



## OPTIMAL FINITE ELEMENT METHODS OF ARBITRARY ORDER IN CURVED DOMAINS APPROXIMATED BY THE UNION OF ORDINARY N-SIMPLICES

**Vitoriano Ruas**

ruas.vitoriano@gmail.com

Chercheur collaborateur,  $\partial'$ Alembert, UPMC, Sorbonne Universités, Paris, France  
& Bolsista pesquisador do CNPq, PósMQI, PUC-Rio

Rua Marquês de São Vicente 225, CEP 22430-060, Gávea, Rio de Janeiro, RJ, Brasil.

**Abstract.** *One of the reasons for the great success of the finite element method is its versatility to deal with problems posed in curved domains. In this case method's isoparametric version for higher order methods, in connection with meshes consisting of curved triangles or tetrahedra, has been mostly employed to recover the optimal approximation properties known to hold for ordinary elements, in the case of polytopic domains. However, besides geometric inconveniences, the isoparametric technique helplessly requires the use of rational shape-functions and the use of numerical integration to compute element matrices. In this work a simple alternative is presented, to bypass these drawbacks in the case of Dirichlet boundary conditions, without eroding qualitative approximation properties. Examples with classical Lagrange finite elements are shown for both two- and three-dimensional problems. More generally this technique is as universal as can be, for it can be easily combined with different types of methods such as Hermite FEs, GFEs or XFEs, in order to maintain their order in the case of a curved domain, whenever it is greater than one for a polygonal or a polyhedral domain.*

**Keywords:** *Curved domain, Finite elements, Optimal order, Ordinary N-simplex, Polynomial shape-function*

## 1 INTRODUCTION

This work deals with a finite element method for solving boundary value problem posed in a two- or three-dimensional domain, with a smooth curved boundary of arbitrary shape. The principle it is based upon is close to the technique called *interpolated boundary conditions* studied in Brenner et al (2008) for two-dimensional problems. Although the latter technique is very intuitive and is known since the seventies (cf. Nitsche, 1972 and Scott, 1973), it has been of limited use so far. Among the reasons for this we could quote its difficult implementation, the lack of an extension to three-dimensional problems, and most of all, restrictions on the choice of boundary nodal points to reach optimal convergence rates. In contrast our method is simple to implement in both in two- and three-dimensional geometries. Moreover optimality is attained very naturally in both cases for various choices of boundary nodal points.

In order to allow an easier description of our methodology, thereby avoiding non essential technicalities, we consider as a model the Poisson equation in an  $N$ -dimensional smooth domain  $\Omega$  with boundary  $\Gamma$ , for  $N = 2$  or  $N = 3$ , with Dirichlet boundary conditions, namely,

$$\begin{cases} -\Delta u = f & \text{in } \Omega \\ u = d & \text{on } \Gamma, \end{cases} \quad (1)$$

where  $f$  and  $d$  are given functions defined in  $\Omega$  and on  $\Gamma$ , having suitable regularity properties. Here (1) is supposed to be solved by different  $N$ -simplex based finite element methods, incorporating degrees of freedom other than function values at the mesh vertices. For instance, if standard quadratic Lagrange finite elements are employed, it is well-known that approximations of an order not greater than 1.5 in the energy norm are generated (Ciarlet, 1978), in contrast to the second order ones that apply to the case of a polygonal or polyhedral domain, assuming that the solution is sufficiently smooth. If we are to recover the optimal second order approximation property something different has to be done. Since long the isoparametric version of the finite element method for meshes consisting of curved triangles or tetrahedra (Zienkiewicz, 1977), has been considered as the ideal way to achieve this. It turns out that, besides a more elaborated description of the mesh, the isoparametric technique inevitably leads to the integration of rational functions to compute the system matrix, which raises the delicate question on how to choose the right numerical quadrature formula in the master element. In contrast, in the technique to be introduced in this paper exact numerical integration can always be used for this purpose, since we only have to deal with polynomial shape-functions. Moreover the element geometry remains the same as in the case of polygonal or polyhedral domains. It is noteworthy that both advantages are conjugated with the fact that no erosion of qualitative approximation properties results from the application of our technique, as compared to the equivalent isoparametric one.

An outline of the paper is as follows. In Section 2 we present our method to solve the model problem with Dirichlet boundary conditions in a smooth curved two-dimensional domain with conforming Lagrange finite elements based on meshes with straight triangles, in connection with the standard Galerkin formulation. A numerical example illustrating technique's potential is given. In Section 3 we extend the approach adopted in Section 2 to the three-dimensional case including also numerical experimentation. We conclude in Section 4 with some comments on possible extensions of the methodology studied in this work.

In the remainder of this paper we will be given partitions  $\mathcal{T}_h$  of  $\Omega$  into (closed) ordinary triangles or tetrahedra, according to the value of  $N$ , satisfying the usual compatibility conditions (Ciarlet, 1978). Every  $\mathcal{T}_h$  is assumed to belong to a uniformly regular family of partitions. We denote by  $\Omega_h$  the set  $\cup_{T \in \mathcal{T}_h} T$  and by  $\Gamma_h$  the boundary of  $\Omega_h$ . Letting  $h_T$  be the diameter of  $T \in \mathcal{T}_h$ , we set  $h := \max_{T \in \mathcal{T}_h} h_T$ , as usual. We also recall that if  $\Omega$  is convex  $\Omega_h$  is a proper subset of  $\Omega$ .

## 2 THE TWO-DIMENSIONAL CASE

To begin with we describe our methodology in the case where  $N = 2$ . In order to simplify the presentation in this section we assume that  $d \equiv 0$ , leaving for the next one its extension to the case of an arbitrary  $d$ .

### 2.1 Method description

Here we make the very reasonable assumption on the mesh that no element in  $\mathcal{T}_h$  has more than one edge on  $\Gamma_h$ .

We also need some definitions regarding the skin  $(\Omega \setminus \Omega_h) \cup (\Omega_h \setminus \Omega)$ . First of all, in order to avoid non essential difficulties, we assume that the mesh is constructed in such a way that convex and concave portions of  $\Gamma$  correspond to convex and concave portions of  $\Gamma_h$ . This property is guaranteed if the points separating such portions of  $\Gamma$  are vertices of polygon  $\Omega_h$ . In doing so, let  $\mathcal{S}_h$  be the subset of  $\mathcal{T}_h$  consisting of triangles having one edge on  $\Gamma_h$ . Now for every  $T \in \mathcal{S}_h$  we denote by  $\Delta_T$  the set delimited by  $\Gamma$  and the edge  $e_T$  of  $T$  whose end-points belong to  $\Gamma$  and set  $T' := T \cup \Delta_T$  if  $\Delta_T$  is not a subset of  $T$  and  $T' := \overline{T} \setminus \Delta_T$  otherwise (see Figure 1). Notice that if  $e_T$  lies on a convex portion of  $\Gamma_h$ ,  $T$  is a proper subset of  $T'$ , while the

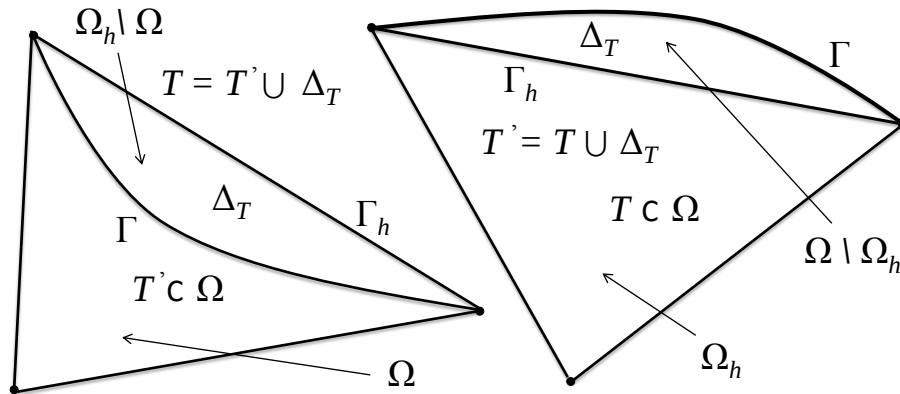


Figure 1: Skin  $\Delta_T$  related to a mesh triangle  $T$  next to a convex (right) or a concave (left) portion of  $\Gamma$

opposite occurs if  $e_T$  lies on a concave portion of  $\Gamma_h$ . With such a definition we can assert that there is a partition  $\mathcal{T}'_h$  of  $\Omega$  associated with  $\mathcal{T}_h$  consisting of non overlapping sets  $T'$  for  $T \in \mathcal{S}_h$ , besides the elements in  $\mathcal{T}_h \setminus \mathcal{S}_h$ .

For convenience henceforth we refer to the  $n_k$  points in a triangle  $T$  which are vertices of the  $k^2$  equal triangles  $T$  can be subdivided into, where  $n_k := (k+2)(k+1)/2$  for  $k > 1$  as the *lagrangian nodes of order k*, as explained in Ciarlet (1978) or Zienkiewicz (1977).

Next we introduce two function spaces  $V_h$  and  $W_h$  associated with  $\mathcal{T}_h$ .  $V_h$  is the standard Lagrange finite element space consisting of continuous functions  $v$  defined in  $\Omega_h$  that vanish on  $\Gamma_h$ , whose restriction to every  $T \in \mathcal{T}_h$  is a polynomial of degree less than or equal to  $k$  for  $k \geq 2$ . For convenience we extend by zero every function  $v \in V_h$  to  $\Omega \setminus \Omega_h$ .  $W_h$  in turn is the space of functions defined in  $\Omega_h \cup \Omega$  having the properties listed below.

1. The restriction of  $w \in W_h$  to every  $T \in \mathcal{T}_h$  is a polynomial of degree less than or equal to  $k$ ;
2. Every  $w \in W_h$  is continuous in  $\Omega_h$  and vanishes at the vertices of  $\Gamma_h$ ;
3. A function  $w \in W_h$  is also defined in  $\Omega \setminus \Omega_h$  in such a way that its polynomial expression in  $T \in \mathcal{S}_h$  also applies to points in  $\Delta_T$ ;
4.  $\forall T \in \mathcal{S}_h$ ,  $w(P) = 0$  for every  $P$  among the  $k - 1$  intersections with  $\Gamma$  of the line passing through the vertex  $O_T$  of  $T$  not belonging to  $\Gamma$  and the points  $M$  different from vertices of  $T$  subdividing the edge opposite to  $O_T$  into  $k$  segments of equal length (cf. Figure 2).

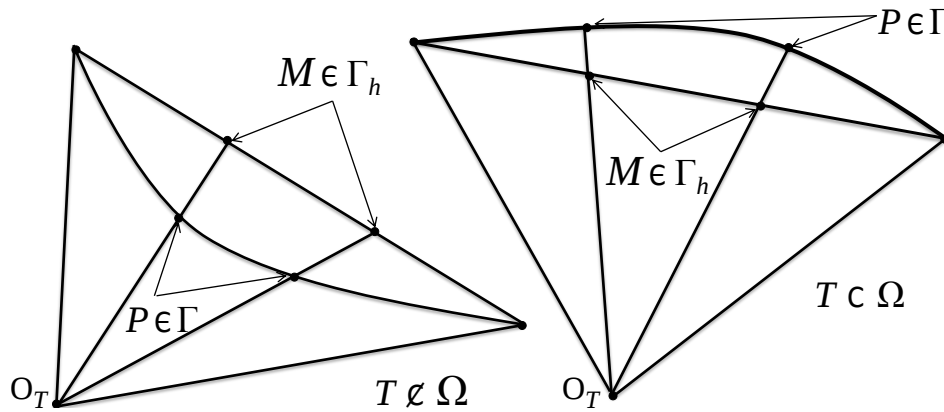


Figure 2: Construction of nodes  $P \in \Gamma$  for space  $W_h$  related to lagrangian nodes  $M \in \Gamma_h$  for  $k = 3$

*Remark 1* The construction of the nodes associated with  $W_h$  located on  $\Gamma$  advocated in item 4 is not mandatory. Notice that it differs from the intuitive construction of such nodes lying on normals to edges of  $\Gamma_h$  commonly used in the isoparametric technique. The main advantage of this proposal is an easy determination of boundary node coordinates by linearity, using a supposedly available analytical expression of  $\Gamma$ . Nonetheless the choice of boundary nodes ensuring our method's optimality is really wide, in contrast to the restrictions inherent to the interpolated boundary condition method (Brenner et al, 2008). ■

The fact that  $W_h$  is a non empty finite-dimensional space was established in Ruas (2017, 1). Furthermore the following result was also proved in the same reference:

**Proposition 2.1** (Ruas, 2017, 1)

Let  $\mathcal{P}_k(T)$  be the space of polynomials defined in  $T \in \mathcal{S}_h$  of degree less than or equal to  $k$ . Provided  $h$  is small enough  $\forall T \in \mathcal{S}_h$ , given a set of  $m_k$  real values  $b_i$ ,  $i = 1, \dots, m_k$  with  $m_k = (k + 1)k/2$ , there exists a unique function  $w_T \in \mathcal{P}_k(T)$  that vanishes at both vertices of  $T$  located on  $\Gamma$  and at the  $k - 1$  points  $P$  of  $\Gamma$  defined in accordance with item 4. of the above definition of  $W_h$ , and takes value  $b_i$  respectively at the  $m_k$  lagrangian nodes of  $T$  not located on  $\Gamma_h$ . ■

Now let us set the problem associated with spaces  $V_h$  and  $W_h$ , whose solution is an approx-

imation of  $u$ , that is, the solution of (1). Denoting by  $f'$  a sufficiently smooth extension of  $f$  to  $\Omega_h \setminus \Omega$  in case this set is not empty, and renaming  $f$  by  $f'$  in  $\Omega$ , we wish to solve,

$$\left\{ \begin{array}{l} \text{Find } u_h \in W_h \text{ such that} \\ a_h(u_h, v) = F_h(v) \quad \forall v \in V_h \\ \text{where } a_h(w, v) := \int_{\Omega_h} \mathbf{grad} w \cdot \mathbf{grad} v \text{ and } F_h(v) := \int_{\Omega_h} f' v. \end{array} \right. \quad (2)$$

The following result is borrowed from Ruas (2017, 1):

**Proposition 2.2** *Provided  $h$  is sufficiently small problem (2) has a unique solution.■*

## 2.2 Method assessment

In order to illustrate the accuracy and the optimal order of the method described in the previous subsection rigorously demonstrated in Ruas (2017,1), we implemented it taking  $k = 2$ . Then we solved equation (1) for several test-cases already reported in different papers, including Ruas (2017, 1). Here we only present results for the following one:

Test-problem in a non convex domain: The behavior of the *new method* in the solution of (1) is assessed for the following data:  $\Omega$  is the annulus delimited by the circles given by  $r = e < 1$  and  $r = 1$  with  $r^2 = x^2 + y^2$ ; for  $f := -\Delta u$  and  $d \equiv 0$  the exact solution  $u$  is given by  $u = (r - e)(1 - r)$ ,  $u$  being extended by  $u'$  defined by the same expression outside  $\Omega$ , and  $f$  by  $f' = -\Delta u'$ . We apply symmetry conditions on  $x = 0$  and  $y = 0$ , take  $e = 0.5$ , and compute with quasi-uniform meshes defined by two integer parameters  $I$  and  $J$ , constructed by subdividing the radial range  $(0.5, 1)$  into  $J$  equal parts and the angular range  $(0, \pi/4)$  into  $I$  equal parts. In this way the mesh of the quarter domain is the polar coordinate counterpart of the  $I \times J$  mesh of the rectangle  $(0, \pi/4) \times (0.5, 1)$  whose edges are parallel to the coordinate axes and to the line  $x = \pi(y - 0.5)/2$ .

In Table 1 we display the absolute errors in the norm  $\|\mathbf{grad}(\cdot)\|_{0,h}$  and in the norm of  $L^2(\Omega_h)$  for  $I = 2J$ , for increasing values of  $I$ , namely  $I = 2^m$  for  $m = 2, 3, 4, 5, 6$ . We also show the evolution of the maximum absolute errors at the (mesh) nodes denoted by *MAEN*. For each criterion we give an estimation of the **ACR**, an acronym for *asymptotic convergence rate*.

As one can observe, the quality of the approximations obtained with the *new method* is in very good agreement with the theoretical result given in Ruas (2017, 1), applying to the energy norm. Indeed, as  $J$  increases the errors in the gradient  $L^2$ -norm decrease roughly as  $h^2$ , as predicted. On the other hand the errors in the  $L^2$ -norm tend to decrease as  $h^3$  and the maximum absolute errors behave like an  $O(h^2)$ . The latter observation indicates that the quality of the maximum absolute errors at the nodes are not affected at all by the way boundary conditions are handled. This is because in case the classical approach is adopted one observes that this error is also an  $O(h^2)$ , while in case  $\Omega$  is a polygon it is known to be an  $O(h^3)$  for sufficiently smooth solutions (Nitsche, 1979).

## 3 THE THREE-DIMENSIONAL CASE

In this section we consider the solution of (1) by our method in case  $N = 3$ . In order to avoid non essential difficulties we make the assumption that no element in  $\mathcal{T}_h$  has more than one face on  $\Gamma_h$ , which is absolutely reasonable.

| $I$                                 | $\rightarrow$ | 4            | 8            | 16           | 32           | 64           | ACR      |
|-------------------------------------|---------------|--------------|--------------|--------------|--------------|--------------|----------|
| $\ \mathbf{grad}(u' - u_h)\ _{0,h}$ | $\rightarrow$ | 0.132906 E-1 | 0.334304 E-2 | 0.838061 E-3 | 0.209734 E-3 | 0.524545 E-4 | $O(h^2)$ |
| $\ u' - u_h\ _{0,h}$                | $\rightarrow$ | 0.400090 E-3 | 0.491773 E-4 | 0.610753 E-5 | 0.761759 E-6 | 0.951819 E-7 | $O(h^3)$ |
| $MAEN$                              | $\rightarrow$ | 0.923815 E-2 | 0.238444 E-2 | 0.600821 E-3 | 0.150500 E-3 | 0.376434 E-4 | $O(h^2)$ |

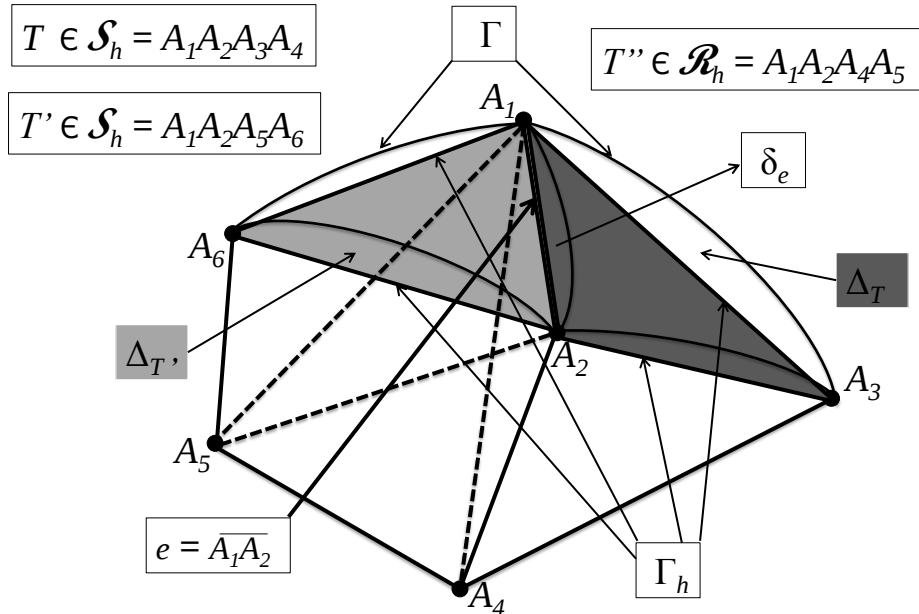
Table 1: Absolute errors in different senses for the test-problem in two-dimension space.

### 3.1 Method description

First of all we need some definitions regarding the set  $(\Omega \setminus \Omega_h) \cup (\Omega_h \setminus \Omega)$ .

Let  $\mathcal{S}_h$  be the subset of  $\mathcal{T}_h$  consisting of tetrahedra having one face on  $\Gamma_h$  and  $\mathcal{R}_h$  be the subset of  $\mathcal{T}_h \setminus \mathcal{S}_h$  of tetrahedra having exactly one edge on  $\Gamma_h$ . Notice that, owing to our initial assumption, no tetrahedron in  $\mathcal{T}_h \setminus [\mathcal{S}_h \cup \mathcal{R}_h]$  has a non empty intersection with  $\Gamma_h$ .

To every edge  $e$  of  $\Gamma_h$  we associate a plane skin  $\delta_e$  containing  $e$ , and delimited by  $\Gamma$  and  $e$  itself. Except for the fact that each skin contains an edge of  $\Gamma_h$ , its plane can be arbitrarily chosen. In Figure 3 we illustrate one out of three such skins corresponding to the edges of a face  $F_T$  or  $F_{T'}$  contained in  $\Gamma_h$ , of two tetrahedra  $T$  and  $T'$  belonging to  $\mathcal{S}_h$ . More precisely in Figure 3 we show the skin  $\delta_e$ ,  $e$  being the edge common to  $F_T$  and  $F_{T'}$ . Further, for every  $T \in \mathcal{S}_h$ , we define a set  $\Delta_T$  delimited by  $\Gamma$ , the face  $F_T$  and the three plane skins associated with the edges of  $F_T$ , as illustrated in Figure 3. In this manner we can assert that, if  $\Omega$  is convex,  $\Omega_h$  is a proper subset of  $\Omega$  and moreover  $\Omega$  is the union of the disjoint sets  $\Omega_h$  and  $\cup_{T \in \mathcal{S}_h} \Delta_T$  (cf. Figure 3). Otherwise  $\Omega_h \setminus \Omega$  is a non empty set that equals the union of certain parts of the sets  $\Delta_T$  corresponding to non convex portions of  $\Gamma$ .

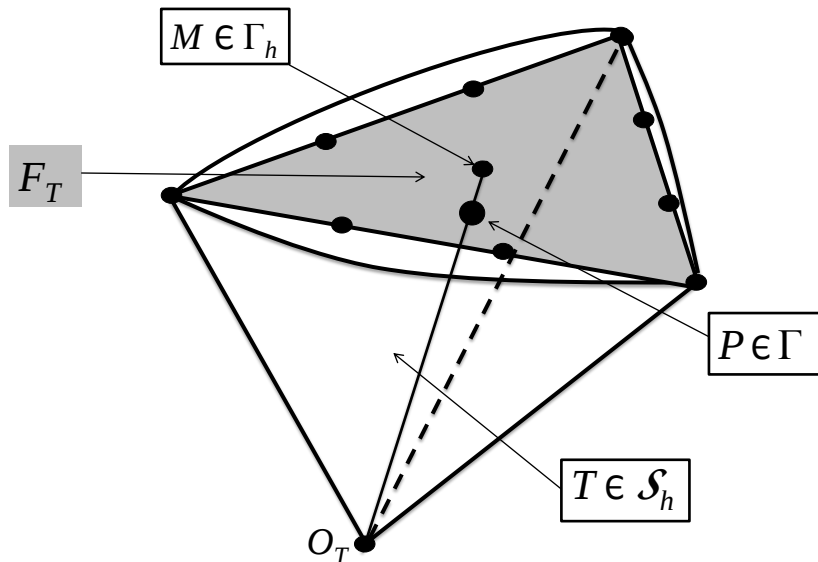
Figure 3: Sets  $\Delta_T$ ,  $\Delta_{T'}$ ,  $\delta_e$  for tetrahedra  $T, T' \in \mathcal{S}_h$  with a common edge  $e$  and a tetrahedron  $T'' \in \mathcal{R}_h$ 

Next we introduce two sets of functions  $V_h$  and  $W_h^d$ , both associated with  $\mathcal{T}_h$ .

$V_h$  is the standard Lagrange finite element space consisting of continuous functions  $v$  defined in  $\Omega_h$  that vanish on  $\Gamma_h$ , whose restriction to every  $T \in \mathcal{T}_h$  is a polynomial of degree less than or equal to  $k$  for  $k \geq 2$ . For convenience we extend by zero every function  $v \in V_h$  to  $\Omega \setminus \Omega_h$ . We recall that a function in  $V_h$  is uniquely defined by its values at the points which are vertices of the partition of each mesh tetrahedron into  $k^3$  equal tetrahedra as described in Ciarlet (1978) or Zienkiewicz (1977). Akin to the two-dimensional case these points will be referred to as the *lagrangian nodes* of order  $k$  of the mesh.

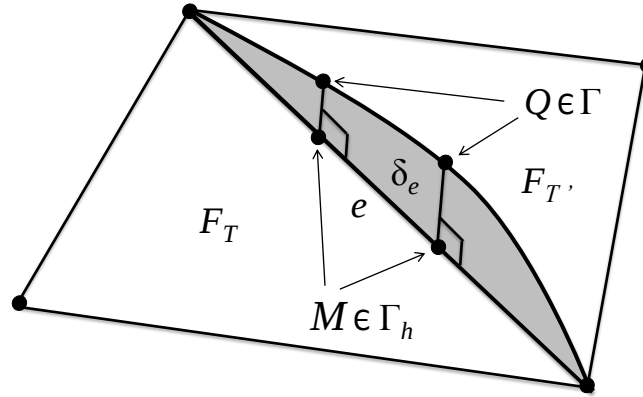
$W_h^d$  in turn is a linear manifold consisting of functions defined in  $\Omega_h \cup \Omega$  satisfying the following conditions:

1. The restriction of  $w \in W_h^d$  to every  $T \in \mathcal{T}_h$  is a polynomial of degree less than or equal to  $k$ ;
2. Every  $w \in W_h^d$  is single-valued at all the inner lagrangian nodes of the mesh, that is all its lagrangian nodes of order  $k$  except those located on  $\Gamma_h$ ;
3. A function  $w \in W_h^d$  is also defined in  $\Omega \setminus \Omega_h$  in such a way that its polynomial expression in  $T \in \mathcal{S}_h$  also applies to points in  $\Delta_T$ ;
4. A function  $w \in W_h^d$  takes the value  $d(S)$  at any vertex  $S$  of  $\Gamma_h$ ;
5.  $\forall T \in \mathcal{S}_h$  and for  $k > 2$ ,  $w(P) = d(P)$  for every point  $P$  among the  $(k-1)(k-2)/2$  intersections with  $\Gamma$  of the line passing through the vertex  $O_T$  of  $T$  not belonging to  $\Gamma$  and the  $(k-1)(k-2)/2$  points  $M$  not belonging to any edge of  $F_T$  among the  $(k+2)(k+1)/2$  points of  $F_T$  that subdivide this face (opposite to  $O_T$ ) into  $k^2$  equal triangles (see illustration in Figure 4 for  $k = 3$ );
6.  $\forall T \in \mathcal{S}_h \cup \mathcal{R}_h$ ,  $w(Q) = d(Q)$  for every  $Q$  among the  $k-1$  intersections with  $\Gamma$  of the line orthogonal to  $e$  in the skin  $\delta_e$ , passing through the points  $M \in e$  different from vertices of  $T$ , subdividing  $e$  into  $k$  equal segments, where  $e$  represents a generic edge of  $T$  contained in  $\Gamma_h$  (see illustration in Figure 5 for  $k = 3$ ).



**Figure 4:** Construction of node  $P \in \Gamma$  of  $W_h^d$  related to the Lagrange node  $M$  in the interior of  $F_T \subset \Gamma_h$

*Remark 2* The construction of the nodes associated with  $W_h^d$  located on  $\Gamma$  advocated in items 5.



**Figure 5:** Construction of nodes  $Q \in \Gamma \cap \overline{\delta_e}$  of  $W_h^d$  related to the Lagrange nodes  $M \in e \subset \Gamma_h$

and 6. is not mandatory. Notice that it differs from the intuitive construction of such nodes lying on normals to faces of  $\Gamma_h$  commonly used in the isoparametric technique. The main advantage of this proposal is the determination by linearity of the coordinates of the boundary nodes  $P$  in the case of item 5. Nonetheless, akin to the two-dimensional case, the choice of boundary nodes ensuring our method's optimality is absolutely very wide. ■

The fact that  $W_h^d$  is a non empty set is a trivial consequence of the two following results proved in Ruas (2017, 2), where  $\mathcal{P}_k(T)$  represents the space of polynomials defined in  $T \in \mathcal{S}_h \cup \mathcal{R}_h$  of degree not greater than  $k$ .

**Proposition 3.1** (Ruas, 2017, 2)

Provided  $h$  is small enough  $\forall T \in \mathcal{S}_h \cup \mathcal{R}_h$  given a set of  $m_k$  real values  $b_i$ ,  $i = 1, \dots, m_k$  with  $m_k = k(k+1)(k+2)/6$  for  $T \in \mathcal{S}_h$  and  $m_k = (k+1)(k+2)(k+3)/6 - (k+1)$  for  $T \in \mathcal{R}_h$ , there exists a unique function  $w_T \in \mathcal{P}_k(T)$  that takes the value of  $d$  at the vertices  $S$  of  $T$  located on  $\Gamma$ , at the points  $P$  of  $\Gamma$  defined in accordance with item 5. for  $T \in \mathcal{S}_h$  only, and at the points  $Q$  of  $\Gamma$  defined in accordance with item 6. of the above definition of  $W_h^d$ , and takes the value  $b_i$  respectively at the  $m_k$  lagrangian nodes of  $T$  not located on  $\Gamma_h$ . ■

A well-posedness result analogous to Proposition 2.2 holds for problem (3), according to Ruas (2017, 2), namely,

**Proposition 3.2** (Ruas, 2017, 2)

As long as  $h$  is sufficiently small problem (3) has a unique solution. ■

*Remark 3* It is important to stress that, in contrast to its two-dimensional counterpart, the set  $W_h^d$  does not necessarily consist of continuous functions. This is because of the interfaces between elements in  $\mathcal{R}_h$  and  $\mathcal{S}_h$ . Indeed a function  $w \in W_h^d$  is not forcibly single-valued at all the lagrangian nodes located on one such an interface, owing to the enforcement of the boundary condition at the points  $Q \in \Gamma$  instead of the corresponding lagrangian node  $M \in \Gamma_h$ , in accordance with item 6. in the definition of  $W_h^d$ . On the other hand  $w$  is necessarily continuous over all other faces common to two mesh tetrahedra. ■

Next we set the problem associated with the space  $V_h$  and the manifold  $W_h^d$ , whose solution is an approximation of  $u$ , that is, the solution of (1). Extending  $f$  by a smooth  $f'$  in  $\Omega_h \setminus \Omega$  if

necessary, and renaming  $f$  by  $f'$  in any case, we wish to solve,

$$\left\{ \begin{array}{l} \text{Find } u_h \in W_h^d \text{ such that} \\ a_h(u_h, v) = F_h(v) \quad \forall v \in V_h \\ \text{where } a_h(w, v) := \int_{\Omega_h} \mathbf{grad} w \cdot \mathbf{grad} v \text{ and } F_h(v) := \int_{\Omega_h} f' v. \end{array} \right. \quad (3)$$

### 3.2 Method assessment

In this section we assess the accuracy of the method studied in the previous subsection - referred to hereafter as the *new method* -, by solving equation (1) in two relevant test-cases, taking  $k = 2$ . A comparison with the approach consisting of shifting boundary conditions from the true boundary to the boundary of the approximating polyhedron is also carried out. Hereafter the latter approach will be called the *simple method*.

**Test-problem with a quadratic solution:** Our model problem (1) was solved with a constant right hand side equal to  $2(a^{-2} + b^{-2} + 1)$  in the ellipsoid centered at the origin, whose equation is  $p(x, y, z) = 1$  where  $p(x, y, z) = (x/a)^2 + (y/b)^2 + z^2$ . The exact solution for  $d \equiv 1$  is the quadratic function  $-p$ , and thus the *new method* is supposed to reproduce it up to machine precision for any mesh. Here we used a mesh consisting of 3072 tetrahedra resulting from the transformation of a standard uniform  $6 \times 8 \times 8 \times 8$  mesh of a unit cube  $\Omega_0$  into tetrahedra having one edge coincident with a diagonal parallel to the line  $x = y = z$  of a cube with edge equal to  $1/8$ , resulting from a first subdivision of  $\Omega_0$  into  $8^3$  equal cubes. The final tetrahedral mesh of the ellipsoid octant corresponding to positive values of  $x, y, z$ , contains the same number of elements and is generated by mapping the unit cube into the latter domain through the transformation of cartesian coordinates into spherical coordinates using a procedure described in Ruas (1986).

It turns out that the absolute error in the  $H^1$ -semi-norm  $\|\mathbf{grad}(\cdot)\|_{0,h}$  resulting from computations with  $a = 0.6$  and  $b = 0.8$ , equals approximately  $0.29896592 \times 10^{-7}$ , for an exact value of ca. 1.0324886. This means that the numerical solution is exact indeed, up to machine precision. On the other hand the absolute error measured in the same way for the *simple method* is about 0.01663104, i.e., a relative error of about 1.6 percent. One might object that this is not so bad for a rather coarse mesh. However substantial gains with the *new method* over the *simple method* will be manifest in the example that follows.

**Test-problem in a convex domain:** We next validate the error estimate in the energy norm given in Ruas (2017, 2), and assess method's accuracy in the  $L^2$ -norm of the error function  $u - u_h$  in  $\Omega_h$ . Here  $\Omega$  is the unit sphere centered at the origin. We take the exact solution  $u = \rho^2 - \rho^4$  where  $\rho^2 = x^2 + y^2 + z^2$ , which means that  $f = -6 + 20\rho^2$ . Owing to symmetry we consider only the octant sub-domain given by  $x > 0, y > 0$  and  $z > 0$  by prescribing Neumann boundary conditions on  $x = 0, y = 0$  and  $z = 0$ . We computed with quasi-uniform meshes defined by a single integer parameter  $J$ , constructed by the procedure proposed in Ruas (1986) and described in main lines at the beginning of this section. Roughly speaking the mesh of the computational sub-domain is the spherical-coordinate counterpart of the standard  $J \times J \times J$  uniform mesh of the unit cube  $(0, 1) \times (0, 1) \times (0, 1)$ . Each tetrahedron of the final mesh results from the transformation of the groups of six tetrahedra generated by the subdivision of each cubic cell

| $h$                                | $\rightarrow$ | 1/4          | 1/8          | 1/12         | 1/16         | 1/20         | ACR      |
|------------------------------------|---------------|--------------|--------------|--------------|--------------|--------------|----------|
| $\ \mathbf{grad}(u - u_h)\ _{0,h}$ | $\rightarrow$ | 0.187649 E-1 | 0.499091 E-2 | 0.225836 E-2 | 0.128114 E-2 | 0.823972 E-3 | $O(h^2)$ |
| $\ u - u_h\ _{0,h}$                | $\rightarrow$ | 0.653073 E-3 | 0.845686 E-4 | 0.253348 E-4 | 0.107516 E-4 | 0.552583 E-5 | $O(h^3)$ |

**Table 2: Absolute errors with the *new method* measured in two different manners.**

| $h$                                | $\rightarrow$ | 1/4          | 1/8          | 1/12           | 1/16         | 1/20         | ACR          |
|------------------------------------|---------------|--------------|--------------|----------------|--------------|--------------|--------------|
| $\ \mathbf{grad}(u - u_h)\ _{0,h}$ | $\rightarrow$ | 0.257134 E-1 | 0.917910 E-2 | 0.50152682 E-2 | 0.326410 E-2 | 0.233854 E-2 | $O(h^{1.5})$ |
| $\ u - u_h\ _{0,h}$                | $\rightarrow$ | 0.454733 E-2 | 0.113568 E-2 | 0.502166 E-3   | 0.281468 E-3 | 0.179698 E-3 | $O(h^2)$     |

**Table 3: Absolute errors with the *simple method* measured in two different manners.**

of the partition of the unit cube using their diagonals parallel to the line  $x = y = z$ . Since the mesh is symmetric with respect to the three cartesian axes only one third of the chosen octant sub-domain was actually taken into account in the computations.

In Table 2 we display the absolute errors in the norms  $\|\mathbf{grad}(\cdot)\|_{0,h}$  and  $\|\cdot\|_{0,h}$  for increasing values of  $J$ , namely,  $J = 4, 8, 12, 16, 20$ . Since the true value of  $h$  equals  $c/J$  for a suitable constant  $c$ , as a reference we set  $h = 1/J$  to simplify things. As one infers from Table 2, the approximations obtained with the *new method* perfectly conform to the theoretical estimate given in Ruas (2017, 2). Indeed as  $J$  increases the errors in the gradient  $L^2$ -norm decrease roughly like  $h^2$ , as predicted. The error in the  $L^2$ -norm in turn tends to decrease as an  $O(h^3)$ . In Table 3 we display the same kind of results obtained with the *simple method*. As one can observe the error in the gradient  $L^2$ -norm decreases roughly like  $h^{1.5}$ , as predicted by the mathematical theory of the finite element method, while the errors in the  $L^2$ -norm seem to behave like an  $O(h^2)$ .

## 4 FINAL COMMENTS

We conclude this work with the following comments or remarks.

1. The method addressed in this work to solve the Poisson equation with Dirichlet boundary conditions in curved domains with classical Lagrange finite elements provides a simple and reliable manner to overcome technical difficulties brought about by more complicated problems and interpolations. For example, Hermite finite element methods to solve fourth order problems in curved domains with normal derivative degrees of freedom can also be dealt with very easily by means of our new method. The author intends to show this in a paper to appear shortly.
2. The technique studied in this paper is also particularly handy, to treat problems posed in curved domains in terms of vector fields, such as the linear elasticity system and

Maxwell's equations of electromagnetism. The same remark applies to multi-field systems such as the Navier-Stokes equations, and more generally mixed formulations of several types with curved boundaries, to be approximated by the finite element method. In the latter respect for instance, the definition of isoparametric finite element analogs is not always so clear or straightforward (Bertrand et al, 2014).

3. As for the Poisson equation with homogeneous Neumann boundary conditions  $\partial u / \partial n = 0$  on  $\Gamma$  (provided  $f$  satisfies the underlying scalar condition) our method coincides with the standard Lagrange finite element method. Notice that if inhomogeneous Neumann boundary conditions are prescribed, optimality can only be recovered if the linear form  $F_h$  is modified, in such a way that boundary integrals for elements  $T \in \mathcal{S}_h$  are shifted to the curved boundary portion sufficiently close to  $\Gamma$  of an extension or reduction of  $T$ . But this is an issue that has nothing to do with our method, which is basically aimed at resolving those related to the prescription of degrees of freedom in the case of Dirichlet boundary conditions.
4. Finally we note that our method leads to linear systems of equations with a non symmetric matrix, even when the original problem is symmetric. Moreover in order to compute the element matrix and right side vector for an element  $T$  in  $\mathcal{S}_h$  or in  $\mathcal{R}_h$ , the inverse of an  $n_k \times n_k$  matrix has to be computed, where  $n_k$  is the dimension of  $P_k(T)$ . However this extra effort is not really a problem nowadays, in view of the significant progress already accomplished in Computational Linear Algebra.

## Acknowledgements

The author gratefully acknowledges the financial support provided by CNPq through grant 307996/2008-5.

## REFERENCES

- Bertrand, F., Münzenmaier, S., & Starke, G., 2014. First-order system least-squares on curved boundaries: higher-order Raviart–Thomas elements. *SIAM J. Numerical Analysis*, Vol. 52, n.6, pp. 3165–3180.
- Brenner S. C., & Scott, L. R., 2008. *The Mathematical Theory of Finite Element Methods*. Texts in Applied Mathematics 15, Springer.
- Ciarlet, P. G., 1978. *The Finite Element Method for Elliptic Problems*, North Holland.
- Nitsche, J. A., 1972. On Dirichlet problems using subspaces with nearly zero boundary conditions. In Aziz, A.K. ed, *The Mathematical Foundations of the Finite Element Method with Applications to Partial Differential Equations*, Academic Press, pp. 603–627.
- Nitsche, J. A., 1979.  $L^\infty$ -convergence of finite element Galerkin approximations for parabolic problems. *ESAIM: Mathematical Modelling and Numerical Analysis*, Vol. 13, n. 1, pp. 31–54.

- Ruas, V., 1986. One-parameter mesh generation for spheroidal domains. In Taylor, C., ed, *Proceedings of the First International Conference on Numerical Grid Generation in Computational Fluid Dynamics, Landshut, Germany*, pp. 71–82.
- Ruas, V., 2017, 1. Optimal simplex finite-element approximations of arbitrary order in curved domains circumventing the isoparametric technique, *arXiv Numerical Analysis*, arXiv:1701.00663 [math.NA].
- Ruas, V., 2017, 2. Methods of arbitrary optimal order with tetrahedral finite-element meshes forming polyhedral approximations of curved domains, *arXiv Numerical Analysis*, arXiv:1706.08004 [math.NA].
- Scott, L. R., 1973. *Finite Element Techniques for Curved Boundaries*, PhD thesis, MIT.
- Zienkiewicz, O. C., 1977. *The Finite Element Method*, 3rd ed., McGraw-Hill, London.