

# A Best Approximation Property of the Moving Finite Element Method

P.K. Jimack  
School of Computer Studies  
University of Leeds  
Leeds LS2 9JT, UK

March 25, 1996

**Key words.** Moving Finite Element, Steady Solutions, Best Free Knot Approximations  
**AMS(MOS) subject classification.** 65M60

## Abstract

The Moving Finite Element method for the solution of time-dependent partial differential equations is a numerical solution scheme which allows the automatic adaption of the finite element approximation space with time. An analysis of how this method models the steady solutions of a general class of parabolic linear source equations is presented. It is shown that under certain conditions the steady solutions of the Moving Finite Element problem can correspond to best free knot spline approximations to the true steady solution of the differential equation when using the natural norm associated with the problem. Hence a quantitative measure of the advantages of the Moving Finite Element method over the usual fixed grid Galerkin method is produced for these equations. A number of numerical examples are included to illustrate these results.

## 1 Introduction

The Moving Finite Element (MFE) method for the solution of time-dependent partial differential equations was first introduced by Miller *et al* ([9], [21], and [22]) in 1981. It is a finite element method in which a spatial mesh with a constant number of degrees of freedom is allowed to deform continuously in time. Unlike in [11], [19] or [26] for example, this is achieved without tying the node positions to individually tracked solution properties such as characteristic speeds or the motions of internal boundaries. Instead these positions are treated as unknown time-dependent variables which, just like the conventional finite element degrees of freedom, must be evaluated as part of the solution procedure. This procedure is designed to simultaneously determine at each time both a suitable spatial mesh and an approximation to the solution on that mesh.

This paper considers the application of the MFE method to the solution of the following family of linear second order evolution equations:

$$\frac{\partial u}{\partial t}(\underline{x}, t) = \frac{\partial}{\partial x_\mu} [p_{\mu\nu}(\underline{x}) \frac{\partial u}{\partial x_\nu}(\underline{x}, t)] - q(\underline{x})u(\underline{x}, t) + r(\underline{x}), \quad \forall \underline{x} \in \Omega \subseteq \Re^d \text{ and } t \in (0, T]. \quad (1.1)$$

Here  $d$  may be any positive integer (but is typically 1,2 or 3),  $T > 0$  and the usual summation convention (summing from 1 to  $d$ ) applies over the repeated suffices  $\mu$  and  $\nu$ . Also we assume that the matrix  $P(\underline{x})$ , whose entries are  $p_{\mu\nu}(\underline{x})$ , is symmetric and positive definite and that

$$\text{each } p_{\mu\nu} \in C^{1,\alpha}(\Omega), \quad q, r \in C^{0,\alpha}(\Omega) \quad \text{and} \quad q \geq 0 \quad \forall \underline{x} \in \Omega,$$

for some exponent  $\alpha$  of Hölder continuity in  $(0, 1)$ . This set of assumptions is sufficient to ensure that the equation (1.1) is well-defined and, provided it is solved subject to suitable conditions on an appropriate boundary  $\partial\Omega$  (satisfying an exterior ball condition say), has a unique classical steady solution (as described in [10, section 6.3] for example).

For the theoretical part of this paper we restrict our consideration to the particular case of homogeneous Dirichlet conditions throughout  $\Omega$  in order to keep the theory as clear as possible. This can easily be extended however and in [14] a wider variety of possible boundary conditions are discussed in some detail. We also assume for simplicity that the initial solution  $u(\underline{x}, 0)$  satisfies the applied boundary conditions and is everywhere twice differentiable.

The main result of this paper shows that when the Moving Finite Element method is used to solve (1.1), if it tends to a steady solution then this solution corresponds to a locally best approximation of the true steady solution of the p.d.e. from the space of all possible free knot linear splines. That is, in a particular norm, the error in the approximation to the true steady solution by the steady MFE solution is at a local minimum in the manifold of free knot linear splines. Such a result was suspected by Miller in one of his original papers on the MFE method ([22]) and it is proved here using a style and framework more akin to that of Delfour *et al* in [6] than that used by Miller.

An alternative way of expressing the result is to say that successful use of the Moving Finite Element method is as good, in terms of approximating the steady solution, as using the fixed grid Galerkin method on a best possible choice of spatial mesh. This is significant because it is one of the few analytical results about the Moving Finite Element method that is able to quantify its advantages over the fixed grid Galerkin method for this type of parabolic equation. One of the only other published results of this type is the slightly weaker result of Dupont [8], who proves that for a certain class of parabolic equations with smooth solutions the Moving Finite Element method, under the influence of sufficiently strong penalty functions of the type used by Miller [22], is asymptotically no worse than a fixed-grid method.

Although it is necessary to restrict the theory in this paper to the particular case of equations of the form (1.1) the MFE method may of course be applied to a much larger variety of problems. It is expected that any insight that it is possible to gain by looking at straightforward linear equations such as (1.1) will be of use when attempting to understand and analyze harder problems.

The following section of the paper briefly introduces the Moving Finite Element method and derives the governing equations when it is applied to (1.1). Only sufficient details to establish enough background and notation for the rest of the paper are included. Further details of the procedure and its implementation can be found elsewhere: in [1], [9] or [27] for example. Section 3 contains the bulk of the theory in which the result outlined above is derived and discussed, then in section 4 a number of numerical examples are given in order to verify the theoretical results and put them in a computational context. The paper ends with a short discussion.

## 2 The Moving Finite Element Method

In this section we give a brief outline of the Moving Finite Element method and how it can be applied to the solution of equation (1.1) with homogeneous Dirichlet boundary conditions on  $\partial\Omega$ . For clarity these simple boundary conditions are considered throughout this paper: see [14] for a discussion of more general boundary conditions.

In order to proceed it will be helpful to introduce some notation. In the first instance we will assume that the spatial domain  $\Omega$  is fixed for all time and is such that its boundary  $\partial\Omega$  can be covered exactly by simplexes of dimension  $d - 1$ . In the case  $d = 2$  this means that the domain boundary is polygonal. Given this, it is possible to discretize  $\Omega$  into a set of non-overlapping simplexes of dimension  $d$  (triangles when  $d = 2$ ). This discretization can be uniquely specified as a

mesh  $\mathcal{M} = (\underline{s}, \mathcal{C})$ . Here

$$\underline{s} = (\underline{s}_1, \dots, \underline{s}_N, \underline{s}_{N+1}, \dots, \underline{s}_{N+B}) \quad (2.1)$$

is an ordered set of the position vectors of the vertices of the  $d$ -dimensional triangulation and  $\mathcal{C}$  is a list of all of its edges. In (2.1) the vertices or knot points are ordered such that there are  $N$  vertices strictly inside  $\Omega$  followed by a further  $B$  vertices on  $\partial\Omega$ .

The Moving Finite Element method seeks to approximate  $u(\underline{x}, t)$ , the solution of (1.1), by a time-dependent piecewise linear function,  $v$  say, defined on a mesh of simplexes  $\mathcal{M}(t) = (\underline{s}(t), \mathcal{C})$  covering the spatial domain  $\Omega$ . Unlike in the conventional Galerkin method, this mesh is allowed to deform smoothly in time by allowing the positions of the internal knot points,  $\underline{s}_1(t), \dots, \underline{s}_N(t)$ , to be time-dependent. Their connectivity  $\mathcal{C}$  remains fixed however.

Because  $\mathcal{C}$  is kept fixed throughout we will generally refer to a mesh  $\mathcal{M}(t) = (\underline{s}(t), \mathcal{C})$  only by the ordered set  $\underline{s}(t)$  for notational convenience. Note that a mesh is only a valid finite element triangulation if the position of the knot points for a given connectivity is such that the measure of each simplex within the mesh is strictly positive. Given that this is the case we can write our approximation  $v$  in the form

$$v(\underline{x}, t) = \sum_{i=1}^N a_i(t) \alpha_i(\underline{x}, \underline{s}(t)) , \quad (2.2)$$

where  $\alpha_i$  is the usual continuous piecewise linear ‘‘hat’’ basis function on the mesh  $\underline{s}(t)$ :

$$\alpha_i(\underline{s}_j(t), \underline{s}(t)) = \delta_{ij} , \quad i = 1, \dots, N ; \quad j = 1, \dots, N + B .$$

The sum only goes from 1 to  $N$  because of the homogeneous Dirichlet boundary conditions on  $\partial\Omega$ .

In order to determine this approximation to  $u(\underline{x}, t)$  we need to find values for the unknowns  $a_1(t), \underline{s}_1(t), \dots, a_N(t), \underline{s}_N(t)$ . The Moving Finite Element method does this by producing a weak form of (1.1) for which the trial solution  $v$  takes the form of (2.2) and the test space is the space in which the function  $\frac{\partial v}{\partial t}$  lies at each instant in time. In order to determine this space we differentiate (2.2) with respect to time to give

$$\begin{aligned} \frac{\partial v}{\partial t} &= \frac{\partial}{\partial t} \sum_{i=1}^N a_i(t) \alpha_i(\underline{x}, \underline{s}(t)) \\ &= \sum_{i=1}^N \dot{a}_i \alpha_i + \sum_{i=1}^N a_i \underline{\nabla}_s \alpha_i \cdot \frac{d\underline{s}}{dt} , \end{aligned} \quad (2.3)$$

where this second term is present due to the time-dependence of each  $\alpha_i$  through the time-dependence of the mesh  $\underline{s}$ , and the gradient operator  $\underline{\nabla}_s$  applies to the  $\underline{s}$  variables only. Hence

$$\begin{aligned} \frac{\partial v}{\partial t} &= \sum_{i=1}^N \dot{a}_i \alpha_i + \underline{\nabla}_s v \cdot \frac{d\underline{s}}{dt} \\ &= \sum_{i=1}^N \dot{a}_i \alpha_i + \sum_{i=1}^N \dot{\underline{s}}_i \cdot \frac{\partial v}{\partial \underline{s}_i} \\ &= \sum_{i=1}^N (\dot{a}_i \alpha_i + \dot{\underline{s}}_i \cdot \underline{\beta}_i) , \end{aligned} \quad (2.4)$$

where  $\underline{\beta}_i = \frac{\partial v}{\partial \underline{s}_i} = (\frac{\partial v}{\partial s_{i1}}, \dots, \frac{\partial v}{\partial s_{id}})^T$  and the dot above a variable denotes differentiation with respect to time. In fact it can be shown (see [15] or [20] for example) that

$$\underline{\beta}_i = \frac{\partial v}{\partial \underline{s}_i} = -\alpha_i \underline{\nabla} v , \quad \text{and hence} \quad \beta_{i\ell} = -\alpha_i \frac{\partial v}{\partial x_\ell} \quad \text{for } \ell = 1, \dots, d . \quad (2.5)$$

Hence in order to minimize the p.d.e. residual over all possible choices of  $\frac{\partial v}{\partial t}$  the Moving Finite Element method takes a weak form of (1.1) for which the test space is the space spanned by the functions

$$\{\alpha_1, \beta_{11}, \dots, \beta_{1d}; \dots, \alpha_N, \beta_{N1}, \dots, \beta_{Nd}\}.$$

The most straightforward weak form of this type is the simple generalization of the Galerkin method given formally by the differential system

$$\left\langle \sum_{i=1}^N (\dot{a}_i \alpha_i + \dot{\underline{s}}_i \cdot \underline{\beta}_i), \alpha_j \right\rangle = \left\langle \frac{\partial}{\partial x_\mu} [p_{\mu\nu} \frac{\partial v}{\partial x_\nu}], \alpha_j \right\rangle - \langle qv - r, \alpha_j \rangle \quad (2.6)$$

and

$$\left\langle \sum_{i=1}^N (\dot{a}_i \alpha_i + \dot{\underline{s}}_i \cdot \underline{\beta}_i), \beta_{jm} \right\rangle = \left\langle \frac{\partial}{\partial x_\mu} [p_{\mu\nu} \frac{\partial v}{\partial x_\nu}], \beta_{jm} \right\rangle - \langle qv - r, \beta_{jm} \rangle \quad (2.7)$$

for the unknowns  $a_1(t), \underline{s}_1(t), \dots, a_N(t), \underline{s}_N(t)$ , with  $j = 1, \dots, N$  and  $m = 1, \dots, d$ . In the above notation  $\langle \cdot, \cdot \rangle$  represents the usual  $L^2$  inner product on  $\Omega$  and summation is again implied over the repeated suffices  $\mu$  and  $\nu$  (as will be the case throughout this paper).

It should be noted at this point however that the second of these sets of equations is not properly defined for a piecewise linear function  $v(\underline{x}, t)$ , even in a distributional sense. To overcome this difficulty it is necessary to express these equations in a formally equivalent form which is well-defined for such functions  $v$ . This can be achieved by applying the following integration by parts argument, similar to that in [23], to the first term on the right-hand-side of (2.7):

$$\begin{aligned} \left\langle \frac{\partial}{\partial x_\mu} [p_{\mu\nu} \frac{\partial v}{\partial x_\nu}], -\alpha_j \frac{\partial v}{\partial x_m} \right\rangle &= \left\langle p_{\mu\nu} \frac{\partial v}{\partial x_\nu}, \alpha_j \frac{\partial^2 v}{\partial x_\mu \partial x_m} + \frac{\partial \alpha_j}{\partial x_\mu} \frac{\partial v}{\partial x_m} \right\rangle \quad (2.8) \\ &= \frac{1}{2} \left\langle \alpha_j p_{\mu\nu}, \left[ \frac{\partial v}{\partial x_\nu} \frac{\partial}{\partial x_m} \left( \frac{\partial v}{\partial x_\mu} \right) + \frac{\partial v}{\partial x_\mu} \frac{\partial}{\partial x_m} \left( \frac{\partial v}{\partial x_\nu} \right) \right] \right\rangle \\ &\quad + \left\langle p_{\mu\nu} \frac{\partial v}{\partial x_\nu}, \frac{\partial \alpha_j}{\partial x_\mu} \frac{\partial v}{\partial x_m} \right\rangle \\ &\quad \text{(using the symmetry } p_{\mu\nu} = p_{\nu\mu} \text{)} \\ &= \frac{1}{2} \left\langle \alpha_j p_{\mu\nu}, \frac{\partial}{\partial x_m} \left( \frac{\partial v}{\partial x_\mu} \frac{\partial v}{\partial x_\nu} \right) \right\rangle + \left\langle \frac{\partial v}{\partial x_m} \frac{\partial v}{\partial x_\nu}, p_{\mu\nu} \frac{\partial \alpha_j}{\partial x_\mu} \right\rangle \\ &= -\frac{1}{2} \left\langle \frac{\partial}{\partial x_m} (\alpha_j p_{\mu\nu}), \frac{\partial v}{\partial x_\mu} \frac{\partial v}{\partial x_\nu} \right\rangle + \left\langle \frac{\partial v}{\partial x_m} \frac{\partial v}{\partial x_\nu}, p_{\mu\nu} \frac{\partial \alpha_j}{\partial x_\mu} \right\rangle. \quad (2.9) \end{aligned}$$

The last line of this expression is defined for piecewise linear functions  $v$  and so can be used in the definition of the Moving Finite Element equations (derived from (2.6) and (2.7)):

$$\sum_{i=1}^N \langle \alpha_i, \alpha_j \rangle \dot{a}_i + \sum_{i=1}^N \sum_{\ell=1}^d \langle \beta_{i\ell}, \alpha_j \rangle \dot{s}_{i\ell} = - \left\langle p_{\mu\nu} \frac{\partial v}{\partial x_\nu}, \frac{\partial \alpha_j}{\partial x_\mu} \right\rangle - \langle qv - r, \alpha_j \rangle \quad (2.10)$$

and

$$\begin{aligned} \sum_{i=1}^N \langle \alpha_i, \beta_{jm} \rangle \dot{a}_i + \sum_{i=1}^N \sum_{\ell=1}^d \langle \beta_{i\ell}, \beta_{jm} \rangle \dot{s}_{i\ell} &= -\frac{1}{2} \left\langle \frac{\partial v}{\partial x_\mu} \frac{\partial v}{\partial x_\nu}, \frac{\partial}{\partial x_m} (\alpha_j p_{\mu\nu}) \right\rangle \\ &\quad + \left\langle \frac{\partial v}{\partial x_m} \frac{\partial v}{\partial x_\nu}, p_{\mu\nu} \frac{\partial \alpha_j}{\partial x_\mu} \right\rangle - \langle qv - r, \beta_{jm} \rangle \quad (2.11) \end{aligned}$$

for  $j = 1, \dots, N$  and  $m = 1, \dots, d$ . Note that our use of homogeneous Dirichlet boundary conditions here has again simplified things by ensuring that there are no boundary integrals present in these

equations. Also, some authors prefer to derive these equations in a slightly different manner, using mollification ([21],[22]) or recovery methods ([16],[7]) to deal with the second order terms. (It is important to realize that the mollification approach of Miller and the above intergration by parts approach are themselves intrinsically related, see [2] for example, where the issue of second order terms is considered in detail.)

As has already been implied, the sets of equations (2.10) and (2.11) are referred to as the Moving Finite Element equations. They form a system of ordinary differential equations which may be written in the form

$$A(\underline{y})\dot{\underline{y}} = \underline{g}(\underline{y}), \quad (2.12)$$

where

$$\begin{aligned} \underline{y} &= (a_1, s_{11}, \dots, s_{1d}; \dots; a_N, s_{N1}, \dots, s_{Nd})^T, \\ \underline{\alpha} &= (\alpha_1, \beta_{11}, \dots, \beta_{1d}; \dots; \alpha_N, \beta_{N1}, \dots, \beta_{Nd})^T, \\ A &= \langle \underline{\alpha}, \underline{\alpha}^T \rangle \end{aligned}$$

and  $\underline{g}$  is the vector of right-hand-sides. The matrix  $A(\underline{y})$  is often called the ‘‘MFE mass matrix’’ by analogy with the usual Galerkin mass matrix.

It should be noted that even though (1.1) is linear, the Moving Finite Element semi-discretization yields a nonlinear system of differential equations. Also, although the matrix  $A$  can be shown to be positive semi-definite, it may become singular for certain values of the solution parameters  $\underline{y}$ . This occurs when the elements of the ordered set  $\underline{\alpha}$ , defined above, form a linearly dependent set. This can be shown to be happen if and only if the MFE solution  $v$  has a directional derivative which is continuous at one or more of the knot points  $\underline{s}_1, \dots, \underline{s}_N$  ([28]). If this is the case (2.12) becomes a differential-algebraic system and the problem is said to be ‘‘degenerate’’. When this is not the case we will refer to the MFE solution as being ‘‘non-degenerate’’ and we note that for such solutions the MFE mass matrix,  $A(\underline{y})$ , is strictly positive definite.

The problem of degeneracy along with the possibility of the measure of one or more of the simplexes in the mesh becoming non-positive as the knot points evolve are often cited as two of the major drawbacks of the MFE method. One approach to overcoming these difficulties is to attempt to influence the nodal motion by using penalty functions in the underlying minimization to which equations (2.10) and (2.11) correspond. This is the approach of Miller *et al* ([9], [21], [22]) and Mueller and Carey [24] for example. However, the work of Baines *et al* ([1], [3], [4], [16], [28]), mainly, but not exclusively, for hyperbolic PDE’s, suggests that the use of these awkward-to-handle penalty functions may not always be necessary. Computational experience of the author ([13]) also suggests that this is the case for certain problems, such as those being considered here.

In the next section we consider the MFE equations (2.10) and (2.11), or (2.12), in more detail. In particular, we investigate their steady solutions and compare them with steady solutions of the continuous equation (1.1).

### 3 Steady Solutions of the Moving Finite Element Equations

As mentioned in section 1, an important property of (1.1) is that with suitable boundary conditions on  $\partial\Omega$  (including the homogeneous Dirichlet conditions being considered here) it always possesses a unique steady solution,  $U(\underline{x})$  say. In this section we show that whenever the MFE equations (2.12) tend to a non-degenerate steady solution this is a best approximation to  $U(\underline{x})$  from the manifold of free knot linear splines on the mesh  $\underline{s}(t)$ , in a particular norm. In order to demonstrate this, theorem 3.4 shows that the stationary equations for a best approximation to  $U(\underline{x})$  are exactly the same as the equations  $\underline{g}(\underline{y}) = \underline{0}$ , with  $\underline{g}(\underline{y})$  as in (2.12). In fact the theorem states a

stronger result than this which enables the stability of the steady MFE solution to imply that the solution of the stationary equations is in fact a local minimum.

Before this theorem can be shown in detail however it is necessary to prove some preliminary lemmas. The first of these is used merely to help prove lemma 3.2 which is used in the proof of theorem 3.4. The third lemma is used in the corollary to this theorem.

**Lemma 3.1** *Consider a  $d$ -dimensional simplex with vertices at  $\hat{\underline{s}}_0, \hat{\underline{s}}_1, \dots, \hat{\underline{s}}_d$  and measure  $A(\hat{\underline{s}}) > 0$ . Let  $\hat{\alpha}_J$  be the local linear basis function on this simplex such that*

$$\hat{\alpha}_J(\hat{\underline{s}}_i) = \delta_{iJ} \quad \text{for } i, J \in \{0, 1, \dots, d\}.$$

Then

$$\frac{\partial A}{\partial \hat{s}_{Jm}} = \frac{\partial \hat{\alpha}_J}{\partial x_m} A \quad \text{for } J \in \{0, 1, \dots, d\} \text{ and } m \in \{1, \dots, d\}.$$

**Proof** First note that the  $d$ -dimensional measure,  $A$ , of the simplex depends only upon the positions of the vertices of the simplex and so is given by

$$A = A(\hat{\underline{s}}) = A(\hat{\underline{s}}_0, \dots, \hat{\underline{s}}_{J-1}, \hat{\underline{s}}_J, \hat{\underline{s}}_{J+1}, \dots, \hat{\underline{s}}_d).$$

(In fact it is possible to write this expression explicitly as

$$A(\hat{\underline{s}}) = \frac{1}{d!} \left| \prod_{i=0, i \neq J}^d (\hat{\underline{s}}_J - \hat{\underline{s}}_i) \right|$$

for any  $J \in \{0, 1, \dots, d\}$ , where  $\prod \times$  represents the  $d + 1$  dimensional vector product of  $d$  vectors in  $\mathbb{R}^d$  – however such a formula is not required in this proof.)

Now observe that the  $d$ -dimensional measure of the simplex obtained by replacing the vertex  $\hat{\underline{s}}_J$  by one at a point  $\underline{x}$  strictly inside the original simplex is  $A(\hat{\underline{s}}_0, \dots, \hat{\underline{s}}_{J-1}, \underline{x}, \hat{\underline{s}}_{J+1}, \dots, \hat{\underline{s}}_d)$ . Hence, since  $\hat{\alpha}_0, \hat{\alpha}_1, \dots, \hat{\alpha}_d$  are area coordinates, we know that

$$\hat{\alpha}_J(\underline{x}) = A(\hat{\underline{s}}_0, \dots, \hat{\underline{s}}_{J-1}, \underline{x}, \hat{\underline{s}}_{J+1}, \dots, \hat{\underline{s}}_d) / A(\hat{\underline{s}}_0, \dots, \hat{\underline{s}}_{J-1}, \hat{\underline{s}}_J, \hat{\underline{s}}_{J+1}, \dots, \hat{\underline{s}}_d).$$

Moreover, since we know that  $\hat{\alpha}_J(\underline{x})$  is affine, we know that  $\frac{\partial \hat{\alpha}_J}{\partial x_m}$  is independent of  $\underline{x}$ . Hence

$$\frac{\partial}{\partial x_m} A(\hat{\underline{s}}_0, \dots, \hat{\underline{s}}_{J-1}, \underline{x}, \hat{\underline{s}}_{J+1}, \dots, \hat{\underline{s}}_d)$$

is independent of  $\underline{x}$ . Thus

$$\frac{\partial}{\partial \hat{s}_{Jm}} A(\hat{\underline{s}}_0, \dots, \hat{\underline{s}}_{J-1}, \hat{\underline{s}}_J, \hat{\underline{s}}_{J+1}, \dots, \hat{\underline{s}}_d)$$

is independent of  $\hat{\underline{s}}_J$  and so

$$\frac{\partial}{\partial x_m} A(\hat{\underline{s}}_0, \dots, \hat{\underline{s}}_{J-1}, \underline{x}, \hat{\underline{s}}_{J+1}, \dots, \hat{\underline{s}}_d) = \frac{\partial}{\partial \hat{s}_{Jm}} A(\hat{\underline{s}}_0, \dots, \hat{\underline{s}}_{J-1}, \hat{\underline{s}}_J, \hat{\underline{s}}_{J+1}, \dots, \hat{\underline{s}}_d).$$

Therefore

$$\frac{\partial \hat{\alpha}_J}{\partial x_m} = \frac{\partial}{\partial \hat{s}_{Jm}} A(\hat{\underline{s}}_0, \dots, \hat{\underline{s}}_{J-1}, \hat{\underline{s}}_J, \hat{\underline{s}}_{J+1}, \dots, \hat{\underline{s}}_d) / A(\hat{\underline{s}}_0, \dots, \hat{\underline{s}}_{J-1}, \hat{\underline{s}}_J, \hat{\underline{s}}_{J+1}, \dots, \hat{\underline{s}}_d)$$

as required. ///

The statement and proof of the next lemma require a small amount of further notation to be established. For each internal knot point,  $\underline{s}_j$ , let  $N(j)$  be the number of elements which have a vertex at  $\underline{s}_j$ . Further, for  $e = 1, \dots, N(j)$ , let  $E(j, e)$  be a unique ordering of these  $N(j)$  elements with a vertex at  $\underline{s}_j$ . Finally, let  $\Omega_{E(j, e)}$  be the region occupied by the simplex numbered  $E(j, e)$  and let  $A_{E(j, e)}$  be the  $d$ -dimensional measure of this region.

**Lemma 3.2** Given  $p(\underline{x}) : \mathfrak{R}^d \rightarrow \mathfrak{R}$ ,  $j \in \{1, \dots, N\}$  and  $m \in \{1, \dots, d\}$  then

$$\frac{\partial}{\partial s_{jm}} \int_{\Omega_{E(j,e)}} p(\underline{x}) d\underline{x} = \int_{\Omega_{E(j,e)}} \frac{\partial}{\partial x_m} [p(\underline{x}) \alpha_j(\underline{x})] d\underline{x}$$

for each  $e \in \{1, \dots, N(j)\}$ .

**Proof** Consider element  $E(j, e)$  for an arbitrary  $e \in \{1, \dots, N(j)\}$ . Let the vertices of this element have positions  $\hat{s}_0, \dots, \hat{s}_d$  and let  $\hat{\alpha}_0, \dots, \hat{\alpha}_d$  be the local linear basis functions on this simplex. Also, let  $J$  be the local vertex of element  $E(j, e)$  which corresponds to node  $j$  (i.e.  $\hat{s}_J = \underline{s}_j$ ). It is now possible to make the following change of variables:

$$\underline{\xi}(\underline{x}) = \sum_{i=0}^d \underline{e}_i \hat{\alpha}_i(\underline{x}),$$

where  $\underline{e}_0 = \underline{0}$  and, for  $i = 1, \dots, d$ ,  $\underline{e}_i$  is the  $d$ -dimensional vector whose entries are all 0 except for the  $i^{\text{th}}$  which is 1. Note that the inverse of this mapping is given by

$$\underline{x}(\underline{\xi}) = \sum_{i=0}^d \hat{s}_i \tilde{\alpha}_i(\underline{\xi}), \quad \text{where} \quad \tilde{\alpha}_i(\underline{e}_k) = \delta_{ik}. \quad (3.1)$$

Now, if we let  $\Delta$  be the simplex in  $\underline{\xi}$ -space with vertices at  $\underline{e}_0, \underline{e}_1, \dots, \underline{e}_d$ , then

$$\int_{\Omega_{E(j,e)}} p(\underline{x}) d\underline{x} = \int_{\Delta} p(\underline{x}(\underline{\xi})) \left| \frac{\partial \underline{x}}{\partial \underline{\xi}} \right| d\underline{\xi}.$$

Hence

$$\frac{\partial}{\partial s_{jm}} \int_{\Omega_{E(j,e)}} p(\underline{x}) d\underline{x} = \frac{\partial}{\partial \hat{s}_{Jm}} \int_{\Omega_{E(j,e)}} p(\underline{x}) d\underline{x} = \frac{\partial}{\partial \hat{s}_{Jm}} \int_{\Delta} d! A_{E(j,e)} p(\underline{x}(\underline{\xi})) d\underline{\xi},$$

and applying lemma 3.1 in order to differentiate the last of these integrands with respect to  $\hat{s}_{Jm}$ , we get

$$\frac{\partial}{\partial s_{jm}} \int_{\Omega_{E(j,e)}} p(\underline{x}) d\underline{x} = \int_{\Delta} d! \left\{ A_{E(j,e)} \sum_{\ell=1}^d \frac{\partial p}{\partial x_\ell} \frac{\partial x_\ell}{\partial \hat{s}_{Jm}} + p(\underline{x}(\underline{\xi})) A_{E(j,e)} \frac{\partial \hat{\alpha}_J}{\partial x_m} \right\} d\underline{\xi}.$$

Given (3.1), this implies that

$$\frac{\partial}{\partial s_{jm}} \int_{\Omega_{E(j,e)}} p(\underline{x}) d\underline{x} = \int_{\Delta} d! A_{E(j,e)} \left\{ \frac{\partial p}{\partial x_m}(\underline{x}(\underline{\xi})) \tilde{\alpha}_J(\underline{\xi}) + p(\underline{x}(\underline{\xi})) \frac{\partial \hat{\alpha}_J}{\partial x_m}(\underline{x}(\underline{\xi})) \right\} d\underline{\xi}$$

which, on changing the variables of integration back to  $\underline{x}$ , gives

$$\begin{aligned} \frac{\partial}{\partial s_{jm}} \int_{\Omega_{E(j,e)}} p(\underline{x}) d\underline{x} &= \int_{\Omega_{E(j,e)}} \left\{ \frac{\partial p}{\partial x_m}(\underline{x}) \hat{\alpha}_J(\underline{x}) + p(\underline{x}) \frac{\partial \hat{\alpha}_J}{\partial x_m}(\underline{x}) \right\} d\underline{x} \\ &= \int_{\Omega_{E(j,e)}} \left\{ \frac{\partial p}{\partial x_m}(\underline{x}) \alpha_j(\underline{x}) + p(\underline{x}) \frac{\partial \alpha_j}{\partial x_m}(\underline{x}) \right\} d\underline{x}, \end{aligned}$$

as required. ///

The final lemma that we state here is useful in the proof that a stable steady MFE solution corresponds to a best local free knot linear spline approximation to the true steady solution of (1.1). Such a result is presented as a corollary to theorem 3.4 below.

**Lemma 3.3** *Let  $A$  be a symmetric positive definite  $m \times m$  matrix and  $B$  be a  $m \times m$  matrix whose eigenvalues all have positive real parts. Then if  $AB$  is symmetric it must also be positive definite.*

**Proof** Since  $AB$  is symmetric  $C = A^{-\frac{1}{2}}ABA^{-\frac{1}{2}} = A^{\frac{1}{2}}BA^{-\frac{1}{2}}$  is as well, where we use the symmetric square root. But the spectrum of  $A^{\frac{1}{2}}BA^{-\frac{1}{2}}$  is the same as that of  $B$ . Thus  $C$  is positive definite and therefore so is  $AB = A^{\frac{1}{2}}CA^{\frac{1}{2}}$ . ///

We are now in a position to prove the main result of this section. In the following theorem we consider minimizing the difference,  $\eta(\underline{x})$  say, between the true steady solution of (1.1),  $U(\underline{x})$ , and the best possible piecewise linear approximation to  $U(\underline{x})$  from all valid meshes  $\underline{s}$ . The norm with respect to which this minimization is performed is defined by

$$\|\eta(\underline{x})\|^2 = \int_{\Omega} \left\{ \frac{\partial \eta}{\partial x_{\mu}} p_{\mu\nu} \frac{\partial \eta}{\partial x_{\nu}} + q\eta^2 \right\} d\underline{x}, \quad (3.2)$$

and in this norm it is shown that any non-degenerate, steady, stable solution of the MFE equations (2.12) must also be a local minimizer of  $\|\eta\|$ .

**Theorem 3.4** *Let  $U(\underline{x})$  be the unique steady solution of (1.1) subject to homogeneous Dirichlet boundary conditions on  $\partial\Omega$ . Let*

$$v(\underline{x}, t) = \sum_{i=1}^N a_i(t) \alpha_i(\underline{x}, \underline{s}(t))$$

be a continuous piecewise linear approximation to  $U(\underline{x})$  on a mesh  $\underline{s}(t)$  with  $N$  free internal knots  $\underline{s}_1(t), \dots, \underline{s}_N(t)$ . Also let

$$\underline{y} = (a_1, s_{11}, \dots, s_{1d}; \dots; a_N, s_{N1}, \dots, s_{Nd})^T$$

and

$$I(\underline{y}) = \int_{\Omega} \left( \frac{\partial}{\partial x_{\mu}} (U - v) p_{\mu\nu} \frac{\partial}{\partial x_{\nu}} (U - v) + q[U - v]^2 \right) d\underline{x}.$$

Then

$$\nabla I(\underline{y}) = -2\underline{g}(\underline{y}), \quad (3.3)$$

for  $\underline{g}(\underline{y})$  as in (2.12).

**Proof** We begin by observing, from (2.10) and (2.11), that  $\underline{g}(\underline{y})$  consists of the following components:

$$- \left\langle p_{\mu\nu} \frac{\partial v}{\partial x_{\nu}}, \frac{\partial \alpha_j}{\partial x_{\mu}} \right\rangle - \langle qv - r, \alpha_j \rangle \quad (3.4)$$

and

$$-\frac{1}{2} \left\langle \frac{\partial v}{\partial x_{\mu}} \frac{\partial v}{\partial x_{\nu}}, \frac{\partial}{\partial x_m} (\alpha_j p_{\mu\nu}) \right\rangle + \left\langle \frac{\partial v}{\partial x_m} \frac{\partial v}{\partial x_{\nu}}, p_{\mu\nu} \frac{\partial \alpha_j}{\partial x_{\mu}} \right\rangle - \left\langle qv - r, \frac{\partial v}{\partial s_{jm}} \right\rangle \quad (3.5)$$

for  $m = 1, \dots, d$  and  $j = 1, \dots, N$ . We will now show that the components of  $\nabla I(\underline{y})$  are as claimed in (3.3) by demonstrating that  $\frac{\partial I}{\partial a_j}$  is  $-2$  times (3.4) and  $\frac{\partial I}{\partial s_{jm}}$  is  $-2$  times (3.5) for  $m = 1, \dots, d$  and  $j = 1, \dots, N$ .

For the first of these two cases,

$$\begin{aligned} \frac{\partial I}{\partial a_j} &= - \int_{\Omega} \left\{ \frac{\partial}{\partial x_{\mu}} (U - v) p_{\mu\nu} \frac{\partial \alpha_j}{\partial x_{\nu}} + \frac{\partial \alpha_j}{\partial x_{\mu}} p_{\mu\nu} \frac{\partial}{\partial x_{\nu}} (U - v) + 2q(U - v)\alpha_j \right\} d\underline{x} \\ &= 2 \int_{\Omega} \frac{\partial \alpha_j}{\partial x_{\mu}} p_{\mu\nu} \frac{\partial v}{\partial x_{\nu}} d\underline{x} + 2 \int_{\Omega} qv\alpha_j d\underline{x} - 2 \int_{\Omega} \left\{ \frac{\partial \alpha_j}{\partial x_{\mu}} p_{\mu\nu} \frac{\partial U}{\partial x_{\nu}} + qU\alpha_j \right\} d\underline{x} \end{aligned}$$

$$\begin{aligned}
& \text{(using the symmetry } p_{\mu\nu} = p_{\nu\mu}\text{)} \\
& = 2 \int_{\Omega} p_{\mu\nu} \frac{\partial v}{\partial x_{\nu}} \frac{\partial \alpha_j}{\partial x_{\mu}} d\mathbf{x} + 2 \int_{\Omega} qv\alpha_j d\mathbf{x} - 2 \int_{\Omega} \left\{ -\frac{\partial}{\partial x_{\mu}} \left( p_{\mu\nu} \frac{\partial U}{\partial x_{\nu}} \right) + qU \right\} \alpha_j d\mathbf{x} \\
& = 2 \int_{\Omega} p_{\mu\nu} \frac{\partial v}{\partial x_{\nu}} \frac{\partial \alpha_j}{\partial x_{\mu}} d\mathbf{x} + 2 \int_{\Omega} qv\alpha_j d\mathbf{x} - 2 \int_{\Omega} r\alpha_j d\mathbf{x}, \tag{3.6}
\end{aligned}$$

which is equal to  $-2$  times (3.4) as required.

For the other case,

$$\begin{aligned}
\frac{\partial I}{\partial s_{jm}} & = \frac{\partial}{\partial s_{jm}} \int_{\Omega} \frac{\partial v}{\partial x_{\mu}} p_{\mu\nu} \frac{\partial v}{\partial x_{\nu}} d\mathbf{x} - \frac{\partial}{\partial s_{jm}} \int_{\Omega} \frac{\partial U}{\partial x_{\mu}} p_{\mu\nu} \frac{\partial v}{\partial x_{\nu}} d\mathbf{x} - \frac{\partial}{\partial s_{jm}} \int_{\Omega} \frac{\partial v}{\partial x_{\mu}} p_{\mu\nu} \frac{\partial U}{\partial x_{\nu}} d\mathbf{x} \\
& \quad - 2 \frac{\partial}{\partial s_{jm}} \int_{\Omega} qUv d\mathbf{x} + \frac{\partial}{\partial s_{jm}} \int_{\Omega} qv^2 d\mathbf{x} \\
& = \sum_{e=1}^{N(j)} \frac{\partial}{\partial s_{jm}} \int_{\Omega_{E(j,e)}} \frac{\partial v}{\partial x_{\mu}} p_{\mu\nu} \frac{\partial v}{\partial x_{\nu}} d\mathbf{x} + \frac{\partial}{\partial s_{jm}} \int_{\Omega} qv^2 d\mathbf{x} - 2 \frac{\partial}{\partial s_{jm}} \int_{\Omega} \left\{ \frac{\partial v}{\partial x_{\mu}} p_{\mu\nu} \frac{\partial U}{\partial x_{\nu}} + qUv \right\} d\mathbf{x} \\
& \text{(using the symmetry } p_{\mu\nu} = p_{\nu\mu}\text{)} \\
& = \sum_{e=1}^{N(j)} \left\{ G_{\mu} G_{\nu} \frac{\partial}{\partial s_{jm}} \int_{\Omega_{E(j,e)}} p_{\mu\nu} d\mathbf{x} + \frac{\partial}{\partial s_{jm}} (G_{\mu} G_{\nu}) \int_{\Omega_{E(j,e)}} p_{\mu\nu} d\mathbf{x} \right\} + 2 \int_{\Omega} qv \frac{\partial v}{\partial s_{jm}} d\mathbf{x} \\
& \quad + 2 \frac{\partial}{\partial s_{jm}} \int_{\Omega} \left\{ \frac{\partial}{\partial x_{\mu}} \left( p_{\mu\nu} \frac{\partial U}{\partial x_{\nu}} \right) - qU \right\} v d\mathbf{x},
\end{aligned}$$

where  $\underline{G}_{\mu}$  ( $\underline{G}_{\nu}$ ) is equal to  $\frac{\partial v}{\partial x_{\mu}}$  ( $\frac{\partial v}{\partial x_{\nu}}$ ) restricted to simplex  $E(j, e)$ . (Note that the values of  $\underline{G}_{\mu}$  and  $\underline{G}_{\nu}$  are independent of  $\mathbf{x}$  but dependent upon  $\underline{s}_j$ .) Hence, using lemma 3.2,

$$\begin{aligned}
\frac{\partial I}{\partial s_{jm}} & = \sum_{e=1}^{N(j)} \left\{ G_{\mu} G_{\nu} \int_{\Omega_{E(j,e)}} \frac{\partial}{\partial x_m} (p_{\mu\nu} \alpha_j) d\mathbf{x} + \left[ G_{\mu} \frac{\partial}{\partial x_{\nu}} \left( \frac{\partial v}{\partial s_{jm}} \right) + \frac{\partial}{\partial x_{\mu}} \left( \frac{\partial v}{\partial s_{jm}} \right) G_{\nu} \right] \int_{\Omega_{E(j,e)}} p_{\mu\nu} d\mathbf{x} \right\} \\
& \quad + 2 \int_{\Omega} (qv - r) \frac{\partial v}{\partial s_{jm}} d\mathbf{x} \\
& = \sum_{e=1}^{N(j)} \left\{ \int_{\Omega_{E(j,e)}} \frac{\partial v}{\partial x_{\mu}} \frac{\partial v}{\partial x_{\nu}} \frac{\partial}{\partial x_m} (p_{\mu\nu} \alpha_j) d\mathbf{x} - \int_{\Omega_{E(j,e)}} \left( \frac{\partial v}{\partial x_{\mu}} \frac{\partial \alpha_j}{\partial x_{\nu}} \frac{\partial v}{\partial x_m} + \frac{\partial \alpha_j}{\partial x_{\mu}} \frac{\partial v}{\partial x_m} \frac{\partial v}{\partial x_{\nu}} \right) p_{\mu\nu} d\mathbf{x} \right\} \\
& \quad + 2 \int_{\Omega} (qv - r) \frac{\partial v}{\partial s_{jm}} d\mathbf{x} \quad \text{(using (2.5))} \\
& = \int_{\Omega} \frac{\partial v}{\partial x_{\mu}} \frac{\partial v}{\partial x_{\nu}} \frac{\partial}{\partial x_m} (p_{\mu\nu} \alpha_j) d\mathbf{x} - 2 \int_{\Omega} \frac{\partial v}{\partial x_m} \frac{\partial v}{\partial x_{\nu}} p_{\mu\nu} \frac{\partial \alpha_j}{\partial x_{\mu}} + 2 \int_{\Omega} (qv - r) \frac{\partial v}{\partial s_{jm}} d\mathbf{x} \tag{3.7} \\
& \text{(again using the symmetry } p_{\mu\nu} = p_{\nu\mu}\text{),}
\end{aligned}$$

which is equal to  $-2$  times (3.5) as required. ///

This theorem tells us that any steady MFE solution (i.e. a solution for which  $\underline{\dot{y}} = \underline{0}$  and so  $\underline{g}(\underline{y}) = \underline{0}$  by (2.12)) is such that  $\underline{\nabla}I(\underline{y}) = \underline{0}$  too. Hence, such a solution satisfies the stationary equations for a best free knot linear spline approximation to  $U(\mathbf{x})$  in the norm (3.2). The following corollary goes on to show that provided the steady MFE solution is non-degenerate and stable, it corresponds to a local minimum of the error  $I(\underline{y})$ .

**Corollary 3.5** *Any non-degenerate, asymptotically stable, steady solution of the Moving Finite Element equations (2.12) is a locally best approximation to the true steady solution of (1.1) in the norm (3.2).*

**Proof** Let  $\underline{y}_0$  be such a solution of (2.12). Then, because it is non-degenerate,  $A(\underline{y}_0)$  is positive definite. If we now consider a small perturbation of  $\underline{y}_0$ , given by  $\underline{y} = \underline{y}_0 + \epsilon \underline{y}_1$ , (2.12) becomes

$$A(\underline{y}_0 + \epsilon \underline{y}_1)(\dot{\underline{y}}_0 + \epsilon \dot{\underline{y}}_1) = \underline{g}(\underline{y}_0 + \epsilon \underline{y}_1).$$

Linearizing this about  $\underline{y}_0$  gives

$$\dot{\underline{y}}_1 = A^{-1}(\underline{y}_0) D\underline{g}(\underline{y}_0) \underline{y}_1,$$

where  $D\underline{g}(\underline{y}_0)$  is the Jacobian of  $\underline{g}$  with respect to  $\underline{y}$  evaluated at  $\underline{y}_0$ . Hence the asymptotic stability of  $\underline{y}_0$  implies that all eigenvalues of  $A^{-1}(\underline{y}_0) D\underline{g}(\underline{y}_0)$  have negative real parts. Moreover, since  $D\underline{g}$  is equal to the Hessian of  $I$  (from (3.3)), it must be symmetric and so we can apply lemma 3.3 with  $A = A(\underline{y}_0)$  and  $B = -A^{-1}(\underline{y}_0) D\underline{g}(\underline{y}_0)$  to deduce that  $-D\underline{g}(\underline{y}_0)$  is positive definite. That is, the Hessian of  $I$  evaluated at  $\underline{y}_0$  is positive definite (again by (3.3)), and so  $\underline{y}_0$  must be a local minimum of  $I(\underline{y})$  as claimed. ///

We have now proved the result claimed: that if the MFE solution tends to a steady state, this is a locally best approximation to the true steady solution. We may also observe that the globally best approximation to the true steady solution in this norm is of course a local minimum too, and so by the converse of the argument in the above proof it follows that this must be a stable steady solution of the MFE equations. Of course this tells us nothing about the domain of attraction of the global minimizer and we have no guarantee that for an arbitrary choice of initial data the MFE equations will tend to this, or any other, steady solution.

In the next section we give a number of numerical examples which serve to confirm the above analysis. One of the examples also confirms the fact that for arbitrary initial data the equations may not always tend to a steady solution due to a breakdown occurring because the measure of one or more of the simplexes becomes non-positive. This difficulty is easily overcome in this case by simply removing appropriate nodes and elements from the mesh – an approach which is studied in much more detail in [17]. From a theoretical point of view this approach of deleting, and possibly adding, nodes and elements presents few problems since the above results all still hold: applying to whatever mesh topology is in use when a steady solution is finally reached. (From an algorithmic point of view this approach may not always be as straightforward of course: again see [17].)

Additional generalizations to the theory can be made by considering a wider class of boundary conditions, as in [14], or by adding constraints to the motion of the nodes. For example, if we choose to impose zero velocity constraints on nodes in a certain region of the domain (to prevent there being too few nodes in a particularly flat region of the solution perhaps) then it is easy to see that the proofs of theorem 3.4 and its corollary still go through, except that now we have a best approximation from a smaller manifold: with only the knots not constrained in the MFE solution being free. Finally it is also possible to allow the constrained motion of some of the nodes on the boundary of the domain, as described in [14].

## 4 Some Numerical Examples

In this section we outline a small number of computational examples which serve to confirm the results of section 3. For simplicity we only consider the use of Dirichlet boundary conditions and we keep the mesh fixed throughout the boundary of the domain. A number of one-dimensional examples appear in [13] so all of the examples in this section are for two-dimensional problems ( $d = 2$ ). Whenever steady MFE solutions are found we confirm that they are local minimizers of the free knot error in the norm (3.2) by using the NAG library routine E04JAF [25].

We start by considering the case where  $p_{\mu\nu} \equiv \delta_{\mu\nu}$  (the usual Kronecker delta) and  $q \equiv 0$  in equation (1.1),

$$\frac{\partial u}{\partial t}(\underline{x}, t) = \frac{\partial^2 u}{\partial x_1^2}(\underline{x}, t) + \frac{\partial^2 u}{\partial x_2^2}(\underline{x}, t) + r(\underline{x}), \quad (4.1)$$

subject to homogeneous Dirichlet boundary conditions on a square spatial domain:  $(0, 1) \times (0, 1)$ . We look at two different choices for  $r(\underline{x})$  – corresponding to the two steady solutions

(i)  $U(\underline{x}) = 64x_1^2(1 - x_1)x_2^2(1 - x_2)$ ,

(ii)  $U(\underline{x}) = \sin(\pi x_1^5) \sin(\pi x_2^5)$ .

For this equation the norm (3.2) simplifies to

$$\|\eta\|^2 = \int_{\Omega} [\nabla\eta]^2 d\underline{x}, \quad (4.2)$$

which is a genuine norm here because the Dirichlet boundary conditions ensure that the error is zero along  $\partial\Omega$ .

Figure 1 shows the evolution of the MFE solution mesh when solving problem (i) with just 15 degrees of freedom. The meshes shown (from top left to bottom right) are the initial mesh, the mesh at  $t = 0.01$ , the mesh at  $t = 0.02$  and the steady solution mesh (at  $t = 1.0$ ). A quadrature rule with degree of precision 5 (see [5]) was used to calculate all of the integrals in (2.10) and (2.11) on each triangular element: because of the choice of  $r(\underline{x})$  this ensured that all integration was exact. The steady solution to this problem was found to be as given in table 1 and it is straightforward to verify that this is indeed a local minimum of

$$\left\| \sum_{i=1}^5 a_i \alpha_i(\underline{x}, \underline{s}) - 64x_1^2(1 - x_1)x_2^2(1 - x_2) \right\|$$

over all choices of  $(a_1, \underline{s}_1, \dots, a_5, \underline{s}_5)$ , where  $\|\cdot\|$  is given by (4.2).

The MFE solution to problem (ii) behaves in a similar manner to that of problem (i), as is illustrated by figure 2. Again only 15 degrees of freedom have been used along with identical initial data, and the meshes shown are the initial mesh and those calculated at  $t = 0.01$ ,  $t = 0.02$  and at the steady state ( $t = 1.0$ ). Table 2 gives the precise values that the 15 degrees of freedom attain when the steady MFE solution is reached. When one attempts to verify that these values represent an optimal approximation to the true steady solution,  $U(\underline{x})$ , a slight discrepancy is observed however. Table 3 shows the optimum values of  $(a_1, \underline{s}_1, \dots, a_5, \underline{s}_5)$  as calculated by E04JAF using a very accurate adaptive quadrature subroutine for all integration. The difference, which is far too small to be discernible on a picture of the meshes, as in figure 2 for example, is due to the inexact quadrature that is used in the MFE code.

In all of the theory presented in section 3 it is assumed that exact integration is used to calculate  $\underline{g}(\underline{y})$  whereas this is not the case in practice. Hence for the remaining examples in this section we regard the results of section 3 as being confirmed whenever the MFE solution and the exact optimum are very close. (In problem (ii) above, for example, the largest difference in any of the degrees of freedom is just 1.0% and the difference between the error using MFE with quadrature and the exact minimum is less than 0.01%.)

The next example that we consider, problem (iii) say, is exactly the same as problem (ii) except uses different initial data on a less coarse mesh: with 66 degrees of freedom and 54 triangular elements. The initial data has been obtained by solving equation (4.1) using an adaptive h-refinement code of the sort described in [12] or [18], and so it gives a reasonably good approximation to the exact steady solution. As can be seen from figure 3 (again the sequence of pictures goes from the

top left to the bottom right) the initial mesh soon starts to deform significantly and by the time  $t = 0.00125$  (the second mesh), 3 node points near the very centre of the mesh are about to run into each other, causing 4 elements to shrink to zero. In order to overcome this potential difficulty it is necessary to delete the offending elements and merge the 3 nodes into one, thus reducing the number of degrees of freedom to 60 and the number of triangular elements to 50 (the third mesh in figure 3). The integration in time can now continue until a steady state is reached (illustrated by the final mesh at  $t = 1.0$ ). Note that in this case the merging of the nodes provides few problems since they come together in a continuous manner. That is, as the nodes get closer together so do their solution values and so replacing the three converged nodes with a single one is very straightforward. A more extensive look at MFE algorithms with node addition and deletion may be found in [17].

As with the previous examples it is possible to use the NAG routine E04JAF to verify that the steady MFE solution is indeed a local optimum over the manifold of free knot linear splines. In this case the optimum is with respect to the final family of meshes, with 20 free knots, rather than the original family, containing 22 free knots. (The free knot at the top right corner of the final mesh appears at first sight to be surprisingly close to the boundary of the domain however numerical experiment confirms that its position is indeed optimal.) In addition it is possible to compare the final approximation to the steady solution, which has an error of 0.6206 in the norm (4.2), with the original approximation, obtained using adaptive h-refinement, which has an error of 0.8200 in this norm. We see a significant improvement using the Moving Finite Element method even though the final approximation uses fewer degrees of freedom.

The final computations that we describe in this section are for the equation

$$\frac{\partial u}{\partial t}(\underline{x}, t) = \Delta u(\underline{x}, t) - u(\underline{x}, t) + r(\underline{x}), \quad (4.3)$$

on the same spatial domain as before:  $(0, 1) \times (0, 1)$ . This corresponds to choosing  $p_{\mu\nu} \equiv \delta_{\mu\nu}$  and  $q \equiv 1$  in (1.1) and so the norm (3.2) becomes the usual  $H^1$  norm on  $\Omega$ . Again  $r(\underline{x})$  may be selected so that the steady solution to the problem is  $U(\underline{x}) = \sin(\pi x_1^5) \sin(\pi x_2^5)$ . When the initial data is chosen as in problems (ii) and (iii) the solution evolves in a very similar manner to the solutions to these problems. In each case the steady MFE solutions are also very similar which is to be expected since they are optimal approximations to the same function in closely related norms.

Figure 4 shows the evolution of the solution mesh for example (iv) which solves equation (4.3) using a third choice of initial data. Again, the meshes shown are at times  $t = 0.0$ ,  $t = 0.01$ ,  $t = 0.02$  and  $t = 1.0$  (the steady state), and on this occasion the problem uses 27 degrees of freedom and 32 elements. Once more it is possible to verify that the steady MFE solution is indeed optimal (allowing for small errors in the numerical quadrature), this time in the  $H^1$  norm on  $\Omega$ .

## 5 Discussion

The results of this paper, whilst applying to problems in an arbitrary number of spatial dimensions with straightforward extensions to a wide variety of boundary conditions, do have a number of restrictions. In particular, we have only considered one specific family of linear equations, (1.1), and no mention is made of the temporal accuracy of the MFE method for these or any other problems: only steady solutions have been considered. Clearly these restrictions are very important since most practical time-dependent problems that we may wish to solve numerically are more complicated than (1.1), containing convection terms or nonlinearities for example. Also we are often interested in the temporal as well as the steady solutions of such problems. Nevertheless, the results of theorem 3.4 and 3.5 are still of some significance. These results demonstrate that there is some potential advantage to be gained by allowing the spatial mesh to deform continuously rather than simply using a fixed finite element mesh or just adding and deleting nodes at discrete times. Practical

adaptive algorithms may combine a number of more complicated features, such as the use of penalty functions or the systematic creation and deletion of elements and nodes, however it is important to try to understand the underlying mechanisms which drive the node motion itself.

Example (iii) in section 4, for example, demonstrates that h-refinement alone will not always be as accurate as a combination of both h-refinement and node movement, and there is no reason to suspect that such a result is not also true for problems other than (1.1). For this reason it seems plausible that further research, applying the MFE method to nonlinear equations for example, is likely to lead to an extension of the results described here. In addition, it would be helpful to understand the precise effects of using numerical quadrature in the assembly of the right-hand-side of equations (2.12).

## Acknowledgements

I would like to thank Keith Miller for his detailed comments on this manuscript which have been of great help. I would also like to thank the referee who provided me with the proof of lemma 3.3, which is a considerable improvement over my original version, and Mike Baines for a number of interesting and useful discussions.

## References

- [1] M J Baines (1985). *Locally Adaptive Moving Finite Elements*. Numerical Methods for Fluid Dynamics II (eds. K W Morton & M J Baines), Oxford University Press.
- [2] M J Baines (1994). *Moving Finite Elements*. OUP.
- [3] M J Baines and A J Wathen (1986). *Moving Finite Element Modelling of Compressible Flow*. Applied Num. Maths., 2, 495–514.
- [4] M J Baines and A J Wathen (1988). *Moving Finite Element Methods for Evolutionary Problems. I. Theory*. J. of Comp. Phys., 79, 245–269.
- [5] G R Cowper (1973) *Gaussian Quadrature Formulas for Triangles*. Int. J. Num. Meth. Eng., 7, 405–410.
- [6] M Delfour, G Payre and J-P Zolésio (1985). *An Optimal Triangulation for Second-Order Elliptic Problems*. Comp. Meth. Appl. Mech. Eng., 50, 231–261.
- [7] J D P Donnelly (1990). *Approximation of the Diffusion Term in the Method of Moving Finite Elements*. Oxford University Computing Laboratory Report 90/9.
- [8] T Dupont (1982) *Mesh Modification for evolution Equations*. Math. Comp., 39, 85–107.
- [9] R J Gelinias, S K Doss and K Miller (1981). *The Moving Finite Element Method: Application to General Partial Differential Equations with Multiple Large Gradients*. J. of Comp. Phys., 40, 202–249.
- [10] D Gilbarg and N S Trudinger (1983) *Elliptic Partial Differential Equations of Second Order*. Grundlehren der mathematischen Wissenschaften 224, Springer-Verlag.
- [11] A Harten and J M Hyman (1983) *Self-Adjusting Grid Methods for One-Dimensional Hyperbolic Conservation Laws*. J. of Comp. Phys., 50, 235–269.

- [12] P K Jimack (1992) *A New Method of Spatial error Control for the Finite Element Method on Convection Dominated Problems*. Advances in Computer Methods for Partial Differential Equations VII (eds. R Vichnevetsky, D Knight & G Richter), IMACS.
- [13] P K Jimack (1992) *On Steady and Large Time Solutions of the Semi-Discrete Moving Finite Element Equations for One-Dimensional Diffusion Equations*. IMA J. Num. Anal., 12, 545–564.
- [14] P K Jimack (1993) *A Best Approximation Property of the Moving Finite Element Method*. School of Computer Studies Research Report 93.35, University of Leeds.
- [15] P K Jimack and A J Wathen(1991) *Temporal Derivatives in the Finite Element Method on Continuously Deforming Grids*. SIAM J. Num. Anal., 28, 990–1003.
- [16] I W Johnson, A J Wathen and M J Baines (1988). *Moving Finite Element Methods for Evolutionary Problems. II. Applications*. J. of Comp. Phys., 79, 270–297.
- [17] A Kuprat (1992). *Creation and Annihilation of Nodes for the Moving Finite Element Method*. Ph.D. Thesis, University of California at Berkeley. Unpublished.
- [18] R Löhner (1987). *iAn Adaptive Finite Element Scheme for Transient Problems in CFD*. Comp. Meth. Appl. Mech. Eng., 61, 323–338.
- [19] B J Lucier (1986). *A Moving Mesh Numerical Method for Hyperbolic Conservation Laws*. Math. Comp., 46, 59–69.
- [20] D R Lynch (1982). *A Unified Approach to Simulation on Deforming Elements with Applications to Phase Change Problems*. J. of Comp. Phys., 47, 387–411.
- [21] K Miller and R Miller (1981). *Moving Finite Elements, Part I*. SIAM J. Num. Anal., 18, 1019–1032.
- [22] K Miller (1981). *Moving Finite Elements, Part II*. SIAM J. Num. Anal., 18, 1033–1057.
- [23] A C Mueller (1983) *Continuously Deforming Finite Element Methods for Transport Problems*. Ph.D. Thesis, University of Austin, Texas. Unpublished.
- [24] A C Mueller and G F Carey (1985). *Continuously Deforming Finite Elements*. Int. J. Num. Meth. Eng., 21, 2099–2126.
- [25] Numerical Algorithms Group Limited (1993) *The NAG Fortran Library Manual – Mark 14*.
- [26] G L Sidén and D R Lynch (1988). *Wave Equation Hydrodynamics on Deforming Elements*. Int. J. Num. Meth. Fluids., 8, 1071–1094.
- [27] A J Wathen (1986) *Mesh Independent Spectra in the Moving Finite Element Equations*. SIAM J. Num. Anal., 23, 797–814.
- [28] A J Wathen and M J Baines (1985). *On the Structure of the Moving Finite Element Equations*. IMA J. Num. Anal., 5, 161–182.

$i$	$a_i$ [Initial $a_i$ ]	$\underline{s}_i$ [Initial $\underline{s}_i$ ]
1	0.1441192496 [0.6]	$(0.2014560973, 0.2014560973)^T$ [0.4, 0.4]
2	0.1120860039 [0.5]	$(0.1317789179, 0.9481600932)^T$ [0.4, 0.9]
3	1.3146998340 [1.4]	$(0.6575676092, 0.6575676092)^T$ [0.7, 0.7]
4	0.1120860039 [0.5]	$(0.9481600932, 0.1317789179)^T$ [0.9, 0.4]
5	1.1199302110 [0.4]	$(0.8051060698, 0.8051060698)^T$ [0.9, 0.9]

Table 1: The values of the 15 degrees of freedom at the steady MFE solution to problem (i) [along with the initial data used].

$i$	$a_i$ [Initial $a_i$ ]	$\underline{s}_i$ [Initial $\underline{s}_i$ ]
1	0.07833177194 [0.6]	$(0.6143033663, 0.6143033663)^T$ [0.4, 0.4]
2	0.03827731410 [0.5]	$(0.4436057237, 0.9509544010)^T$ [0.4, 0.9]
3	0.8054015159 [1.4]	$(0.8654844853, 0.8654844853)^T$ [0.7, 0.7]
4	0.03827731410 [0.5]	$(0.9509544010, 0.4436057237)^T$ [0.9, 0.4]
5	0.7698683170 [0.4]	$(0.9128451098, 0.9128451098)^T$ [0.9, 0.9]

Table 2: The values of the 15 degrees of freedom at the steady MFE solution to problem (ii) [along with the initial data used].

$i$	$a_i$	$\underline{s}_i$
1	0.07755135662	$(0.6137001656, 0.6137000912)^T$
2	0.03792676572	$(0.4437757351, 0.9519518425)^T$
3	0.8008348341	$(0.8658795118, 0.8658795095)^T$
4	0.03792675588	$(0.9519518557, 0.4437757363)^T$
5	0.7632101319	$(0.9132093670, 0.9132093662)^T$

Table 3: The optimum values of the 15 degrees of freedom for problem (ii) as calculated by E04JAF using highly accurate quadrature.

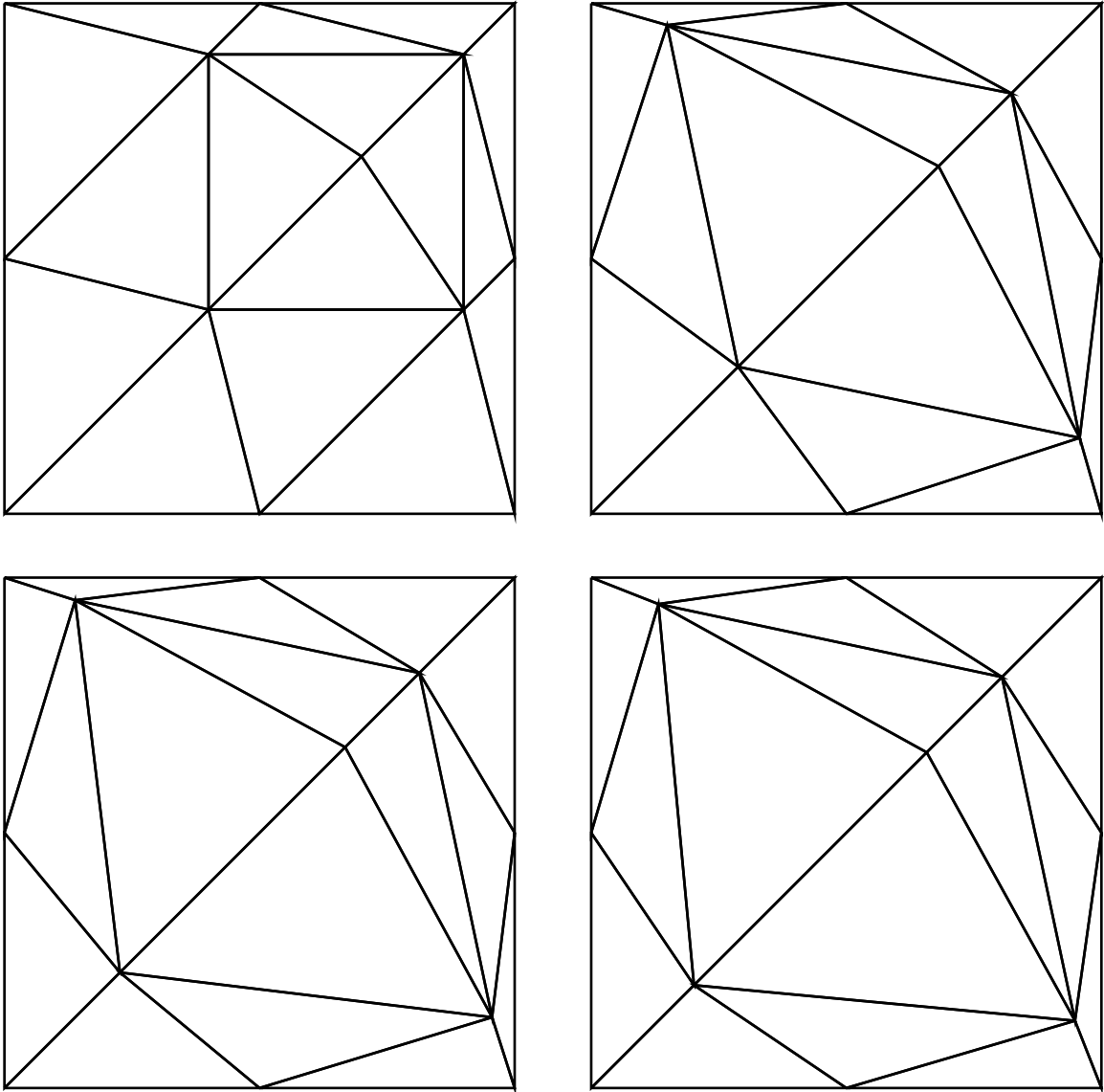


Figure 1: Evolution of the MFE solution mesh when solving problem (i).

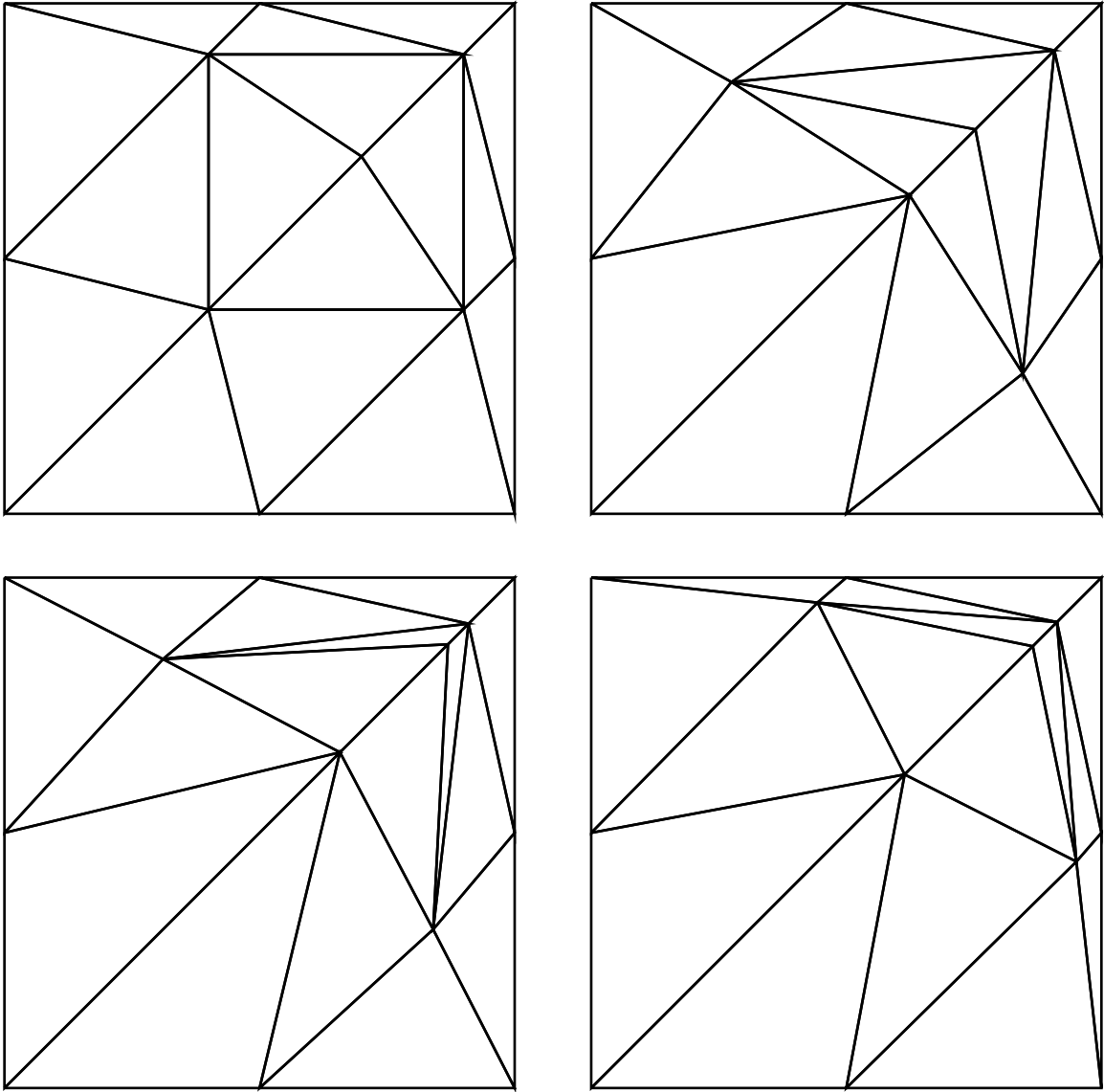


Figure 2: Evolution of the MFE solution mesh when solving problem (ii).

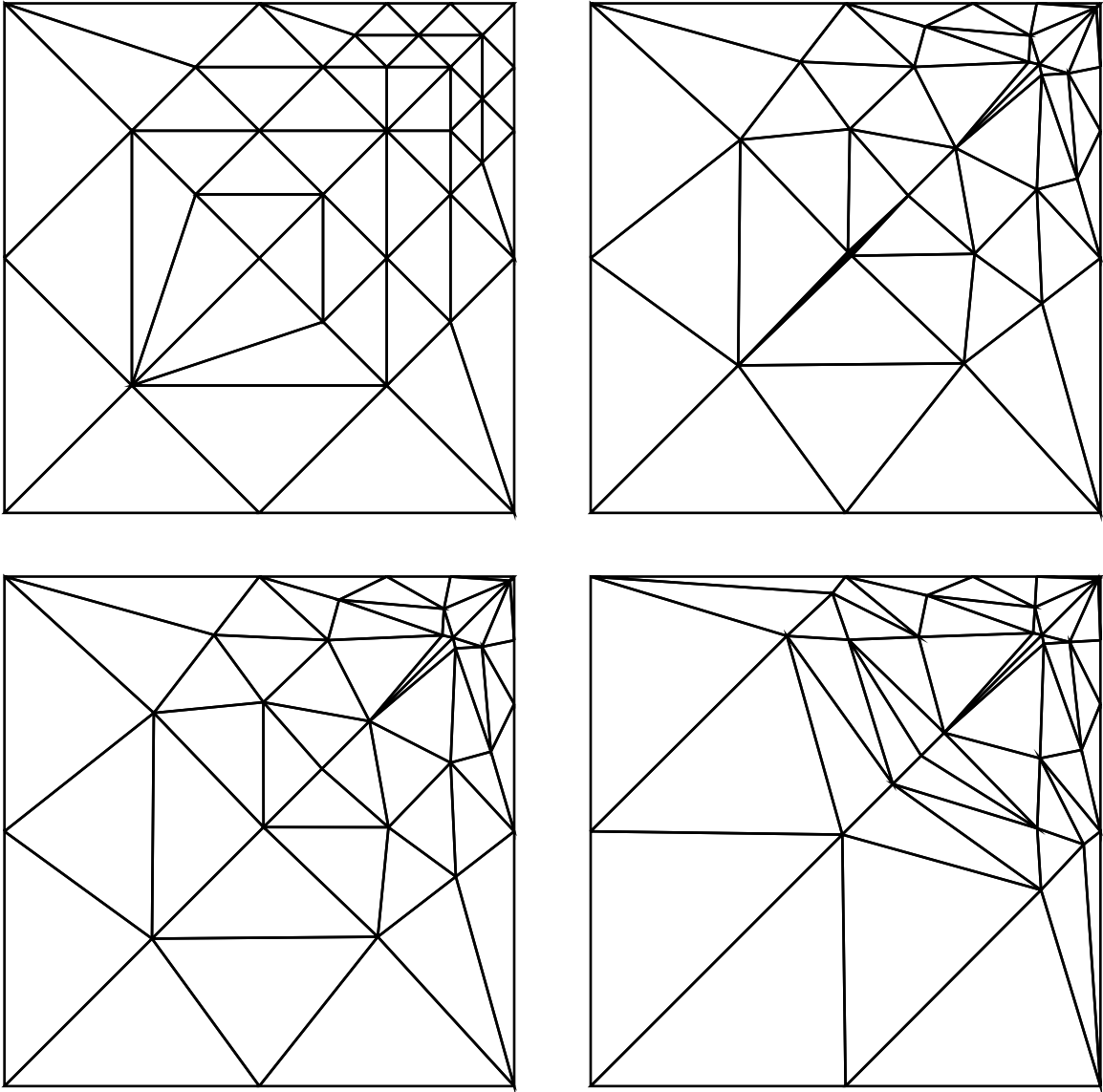


Figure 3: Evolution of the MFE solution mesh when solving problem (iii).

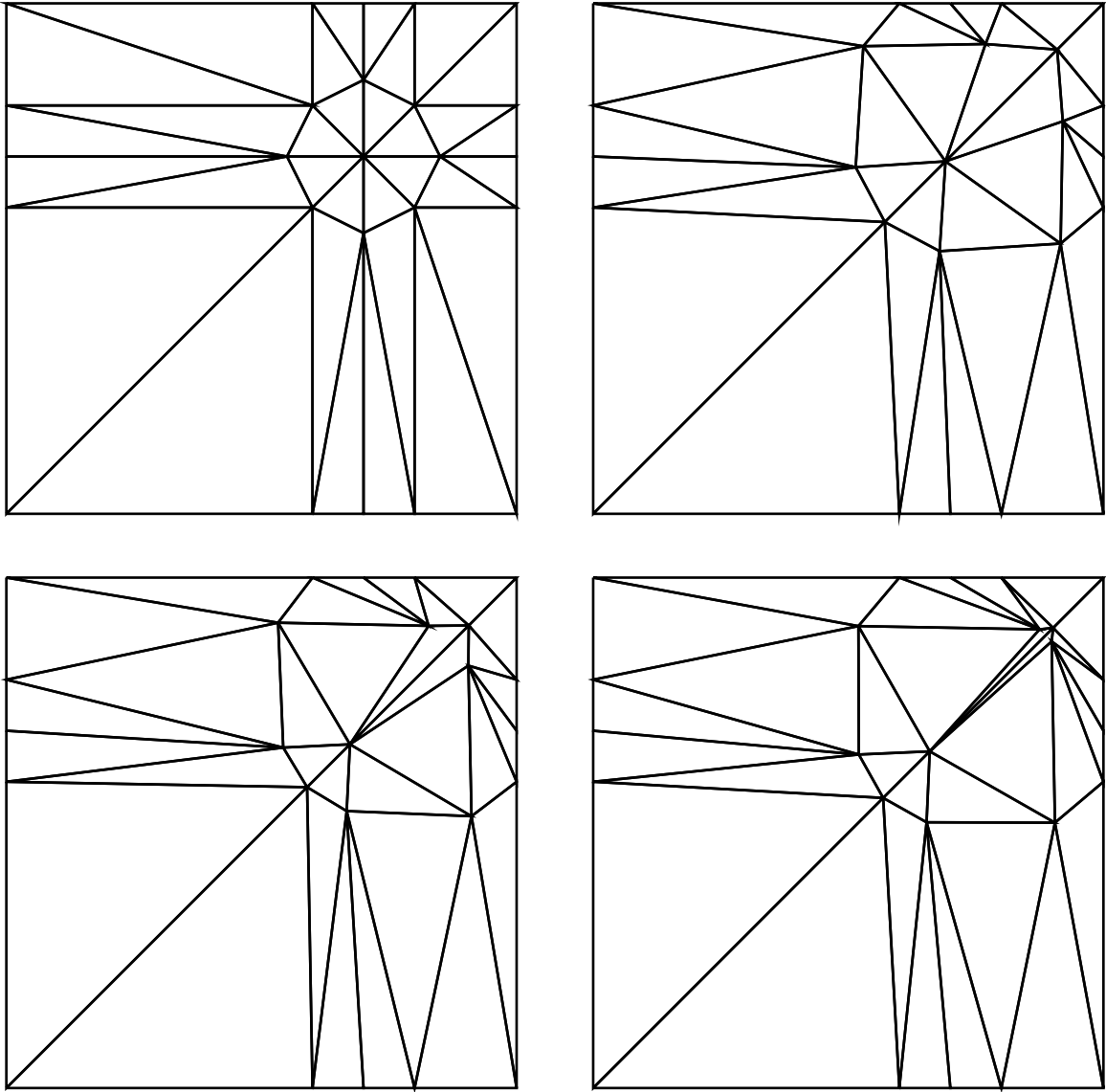


Figure 4: Evolution of the MFE solution mesh when solving problem (iv).