) and r is the order of the rotational space (R_r).

The first example is a uniform length of parallel-plate waveguide (Fig. 2), short-circuited at one end ($z = 1m$) and driven with a unit electric field at the other ($z = 0$). The conducting plates are at $x = 0$ and $x = a$. The boundary surfaces $y = 0$ and $y = b$ are left free; the natural boundary condition for such surfaces is that the tangential magnetic field vanishes there, which is correct, since the magnetic field of the TEM mode of the waveguide has only a y component.

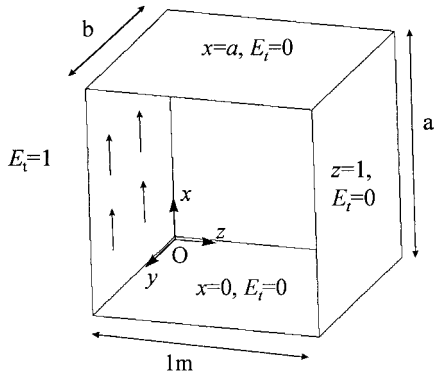


Fig. 2. A short-circuited length of parallel-plate waveguide, excited by a unit electric field. Surfaces with no specified E_t are left free, implying that the tangential magnetic field vanishes there.

The problem is meshed by subdividing it into $N_x \times N_y \times N_z$ equal-sized blocks, and then subdividing each block into six tetrahedra. Since in this problem there is no variation in the x and y directions, N_x and N_y are both set to 1. As N_z is increased, the dimensions a and b are scaled to keep the blocks cubical, in order to avoid any effects that might arise from severely distorted elements.

Once the problem is solved and the electric field calculated, the phase of the reflection coefficient at $z = 0$ is found as follows. The normalized susceptance B at $z = 0$ is obtained from the finite-element functional

$$B = -\frac{1}{k_o ab}\int_{\Omega}\left(\frac{1}{\mu_r}\nabla\times\boldsymbol{E}\cdot\nabla\times\boldsymbol{E} - k_o^2\varepsilon_r\boldsymbol{E}\cdot\boldsymbol{E}\right)d\Omega = 0 \tag{38}$$

and the phase of the reflection coefficient is then

$$\angle\Gamma = -2\tan^{-1} B. \tag{39}$$

The exact value for the phase is $-2k_o - \pi$ radians.

Fig. 3 shows the reduction in the error of the phase as the number of subdivisions N_z is increased, for "full-order" elements, $(g, r) = (1,1), (2,2), (3,3), (4,4)$. Note that the frequency is set higher for the higher orders, so that the asymptotic behavior can be seen on the same figure. The results reveal that, asymptotically, the phase error decreases as h^{2p} for the (p, p) element, where h is the element size. This is to be expected. A field that is complete to order p has a curl that is only complete to order $p-1$, which therefore has $O(h^p)$ errors. Since the functional depends on the square of the curl of the field, and is stationary at the true solution, the resulting error in the functional is $O(h^{2p})$.

Since the convergence rate is limited by the order of the curl, reducing the gradient space by one order should leave it unchanged. This is the rationale behind the "reduced-gradient" spaces: $(g, r) = (0,1), (1,2), (2,3), (3,4)$. Fig. 4 gives the results in this case. Indeed, the $(p-1, p)$ element is found to have the same asymptotic convergence rate as the (p, p) element.

A discussion of convergence rates for high-order edge elements (up to order 2) can also be found in [28].

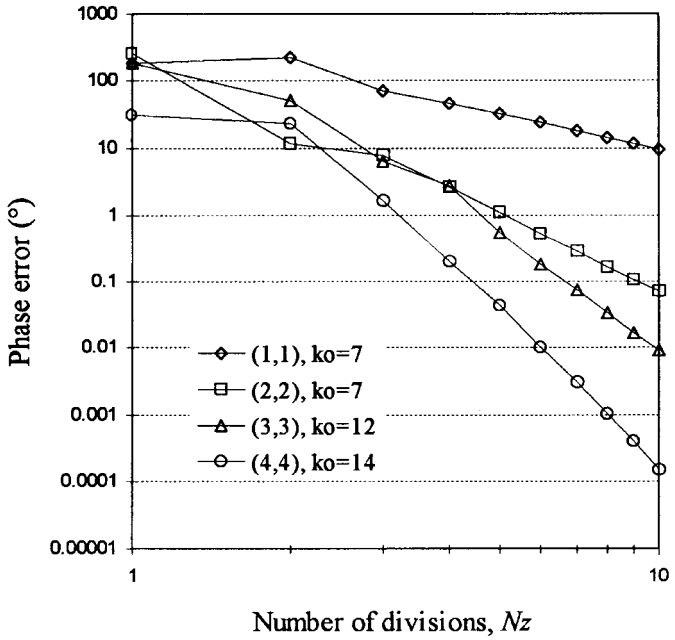


Fig. 3. Error in the phase of the reflection coefficient of a short-circuited piece of waveguide, versus number of subdivisions, for full-order elements, orders 1 to 4. k_o is in rads/m.

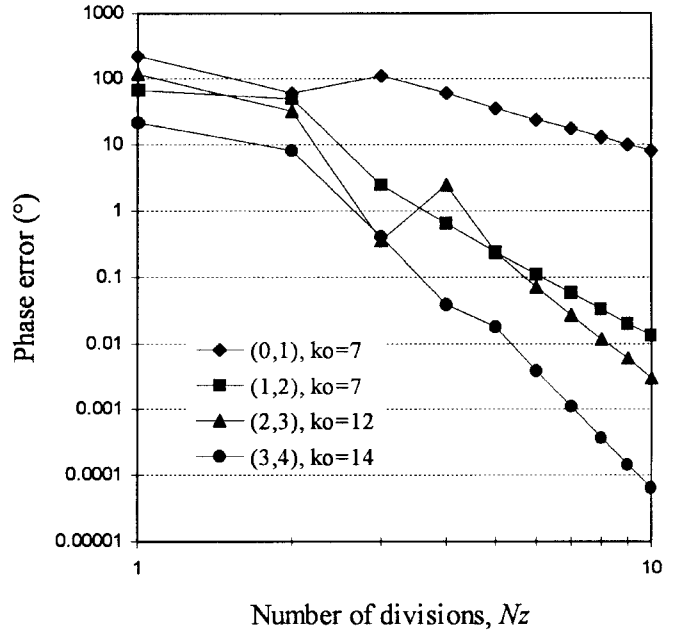


Fig. 4. Error in the phase of the reflection coefficient of a short-circuited piece of waveguide, versus number of subdivisions, for reduced-gradient elements, orders 1 to 4. k_o is in rads/m.

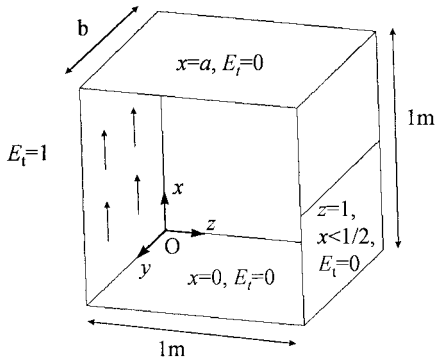


Fig. 5. Parallel-plate waveguide, excited by a unit electric field, and partially blocked by a conducting plate (at $z = 1$). Half the problem is modeled; $z = 1$ is a plane of symmetry. Surfaces with no specified E_t are left free, implying that the tangential magnetic field vanishes there.

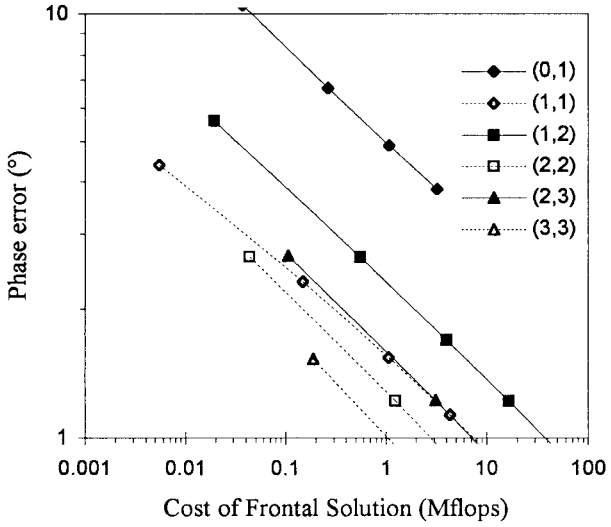


Fig. 6. Phase error versus cost of solution, as the number of tetrahedra is increased, for the capacitive obstacle shown in Fig. 5. $k_o = 1$ rads/m.

It might seem, then, that the $(p-1, p)$ element should always be used instead of the (p, p) element. That this is not the case is demonstrated by the next example. A parallel-plate waveguide is half blocked by a thin conducting plate, oriented perpendicular to the electric field and therefore creating a capacitive loading. When excited by equal electric fields on either side, the blocking plate lies in a plane of symmetry and one half of the problem can be solved, as shown in Fig. 5. The susceptance, B, at the input plane $z = 0$, and the corresponding reflection coefficient, are calculated in the same way as before.

Once again the problem is meshed by subdividing it into $N_x \times N_y \times N_z$ equal-sized blocks, and then subdividing each block into six tetrahedra. This time, however, N_x, N_y, and N_z are all set equal to the same value, N.

Fig. 6 shows the phase error for a number of different element types, as N is increased. To compare the computational efficiencies fairly, the errors are plotted against the number of floating-point operations required to solve the resulting matrix problem using a well-known sparse-matrix technique, the frontal method [29]. The exact phase for this problem is unknown, but a reference value obtained from extrapolation of the $(4, 4)$ results is believed to be accurate to within 0.1°, and is used as the true value for the purpose of calculating the phase errors.

Fig. 6 clearly demonstrates the superior computational efficiency of higher order elements, at least when a frontal solver is used. An element such as $(3, 3)$ can achieve the same phase accuracy as $(1, 1)$ in almost one tenth the CPU time. More interestingly, the full-order elements significantly outperform the reduced-gradient elements: the $(2, 3)$ element does only

about as well as $(1,1)$! (We are comparing here accuracy *per flop*; it is still true that $(2,3)$ gives significantly higher phase accuracy than $(1,1)$ with the same number of tetrahedra.) The reason for the superior performance of the full-order elements is that, because of the relatively low frequency and the orientation of the sharp conducting edge, the electric field in this problem is strongly gradient in nature. The curl of the electric field, i.e., the magnetic field, is relatively unimportant, and so the $(p-1,p)$ element, which favors the rotational space, is much less efficient than the (p,p) element. Of course, this will not always be the case, but it is important to realize that the best element type is problem (and region) dependent.

VII. Conclusion

Basis functions of arbitrary order for vector triangles and tetrahedra have been presented. Convergence tests demonstrate that the new elements have the expected characteristics, at least to fourth order, in three dimensions. Earlier hierarchal elements of this kind were limited to second order. The increased polynomial range will increase the power of p-adaptive methods in 3-D electromagnetics.

The question of what are the "correct" higher order extensions of the elementary Whitney element has long been debated. The natural extension appears to some to be the reduced-gradient series, because of the importance of the curl of the field in electromagnetics. Others have argued for function spaces that are complete to a given polynomial order. The results presented in this paper suggest that both sides may be right: where the curl dominates, the reduced-gradient element is more efficient; where the gradient is more important, the full-order element works better. This insight suggests that the balance of gradient to curl in the element should in fact be determined not *a priori*, but through an adaptive criterion.

References

[1] C. W. Crowley, P. P. Silvester, and H. Hurwitz, "Covariant projection elements for 3D vector field problems," *IEEE Trans. Magn.*, vol. 24, no. 1, pp. 397–400, Jan. 1988.

[2] P. T. S. Liu and J. P. Webb, "Hierarchal vector finite elements for 3D electromagnetics," *Proc. Inst. Elec. Eng., Microwaves, Antennas Propagat.*, vol. 142, no. 5, pp. 373–378, Oct. 1995.

[3] J. P. Webb, "Edge elements and what they can do for you," *IEEE Trans. Magn.*, vol. 29, pp. 1460–1465, Mar. 1993.

[4] A. Bossavit, "A rationale for 'edge-elements' in 3-D fields computations," *IEEE Trans. Magn.*, vol. 24, pp. 74–79, Jan. 1988.

[5] G. Mur, "A finite-element method for computing three-dimensional electromagnetic fields in inhomogeneous media," *IEEE Trans. Magn.*, vol. MAG-21, no. 6, p. 2188, 1985.

[6] J. F. Lee, "Analysis of passive microwave devices by using three-dimensional tangential vector finite elements," *Int. J. Numerical Modeling: Electronic Networks* (*Devices and Fields*), vol. 3, no. 4, pp. 235–246, Dec. 1990.

[7] J. F. Lee, D. K. Sun, and Z. J. Cendes, "Full-wave analysis of dielectric waveguides using tangential vector finite elements," *IEEE Trans. Microwave Theory Tech.*, vol. 39, pp. 1262–1271, Aug. 1991.

[8] J. P. Webb and B. Forghani, "Hierarchal scalar and vector tetrahedra," *IEEE Trans. Magn.*, vol. 29, pp. 1495–1498, Mar. 1993.

[9] A. Ahagon and T. Kashimoto, "Three-dimensional electromagnetic wave analysis using high order edge elements," *IEEE Trans. Magn.*, vol. 21, pp. 1753–1756, May 1995.

[10] A. Kameari, "Symmetric second order edge elements in triangles and tetrahedrons," presented at the Eighth Bienniel IEEE Conference on Electromagnetic Field Computation, Tucson, AZ, June 1–3, 1998.

[11] I. Babuska, B. A. Szabo, and I. N. Katz, "The p-version of the finite element method," *SIAM J. Numer. Anal.*, vol. 18, pp. 515–545, 1981.

[12] S. McFee and J. P. Webb, "Adaptive finite element analysis of microwave and optical devices using hierarchal triangles," *IEEE Trans. Magn.*, vol. 28, pp. 1708–1711, Mar. 1992.

[13] O. C. Zienkiewicz, J. Z. Zhu, and N. G. Gong, "Effective and practical h-p version adaptive analysis procedures for the finite element method," *Int. J. Numer. Methods in Eng.*, vol. 28, pp. 879–891, 1989.

[14] J. C. Nedelec, "Mixed finite elements in R3," *Numerische Mathematik*, vol. 35, pp. 315–341, 1980.

[15] _____, "A new family of mixed finite elements in R3," *Numerische Mathematik*, vol. 50, pp. 57–81, 1986.

[16] J. P. Webb and B. Forghani, "A T-Omega method using hierarchal edge elements," *Proc. Inst. Elec. Eng.,, Sci. Meas. Technol.*, vol. 142, no. 2, pp. 133–141, Mar. 1995.

[17] G. Peng, R. Dyczij-Edlinger, and J.-F. Lee, "Hierarchical methods for solving matrix equations from TVFEM's for microwave components," *IEEE Trans. Magn.*, vol. 35, pp. 1474–1477, May 1999.

[18] R. D. Graglia, D. R. Wilton, and A. F. Peterson, "Higher order interpolatory vector bases for computational electromagnetics," *IEEE Trans. Antennas Propagat.*, vol. 45, pp. 329–342, Mar. 1997.

[19] T. V. Yioultsis and T. D. Tsiboukis, "Development and implementation of second and third order vector finite elements in various 3-D electromagnetic field problems," *IEEE Trans. Magn.*, vol. 33, pp. 1812–1815, Mar. 1997.

[20] J. S. Wang and N. Ida, "Curvilinear and higher order 'edge' finite elements in electromagnetic field computation," *IEEE Trans. Magn.*, vol. 29, pp. 1491–1494, Mar. 1993.

[21] D. Sun, J. Manges, X. Yuan, and Z. Cendes, "Spurious modes in finite-element methods," *IEEE Antennas Propagat. Mag.*, vol. 37, pp. 12–24, Oct. 1995.

[22] P. Silvester, "High-order polynomial triangular elements for potential problems," *Int. J. Eng. Sci.*, vol. 7, pp. 849–861, 1969.

[23] L. S. Andersen and J. L. Volakis, "Hierarchical tangential vector finite elements for tetrahedra," *IEEE Microwave Guided Wave Lett.*, vol. 8, pp. 127–129, Mar. 1998.

[24] P. Silvester and R. L. Ferrari, *Finite Elements for Electrical Engineers*, 3rd ed. Cambridge, U.K.: Cambridge Univ. Press, 1996.

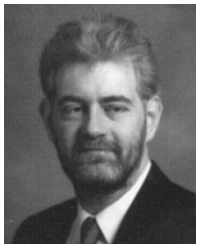
[25] R. Albanese and C. Rubinacci, "Solution of three dimensional eddy current problems by integral and differential methods," *IEEE Trans. Magn.*, vol. 24, pp. 98–101, Jan. 1988.

[26] J. P. Webb and R. Abouchacra, "Hierarchal triangular elements using orthogonal polynomials," *Int. J. Numer. Methods in Eng.*, vol. 38, no. 2, pp. 245–257, Jan. 1995.

[27] P. P. Silvester, "Universal finite element matrices for tetrahedra," *Int. J. Num. Met. Eng.*, vol. 18, pp. 1055–1061, 1982.

[28] C. Geuzaine, B. Meys, P. Dular, and W. Legros, "Convergence of high order curl-conforming finite elements," *IEEE Trans. Magn.*, vol. 35, pp. 1442–1445, May 1999.

[29] B. M. Irons, "A frontal solution program for finite element analysis," *Int. J. Num. Met. Eng.*, vol. 2, pp. 5–32, 1970.



Jon P. Webb (M'83) received the Ph.D. degree from Cambridge University, U.K., in 1981.

Since 1982, he has been Professor in the Electrical Engineering Department of McGill University, Montreal, Quebec, Canada. His area of research is computer methods in electromagnetics, especially the application of the finite-element method.