

EXPLICIT FINITE ELEMENT METHODS FOR LINEAR HYPERBOLIC SYSTEMS

RICHARD S. FALK¹ and GERARD R. RICHTER² *

¹ Department of Mathematics, Rutgers University, Piscataway, NJ 08854,
falk@math.rutgers.edu

² Department of Computer Science, Rutgers University, Piscataway, NJ 08854,
richter@cs.rutgers.edu

Abstract. Our focus is on explicit finite element discretization of transient, linear hyperbolic systems in arbitrarily many space dimensions. We propose several ways of generating suitable “explicit” meshes, and sketch an $O(h^{n+1/2})$ error estimate for a discontinuous Galerkin method. Continuous methods are also considered briefly. This paper parallels [2] in large part, while using a different approach in the analysis.

1 Introduction

The problem of interest to us here is a linear, symmetric hyperbolic system

$$\mathcal{L}\mathbf{u} \equiv \frac{\partial \mathbf{u}}{\partial t} + \sum_{i=1}^N A_i \frac{\partial \mathbf{u}}{\partial x_i} + B\mathbf{u} = \mathbf{f}, \quad (\mathbf{x}, t) \in \Omega_T \equiv \Omega \times [0, T], \quad (1)$$

where \mathbf{u} is an m -vector and the matrices A_i are $m \times m$, symmetric, and constant. We assume Ω is a bounded polyhedral domain in R^N and denote its boundary by $\Gamma(\Omega)$. Likewise, we denote the boundary of the space-time domain Ω_T by $\Gamma(\Omega_T)$. Along $\Gamma(\Omega_T)$, the unit outer normal $\mathbf{n} = (n_x, n_t) = (n_1, \dots, n_N, n_t)$ has either $n_x = 0$ or $n_t = 0$.

An appropriate set of initial and boundary conditions for (1) is

$$\begin{aligned} \mathbf{u} &= \mathbf{g} && \text{at } t = 0, \\ (D - \mathcal{N})\mathbf{u} &= 0 && \text{on } \Gamma(\Omega) \times [0, T], \end{aligned} \quad (2)$$

where $D = \sum_{i=1}^N n_i A_i$ and $\mathcal{N} + \mathcal{N}^* \geq 0$. Problem (1)-(2) has the form of a Friedrichs system [4] for which a unique solution is guaranteed under certain restrictions.

An example of (1)-(2) is the wave equation in two space dimensions:

$$\begin{aligned} w_{tt} - w_{xx} - w_{yy} &= f, \\ w, w_t &\text{ given at } t = 0, \\ w &= 0 \text{ on } \Gamma(\Omega) \times [0, T]. \end{aligned}$$

* The authors were supported in part by NSF grant DMS-9704556 and DARPA grant 4-23685, respectively .

With $u_1 = w_x, u_2 = w_y, u_3 = w_t$, this can be written as

$$\mathbf{u}_t + \begin{pmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix} \mathbf{u}_x + \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & -1 & 0 \end{pmatrix} \mathbf{u}_y = \begin{pmatrix} 0 \\ 0 \\ f \end{pmatrix}.$$

For this system,

$$D = \begin{pmatrix} 0 & 0 & -n_1 \\ 0 & 0 & -n_2 \\ -n_1 & -n_2 & 0 \end{pmatrix},$$

and we may take

$$\mathcal{N} = \begin{pmatrix} 0 & 0 & n_1 \\ 0 & 0 & n_2 \\ -n_1 & -n_2 & 2 \end{pmatrix}.$$

Of the many previous finite element treatments of the general problem (1)-(2) and its related non-transient counterpart (e.g., [1], [5], [7], [8]), we know of none which is explicit, i.e., develops an approximate solution in an element by element fashion. This is our focus, in the setting of arbitrarily large m and N .

The key mesh requirement for explicitness is that

$$M \equiv n_t I + \sum_{i=1}^N n_i A_i$$

be definite on all interior (i.e., $\not\subset \Gamma(\Omega) \times [0, T]$) faces of each element K where (\mathbf{n}_x, n_t) denotes the unit outer normal to K . We denote by $\Gamma_{\text{in}}(K)$ ($\Gamma_{\text{out}}(K)$) the portion of $\Gamma(K)$ for which M is negative (positive) definite. The above explicitness condition will hold if all element faces are inclined sufficiently toward the x -hyperplane to make $\|\sum_{i=1}^N n_i A_i\| < |n_t|$. The sign of n_t will then indicate the direction of explicitness. In addition to the definiteness condition, we assume the eigenvalues $\lambda(M)$ of M are bounded away from zero:

$$|\lambda(M)| \geq \gamma > 0. \quad (3)$$

With an “explicit” mesh, we can discretize (1)-(2) via the following extension of the discontinuous Galerkin method:

$$a(\mathbf{u}_h, \mathbf{v}_h)_K = (\mathbf{f}, \mathbf{v}_h)_K, \quad \text{all } \mathbf{v}_h \in \mathcal{S}_h(K), \quad (4)$$

$$a(\mathbf{u}, \mathbf{v})_K \equiv (\mathcal{L}\mathbf{u}, \mathbf{v})_K - \int_{\Gamma_{\text{in}}(K)} [\mathbf{u}]^T M \mathbf{v} + \frac{1}{2} \int_{\Gamma^*(K)} \mathbf{u}^T (\mathcal{N} - D) \mathbf{v}. \quad (5)$$

The approximation subspace $\mathcal{S}_h(K)$ is comprised of polynomials of total degree $\leq n$ over K or is an n th degree tensor product space, $\mathbf{u}_h \in \mathcal{S}_h(K)$ is the finite element approximation, and $(\cdot, \cdot)_K$ denotes the $L^2(K)$ inner product; $\Gamma^*(K)$ denotes the intersection, if any, of $\Gamma(K)$ with $\Gamma(\Omega) \times [0, T]$. In general, \mathbf{u}_h will be discontinuous on interelement boundaries. We denote by \mathbf{u}_h^- and \mathbf{u}_h^+ its upstream and downstream limits, respectively, and use the notation $[\mathbf{u}_h] = \mathbf{u}_h^+ - \mathbf{u}_h^-$.

We will sketch the derivation of the following error estimate for \mathbf{u}_h as defined by (4)-(5):

$$\begin{aligned} & \|\mathbf{u}_h^- - \mathbf{u}\|_{T_{\text{out}}(\Omega_T)}^2 + \|\mathbf{u}_h - \mathbf{u}\|_{\Omega_T}^2 + h \sum_{K \subset \Omega_T} \|\mathcal{L}_0(\mathbf{u}_h - \mathbf{u})\|_K^2 \\ & + \sum_{K \subset \Omega_T} \|[\mathbf{u}_h - \mathbf{u}]\|_{\Gamma_{\text{in}}(K)}^2 \leq O(h^{2n+1}). \end{aligned} \quad (6)$$

This is an extension of the standard error estimate, first given in [6], for the discontinuous Galerkin method. We use the notation $\|\cdot\|_{k,D}$ for the norm on $H^k(D)$, $D \subset \Omega_T$, omitting k when it has value zero, and denote "surface" L^2 norms (e.g., over $\Gamma_{\text{in}}(K)$) by $|\cdot|$. The principal part of \mathcal{L} is denoted by \mathcal{L}_0 , and C denotes a generic constant independent of h but which, in general, is different at each occurrence.

The estimate (6) is essentially the same as that obtained in [2]. We use a different approach here, however, establishing (6) directly without first showing existence and stability of \mathbf{u}_h . Also, we employ an additional test function ($v_h = \mathcal{L}_0(\mathbf{u}_h - \mathbf{u}_I)$, where \mathbf{u}_I is an interpolant of \mathbf{u}) not used in [2], which eliminates the need for a technical assumption made in [2] (that each element be convex or have "sufficiently many" faces in comparison to n , the degree of approximation). It also allows (6) to be obtained with an arbitrary optimal order interpolant.

An outline of the paper is as follows. In §2 we detail the requirements for an explicit mesh, and in §3 we describe some ways to generate such a mesh. In §4 we outline the derivation of the estimate (6) for the discontinuous Galerkin method. Finally, in §5, we briefly consider a pair of continuous explicit finite element methods for (1)-(2), which work well for the simplest case $\Omega \subset \mathbb{R}^1$, but have significant shortcomings when generalized to higher dimension.

2 Requirements for explicitness

To elucidate the domain of dependence properties of (1), we consider the homogeneous equation

$$\mathcal{L}_0 \mathbf{u} = 0, \quad \mathcal{L}_0 \equiv \frac{\partial}{\partial t} + \sum_{i=1}^N A_i \frac{\partial}{\partial x_i}$$

in a generic polyhedral element K in the interior of the space-time domain. Integrating against \mathbf{u} , we get

$$\oint_{\Gamma(K)} \mathbf{u}^T M \mathbf{u} = 0, \quad M = n_t I + \sum_{i=1}^N n_i A_i. \quad (7)$$

We require that M be *definite* on each face of $\Gamma(K)$. This will be the case if K can be chosen so that $\|\sum_{i=1}^N n_i A_i\| < |n_t|$ on $\Gamma(K)$. A sufficient condition for this is

$$\frac{|\mathbf{n}_x|}{|n_t|} \leq \frac{1}{\sqrt{N} \max_i \|A_i\|}. \quad (8)$$

The sign of n_t will then indicate the inflow and outflow portions of $\Gamma(K)$, and the flow direction will be that of increasing t . We also require the inflow and outflow portions of $\Gamma(K)$ to be connected; otherwise explicit development of a solution will not be possible.

Assuming M is definite on each face of $\Gamma(K)$, we may recast (7) as

$$\int_{\Gamma_{\text{out}}(K)} \mathbf{u}^T M \mathbf{u} = \int_{\Gamma_{\text{in}}(K)} \mathbf{u}^T (-M) \mathbf{u},$$

where $\Gamma_{\text{out}}(K)$ ($\Gamma_{\text{in}}(K)$) denotes the portion of $\Gamma(K)$ where M ($-M$) is positive definite. Applying (3),

$$|\mathbf{u}|_{\Gamma_{\text{out}}(K)}^2 \leq C\gamma^{-1} \left(\max_{\Gamma_{\text{in}}(K)} \|M\| \right) |\mathbf{u}|_{\Gamma_{\text{in}}(K)}^2.$$

Thus $\mathbf{u} \equiv 0$ on $\Gamma_{\text{in}}(K)$ implies $\mathbf{u} \equiv 0$ on $\Gamma_{\text{out}}(K)$. Now for an arbitrary point $(\mathbf{x}^*, t^*) \in K$, we may construct a smaller polyhedral element $K' \subset K$ such that $\Gamma_{\text{in}}(K') \subset \Gamma_{\text{in}}(K)$ and $(\mathbf{x}^*, t^*) \in \Gamma_{\text{out}}(K')$. Integrating against \mathbf{u} over K' , we infer that $\mathbf{u}(\mathbf{x}^*, t^*) = 0$ if $\mathbf{u} \equiv 0$ on $\Gamma_{\text{in}}(K)$. Thus $\mathbf{u} \equiv 0$ on $\Gamma_{\text{in}}(K)$ implies $\mathbf{u} \equiv 0$ throughout K .

In a similar way, we may obtain a local stability result for a discrete model of (1). Suppose

$$\mathcal{L}_0 \mathbf{u}_h = \mathbf{f}_h \quad \text{in } K,$$

where $\mathbf{u}_h \in \mathcal{S}_h(K)$ is given on $\Gamma_{\text{in}}(K)$. Reasoning as before, we conclude that if $\mathbf{f}_h \equiv 0$ in K and $\mathbf{u}_h \equiv 0$ on $\Gamma_{\text{in}}(K)$, then $\mathbf{u}_h \equiv 0$ in K . Since \mathbf{u}_h in K may be regarded as the solution of a linear algebraic system with data \mathbf{f}_h and \mathbf{u}_h on $\Gamma_{\text{in}}(K)$, we infer that $\|\mathbf{u}_h\|_K$ can be bounded by a linear combination of $|\mathbf{u}_h|_{\Gamma_{\text{in}}(K)}$ and $\|\mathbf{f}_h\|_K$. Applying the appropriate scaling, we get for this bound:

$$\|\mathbf{u}_h\|_K \leq C \left(\sqrt{h} |\mathbf{u}_h|_{\Gamma_{\text{in}}(K)} + h \|\mathbf{f}_h\|_K \right).$$

Equivalently, for any $\mathbf{v}_h \in \mathcal{S}_h(K)$,

$$\|\mathbf{v}_h\|_K \leq C \left(\sqrt{h} |\mathbf{v}_h|_{\Gamma_{\text{in}}(K)} + h \|\mathcal{L}_0 \mathbf{v}_h\|_K \right). \quad (9)$$

We will use this bound later.

We briefly consider the wave equation example. Here

$$M = \begin{pmatrix} n_t & 0 & -n_1 \\ 0 & n_t & -n_2 \\ -n_1 & -n_2 & n_t \end{pmatrix},$$

whose eigenvalues are $\lambda = n_t, n_t \pm \sqrt{n_1^2 + n_2^2} = n_t \pm |\mathbf{n}_x|$. Thus M will be definite if $|\mathbf{n}_x| < |n_t|$. Condition (8) is more restrictive: $|\mathbf{n}_x| \leq |n_t|/\sqrt{2}$.

3 Mesh construction

We now consider the problem of generating an explicitly configured mesh of polyhedral space-time elements for $\Omega \times [0, T]$. As our starting point, we assume an appropriate face-conforming mesh of elements \mathcal{T}_h is given for the spatial domain Ω . Let $\mathcal{X}_h = \{\mathbf{x}_i\}$ denote the nodes of \mathcal{T}_h and $\mathcal{N}(\mathbf{x}_i)$ the set of neighboring vertices that share a common element with \mathbf{x}_i . The space-time mesh will be created incrementally, in the direction of increasing t . Its forward extent at $\mathbf{x}_i \in \mathcal{X}$ at any stage in its development will be denoted by $t_{\max}(\mathbf{x}_i)$. Each space-time element in our construction will be centered about a particular \mathbf{x}_i , will have \mathbf{x}_i and $\mathcal{N}(\mathbf{x}_i)$ as x coordinates of its vertices, and will advance $t_{\max}(\mathbf{x}_i)$ to its next value while leaving $t_{\max}(\mathbf{x}_j), j \neq i$, unaltered. To elucidate the parallelism possibilities, we shall assign each spatial vertex $\mathbf{x}_i \in \mathcal{X}$ a “color” $C(\mathbf{x}_i) \in \{1, 2, \dots\}$ subject to the condition $\mathbf{x}_j \in \mathcal{N}(\mathbf{x}_i) \implies C(\mathbf{x}_j) \neq C(\mathbf{x}_i)$.

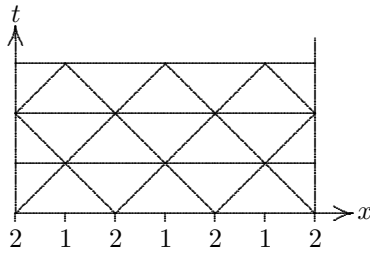


Fig. 3.1a

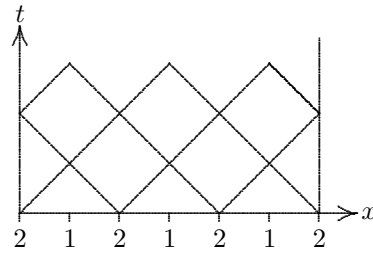


Fig. 3.1b

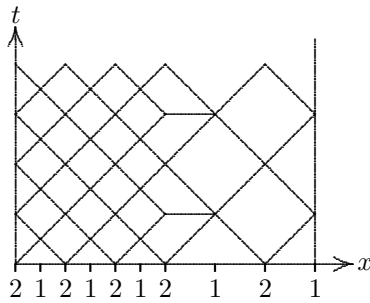


Fig. 3.1c

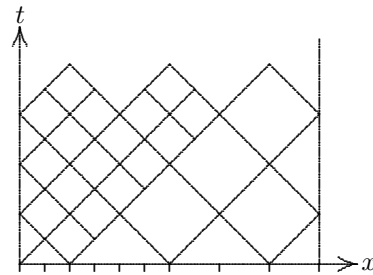


Fig. 3.1d

We first consider the case $\Omega = [0, 1] \subset R^1$, with Ω divided into uniform subintervals of width h . As indicated in Fig. 3.1a, two colors suffice for the spatial nodes $\{\mathbf{x}_i\}$. In the first step, we may advance t_{\max} at nodes of color 1 to $t = \Delta t$. Assuming Δt is chosen sufficiently small in comparison to h to ensure explicitness, the PDE solution can be developed concurrently in all such space-time elements centered about vertices of color 1. If in steps 2, 3, 4, 5, 6, 7, ..., we advance t_{\max} at nodes

of color 2, 2, 1, 1, 2, 2, ... to $t = \Delta t, 2\Delta t, 2\Delta t, 3\Delta t, 3\Delta t, 4\Delta t, \dots$, respectively, we obtain the space-time mesh shown in Fig. 3.1a. A second alternative is to follow step 1 above by steps $2', 3', 4', 5', \dots$ in which t_{\max} at nodes of color 2, 1, 2, 1, ... is advanced to $t = 2\Delta t, 3\Delta t, 4\Delta t, 5\Delta t, \dots$, resulting in the mesh of Fig. 3.1b. This mesh is comprised of a single generic element, a rhombus, and is twice as efficient at “consuming space” as the first scheme. One could, of course, bring the solution back to a common t at a subsequent time if desired.

Next suppose our spatial discretization of $\Omega = [0, 1]$ is nonuniform. We consider a simple case of a two-for-one mesh refinement in Fig 3.1c. The space-time mesh depicted there results if, in steps 1, 2, 3, 4, ..., nodes of color 1, 2, 1, 2, ... are advanced to their maximum t values consistent with explicitness, but coarse mesh nodes are not updated in steps 3, 4 (also 7, 8, ...). In general, one would like the frequency of update to vary inversely to the spatial grid size. This mesh illustrates the possibility of achieving two potentially desirable objectives: an explicit mesh, and a locally varying time step tailored to the degree of spatial refinement needed. (The more common, more rigid, alternative is to not have any spatial variation in time step). Another possible mesh generation technique would be to start with a uniform coarse mesh, consisting of congruent rhombuses and refine on a four-for-one basis where needed. Fig. 3.1d illustrates this for a case of a moving mesh.

We now turn to the more interesting case $\Omega \subset R^2$. Suppose, initially, that our spatial mesh consists of equilateral triangles of side length h . This illustrated in Fig. 3.2, where a 3-coloring of the corresponding nodes is also indicated. In analogy with the $\Omega \subset R^1$ case, we may, in steps 1,2,3,4,5,6,..., advance nodes of color 1,2,3,3,2,1, ... to $t = \Delta t, \Delta t, \Delta t, 2\Delta t, 2\Delta t, 2\Delta t, \dots$. Now, however, the elements so generated in each step are each unions of 6 tetrahedra and have either 7 or 12 (in steps 2, 5, 8, ...) faces. For this scheme $|n_x|/|n_t| \leq (2\Delta t)/(\sqrt{3}h)$.

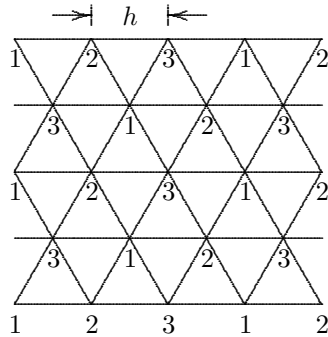


Fig. 3.2

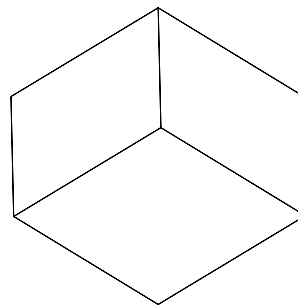


Fig. 3.3

A second alternative is: in steps 1,2,3,4,5,6,..., advance nodes of color 1,2,3,1,2,3, ... to $t = \Delta t, 2\Delta t, 3\Delta t, 4\Delta t, 5\Delta t, 6\Delta t, \dots$. For this scheme, $|\mathbf{n}_x|/|n_t| \leq 2\Delta t/h$; thus the maximum time step Δt for this scheme must be smaller than the previous one. Overall, however, the second scheme uses fewer elements to fill a given volume in the space-time domain. Moreover, the second scheme uses a single generic element (apart from boundary effects) which is, in fact, a hexahedron with two opposite vertices lying along a line of constant \mathbf{x} , parallel to the t -axis, as pictured in Fig. 3.3. Thus it may be viewed as a higher dimensional analog of the mesh depicted in Fig. 3.1b. Perhaps the simplest way to generate an explicit mesh for the case $\Omega \subset R^2$ would be use a coarse mesh of such hexahedra and then refine locally on an eight-for-one basis as appropriate. This scheme, as well as the first one, generalize readily to higher dimension.

4 Analysis

We shall restrict our attention here to interior elements K for which $\Gamma^*(K) = \emptyset$ in (5). The more general case is dealt with in [2]. We also assume the mesh is quasiuniform and nondegenerate (allowing the use of inverse inequalities).

We begin by giving a pair of identities for the bilinear form $a(\mathbf{u}, \mathbf{v})$ defined in (5). By integrating $a(\mathbf{u}, \mathbf{v})$ by parts, then performing some manipulations on the $\Gamma_{\text{in}}(K)$ integrals, we get:

$$\begin{aligned} a(\mathbf{u}, \mathbf{v})_K &= -a(\mathbf{v}, \mathbf{u})_K + ((B + B^*)\mathbf{u}, \mathbf{v})_K + \oint_{\Gamma(K)} (\mathbf{u}^-)^T M \mathbf{v}^- \\ &\quad - \int_{\Gamma_{\text{in}}(K)} [\mathbf{u}]^T M [\mathbf{v}]. \end{aligned} \quad (10)$$

Taking $\mathbf{v} = \mathbf{u}$ in (10), then using (3), we obtain:

$$\begin{aligned} a(\mathbf{u}, \mathbf{u})_K &= \frac{1}{2} \oint_{\Gamma(K)} (\mathbf{u}^-)^T M \mathbf{u}^- + \frac{1}{2} \int_{\Gamma_{\text{in}}(K)} [\mathbf{u}]^T (-M) [\mathbf{u}] \\ &\quad + \frac{1}{2} ((B + B^*)\mathbf{u}, \mathbf{u})_K \\ &\geq \frac{1}{2} \oint_{\Gamma(K)} (\mathbf{u}^-)^T M \mathbf{u}^- + \frac{\gamma}{2} |[\mathbf{u}]|_{\Gamma_{\text{in}}(K)}^2 + \frac{1}{2} ((B + B^*)\mathbf{u}, \mathbf{u})_K. \end{aligned} \quad (11)$$

We now assume the continuous problem and its discrete counterpart have solutions \mathbf{u} and \mathbf{u}_h , respectively, and estimate the difference between the two. From the derived estimate, it will follow that if the continuous problem has a solution, then \mathbf{u}_h is well-defined for h sufficiently small. It will be convenient to use an interpolant $\mathbf{u}_I \in \mathcal{S}_h(K)$ for \mathbf{u} that, we assume, will give optimal order accuracy if \mathbf{u} is sufficiently smooth, i.e.,

$$\|\mathbf{u}_I - \mathbf{u}\|_K + \sqrt{h} |\mathbf{u}_I - \mathbf{u}|_{\Gamma(K)} \leq Ch^{n+1} \|\mathbf{u}\|_{n+1, K}. \quad (12)$$

Subtracting $a(\mathbf{u}, \mathbf{v}_h)_K = (\mathbf{f}, \mathbf{v}_h)_T$ from (4) and introducing \mathbf{u}_I , we obtain:

$$a(\mathbf{e}_h, \mathbf{v}_h)_K = a(\mathbf{u} - \mathbf{u}_I, \mathbf{v}_h)_K, \quad \mathbf{e}_h \equiv \mathbf{u}_h - \mathbf{u}_I. \quad (13)$$

In what follows, we shall denote by $I(K)$ the ‘‘inflow’’ elements to K , lying immediately upstream from K .

The basic ingredients of the error estimate for \mathbf{u}_h are expressed in the following:

Lemma 1. (i) *The choice $\mathbf{v}_h = \mathbf{e}_h$ in (13) yields, for arbitrary $\epsilon > 0$:*

$$\begin{aligned} & \frac{1}{2} \oint_{\Gamma(K)} ((\mathbf{u}_h - \mathbf{u})^-)^T M(\mathbf{u}_h - \mathbf{u})^- + \frac{\gamma}{2} \|\mathbf{e}_h\|_{\Gamma_{in}(K)}^2 \\ & \leq \epsilon \left(\|\mathbf{e}_h\|_{\Gamma_{in}(K)}^2 + h \|\mathcal{L}_0 \mathbf{e}_h\|_K^2 \right) + C \left(\|\mathbf{e}_h\|_K^2 + \epsilon^{-1} h^{2n+1} \|\mathbf{u}\|_{n+1, K \cup I(K)}^2 \right). \end{aligned} \quad (14)$$

(ii) *The choice $\mathbf{v}_h = \mathcal{L}_0 \mathbf{e}_h$ in (13) yields:*

$$\|\mathcal{L}_0 \mathbf{e}_h\|_K^2 \leq C \left(\|\mathbf{e}_h\|_K^2 + h^{-1} \|\mathbf{e}_h\|_{\Gamma_{in}(K)}^2 + h^{2n} \|\mathbf{u}\|_{n+1, K \cup I(K)}^2 \right). \quad (15)$$

(iii) *\mathbf{e}_h satisfies:*

$$\|\mathbf{e}_h\|_K^2 \leq C \left(h \|\mathbf{e}_h^-\|_{\Gamma_{in}(K)}^2 + h \|\mathbf{e}_h\|_{\Gamma_{in}(K)}^2 + h^2 \|\mathcal{L}_0 \mathbf{e}_h\|_K^2 \right). \quad (16)$$

Proof. (i) By taking $\mathbf{v}_h = \mathbf{e}_h$ in (13), applying (10) and (11), then the Schwarz inequality, arithmetic-geometric mean inequality, and inverse inequalities, we get the following bounds:

$$\begin{aligned} a(\mathbf{e}_h, \mathbf{e}_h)_K & \geq \frac{1}{2} \oint_{\Gamma(K)} (\mathbf{e}_h^-)^T M \mathbf{e}_h^- + \frac{\gamma}{2} \|\mathbf{e}_h\|_{\Gamma_{in}(K)}^2 - C \|\mathbf{e}_h\|_K^2, \\ a(\mathbf{u} - \mathbf{u}_I, \mathbf{e}_h)_K & = - \left((\mathcal{L}_0 \mathbf{e}_h + B \mathbf{e}_h, \mathbf{u} - \mathbf{u}_I)_K - \int_{\Gamma_{in}(K)} [\mathbf{e}_h]^T M (\mathbf{u} - \mathbf{u}_I)^+ \right) \\ & + ((B + B^*) (\mathbf{u} - \mathbf{u}_I), \mathbf{e}_h)_K + \oint_{\Gamma(K)} ((\mathbf{u} - \mathbf{u}_I)^-)^T M \mathbf{e}_h^- \\ & \quad - \int_{\Gamma_{in}(K)} [\mathbf{u} - \mathbf{u}_I]^T M [\mathbf{e}_h] \\ & \leq \oint_{\Gamma(K)} ((\mathbf{u} - \mathbf{u}_I)^-)^T M \mathbf{e}_h^- + \epsilon \left(\|\mathbf{e}_h\|_{\Gamma_{in}(K)}^2 + \|\mathbf{e}_h\|_K^2 + h \|\mathcal{L}_0 \mathbf{e}_h\|_K^2 \right) \\ & \quad + C \epsilon^{-1} \left(h^{-1} \|\mathbf{u} - \mathbf{u}_I\|_K^2 + \|\mathbf{u}_h - \mathbf{u}_I\|_{\Gamma_{in}(K)}^2 + \|(\mathbf{u} - \mathbf{u}_I)^+\|_{\Gamma_{in}(K)}^2 \right). \end{aligned}$$

Combining these bounds, then completing the square on the $\Gamma(K)$ integral, we get (14)

(ii) Using similar techniques, for $\mathbf{v}_h = \mathcal{L}_0 \mathbf{e}_h$ in (13), we get:

$$\begin{aligned}
 a(\mathbf{e}_h, \mathcal{L}_0 \mathbf{e}_h)_K &= (\mathcal{L}_0 \mathbf{e}_h + B \mathbf{e}_h, \mathcal{L}_0 \mathbf{e}_h)_K - \int_{\Gamma_{\text{in}}(K)} [\mathbf{e}_h]^T M(\mathcal{L}_0 \mathbf{e}_h)^+ \\
 &\geq \|\mathcal{L}_0 \mathbf{e}_h\|_K^2 - \|B\| \|\mathbf{e}_h\|_K \|\mathcal{L}_0 \mathbf{e}_h\|_K - C \|\mathbf{e}_h\|_{\Gamma_{\text{in}}(K)} \frac{\|\mathcal{L}_0 \mathbf{e}_h\|_K}{\sqrt{h}} \\
 &\geq \frac{1}{2} \|\mathcal{L}_0 \mathbf{e}_h\|_K^2 - C(\|\mathbf{e}_h\|_K^2 + h^{-1} \|\mathbf{e}_h\|_{\Gamma_{\text{in}}(K)}^2), \\
 a(\mathbf{u} - \mathbf{u}_I, \mathcal{L}_0 \mathbf{e}_h)_K &= (\mathcal{L}(\mathbf{u} - \mathbf{u}_I), \mathcal{L}_0 \mathbf{e}_h)_K - \int_{\Gamma_{\text{in}}(K)} [\mathbf{u} - \mathbf{u}_I]^T M(\mathcal{L}_0 \mathbf{e}_h)^+ \\
 &\leq C \|\mathbf{u} - \mathbf{u}_I\|_{1,K} \|\mathcal{L}_0 \mathbf{e}_h\|_K + C \|\mathbf{u} - \mathbf{u}_I\|_{\Gamma_{\text{in}}(K)} \frac{\|\mathcal{L}_0 \mathbf{e}_h\|_K}{\sqrt{h}} \\
 &\leq \frac{1}{4} \|\mathcal{L}_0 \mathbf{e}_h\|_K^2 + C \left(\|\mathbf{u} - \mathbf{u}_I\|_{1,K}^2 + h^{-1} \|\mathbf{u} - \mathbf{u}_I\|_{\Gamma_{\text{in}}(K)}^2 \right).
 \end{aligned}$$

The result of these bounds is (15)

(iii) We may provide for a jump discontinuity in \mathbf{v}_h across $\Gamma_{\text{in}}(K)$ in (9) by writing $|\mathbf{v}_h^+|_{\Gamma_{\text{in}}(K)} \leq |\mathbf{v}_h^-|_{\Gamma_{\text{in}}(K)} + \|[\mathbf{v}_h]\|_{\Gamma_{\text{in}}(K)}$. Applying the resulting bound to \mathbf{u}_h , we get (16)

Multiplying (14)-(16) by 1, μh , ν , respectively, then adding, then applying the bound $|\mathbf{e}_h^-|_{\Gamma_{\text{in}}(K)} \leq |(\mathbf{u}_h - \mathbf{u})^-|_{\Gamma_{\text{in}}(K)} + Ch^{n+1/2} \|\mathbf{u}\|_{n+1, I(K)}$, gives:

$$\begin{aligned}
 &\frac{1}{2} \oint_{\Gamma(K)} ((\mathbf{u}_h - \mathbf{u})^-)^T M(\mathbf{u}_h - \mathbf{u})^- + \left(\frac{\gamma}{2} - \epsilon - C\mu - C\nu h\right) \|\mathbf{e}_h\|_{\Gamma_{\text{in}}(K)}^2 \\
 &\quad + (\mu h - \epsilon h - C\nu h^2) \|\mathcal{L}_0 \mathbf{e}_h\|_K^2 + (\nu - C - C\mu h) \|\mathbf{e}_h\|_K^2 \\
 &\leq C \left(h |(\mathbf{u}_h - \mathbf{u})^-|_{\Gamma_{\text{in}}(K)}^2 + (1 + \epsilon^{-1}) h^{2n+1} \|\mathbf{u}\|_{n+1, K \cup I(K)}^2 \right).
 \end{aligned}$$

We next take $\mu < \gamma/4C$ to allow coercivity of $\|\mathbf{e}_h\|_{\Gamma_{\text{in}}(K)}^2$, then take $\nu > C\mu h + C + 1$ to coerce $\|\mathbf{e}_h\|_K^2$, then choose ϵ small enough to coerce $\|\mathbf{e}_h\|_{\Gamma_{\text{in}}(K)}^2$ and $\|\mathcal{L}_0 \mathbf{e}_h\|_K^2$ for h sufficiently small. We can write the result as follows:

Lemma 2. *There exist positive constants γ_1 and γ_2 such that for h sufficiently small,*

$$\begin{aligned}
 &\oint_{\Gamma(K)} ((\mathbf{u}_h - \mathbf{u})^-)^T M(\mathbf{u}_h - \mathbf{u})^- + \|\mathbf{e}_h\|_K^2 + \gamma_1 \|\mathbf{e}_h\|_{\Gamma_{\text{in}}(K)}^2 + \gamma_2 h \|\mathcal{L}_0 \mathbf{e}_h\|_K^2 \\
 &\leq C \left(h |(\mathbf{u}_h - \mathbf{u})^-|_{\Gamma_{\text{in}}(K)}^2 + h^{2n+1} \|\mathbf{u}\|_{n+1, K \cup I(K)}^2 \right). \quad (17)
 \end{aligned}$$

Assuming sufficient smoothness in \mathbf{u} and applying this bound over all $K \subset \Omega_T$ then yields (3)(cf. [2]).

5 Continuous explicit finite element methods

We briefly consider the possibility of explicitly generating a continuous finite element method over an appropriate mesh. The generic form of such a method is

$$\begin{aligned} \mathbf{u}_h &\in \mathcal{S}_h(K) \\ (\mathcal{L}\mathbf{u}_h, \mathbf{v}_h)_K &= (\mathbf{f}, \mathbf{v}_h)_K, \quad \text{all } \mathbf{v}_h \in \mathcal{T}_h(K) \end{aligned}$$

Here $\mathcal{T}_h(K)$ must have dimension less than that of $\mathcal{S}_h(K)$ because \mathbf{u}_h will already be known on $T_{\text{in}}(K)$ at the time when it is to be computed in K . A potential advantage of a continuous method is the smaller number of degrees of freedom in \mathbf{u}_h , hence fewer unknowns to be solved for.

We first mention a method [9] which can be applied over a mesh of triangles like that depicted in Fig. 3.1a.

$$\begin{aligned} \mathbf{u}_h &\in \mathcal{P}_n(K) \\ (\mathcal{L}\mathbf{u}_h, \mathbf{v}_h)_K &= (\mathbf{f}, \mathbf{v}_h)_K, \quad \text{all } \mathbf{v}_h \in \mathcal{P}_{n-\rho(K)}(K). \end{aligned} \quad (18)$$

Here \mathcal{P}_n consists of polynomials of total degree $\leq n$ and $\rho(K)$ denotes the number of in-flow sides that K has (either one or two). This method typically gives $O(h^{n+1})$ convergence, like the discontinuous Galerkin method; an analysis appears in [3].

The method (18) extends directly to higher dimension over simplices K . For the case $\Omega \subset R^2$, the elements are tetrahedra which may have either 1, 2, or 3 in-flow faces (i.e., $\rho(K) = 1, 2$ or 3). Thus there are three possible test spaces for (18), and, not surprisingly, no analysis. In addition, n must be at least 2 (otherwise two of the three possible test spaces in (18) will be void), and an explicit tetrahedral mesh seems impractical to construct and manage. Thus (18) does not look promising for $N \geq 2$.

We also mention a continuous method for $\Omega \subset R^1$ due to Winther [10], which can be applied over a mesh of parallelograms like that depicted in Fig. 3.1b. It is:

$$\begin{aligned} \mathbf{u}_h &\in \Pi_n(K) \\ (\mathcal{L}\mathbf{u}_h, \mathbf{v}_h)_K &= (\mathbf{f}, \mathbf{v}_h)_K, \quad \text{all } \mathbf{v}_h \in \Pi_{n-1}(K). \end{aligned} \quad (19)$$

Here $\Pi_n(K)$ is a tensor product space of polynomials of degree $\leq n$ in coordinates ξ, η aligned with the parallelogram sides. Optimal order error estimates are derived in [10]. This method, too, extends immediately to higher dimension. However, a simple calculation reveals that for the simplest case of (1), $\mathbf{u}_t = 0$, in two space dimensions and time, with $n = 1$ (linear approximation), (19) has an algebraic instability arising from a nondecaying spurious root of multiplicity 2. This casts doubt on the usefulness of (19) for $N \geq 2$.

By contrast, the discontinuous Galerkin method is stable regardless of N and very flexible in terms of applicability.

References

1. Du, Q., Gunzburger, M., Layton, W.: A low dispersion, high accuracy finite element method for first order hyperbolic systems in several space variables. Hyperbolic partial differential equations V, Comput. Math. Appl. **15** (1988) 447–457

2. Falk, R.S., Richter, G.R.: Explicit finite element methods for symmetric hyperbolic equations. *SIAM J. Numer. Anal.*, to appear.
3. Falk, R.S., Richter, G.R.: Analysis of a continuous finite element method for hyperbolic equations. *SIAM J. Numer. Anal.* **24** (1987) 257–278
4. Friedrichs, K. O.: Symmetric positive linear differential equations. *Comm. Pure Appl. Math.* **11** (1958) 333–418
5. Johnson, C., Nävert, U., Pitkäranta, J.: Finite element methods for linear hyperbolic problems. *Comput. Methods Appl. Mech. Engrg.* **45** (1984) 285–312
6. Johnson, C., Pitkäranta, J.: An analysis of the discontinuous Galerkin method for a scalar hyperbolic equation. *Math. Comp.* **46** (1986) 1–26
7. Layton, W.: High-accuracy finite-element methods for positive symmetric systems. *Hyperbolic partial differential equations III*, *Comput. Math. Appl. Part A* **12** (1986) 565–579
8. Lesaint, P.: Finite element methods for symmetric hyperbolic equations. *Numer. Math.* **21** (1973) 244–255
9. Reed, W. H., Hill, T. R.: Triangular mesh methods for the neutron transport equation. Los Alamos Scientific Laboratory Technical Report LA-UR-73-479 (1973)
10. Winther, R.: A stable finite element method for first-order hyperbolic systems *Math. Comp.* **36** (1981) 65–86