M2AN. MATHEMATICAL MODELLING AND NUMERICAL ANALYSIS
- MODÉLISATION MATHÉMATIQUE ET ANALYSE NUMÉRIQUE

K. W. MORTON
A. PRIESTLEY
E. SULI

Stability of the Lagrange-Galerkin method with non-exact integration

M2AN. Mathematical modelling and numerical analysis - Modélisation mathématique et analyse numérique, tome 22, n° 4 (1988), p. 625-653

<http://www.numdam.org/item?id=M2AN_1988__22_4_625_0>

MATHEMATICAL MODELLING AND NUMERICAL ANALYSIS
MODÉLISATION MATHÉMATIQUE ET ANALYSE NUMÉRIQUE
(Vol. 22, n° 4, 1988, p. 625 à 653)

STABILITY OF THE LAGRANGE-GALERKIN METHOD WITH NON-EXACT INTEGRATION (*)

by K. W. MORTON (1), A. PRIESTLEY (1) and E. SULI (1)

Communiqué par R. Temam

Abstract. — *The Lagrange-Galerkin finite element method for a linear advection problem is unconditionally stable if exact integration is used for the evaluation of the inner products However, great care must be taken when non-exact integration is performed Large classes of well-known quadrature rules lead to conditionally unstable schemes An alternative technique is presented and shown to be stable*

Résumé. — *La méthode d'éléments finis de Lagrange-Galerkin est inconditionnellement stable pour le problème d'advection linéaire, si le calcul des produits internes est obtenu à l'aide d'une intégration exacte Cependant il faut être prudent dans le cas où une intégration non exacte est utilisée Les méthodes deviennent conditionnellement instables dans le cas de nombreuses catégories d'intégrations numériques. Une autre technique est donc présentée et démontrée stable*

0. INTRODUCTION

In the past ten years there has been an increased effort aimed at solving advection-dominated diffusion problems. One of the new numerical techniques designed for dealing with such problems is the Lagrange-Galerkin method which is based on combining the method of characteristics with a standard finite element procedure (see, Benqué *et al.* [2], Bercovier & Pironneau [3], Douglas & Russell [6], Lesaint [12], Pironneau [15], Russell [16], Suli [17]). In these papers the unconditional stability and the convergence of the Lagrange-Galerkin method has been demonstrated for a wide class of problems. However, most of the analysis is based on the

(*) Received in June 1987, revised in April 1988. The work reported here forms part of the research programme of the Oxford/Reading Institute for Computational Fluid Dynamics

(1) Oxford University Computing Laboratory, Numerical Analysis Group, 8-11 Keble Road, Oxford OX1 3QD, England.

M² AN Modélisation mathématique et Analyse numérique 0399-0516/88/04/625/29/$ 4.90

assumption that all inner products are evaluated exactly, though some study of the influence of quadrature has been performed by Lesaint [12] and Russell [16].

The purpose of this paper is to show that if standard numerical quadrature techniques are used to evaluate the inner products, the method becomes conditionally unstable. Moreover, in most cases the stability condition cannot be met by limiting the time-step : Lobatto quadrature is a notable exception. An alternative quadrature technique, derived from particle methods and called area-weighting, is presented and shown to be stable.

In the next section two formulations of the Lagrange-Galerkin method are given for a linear advection equation and some basic properties of the exactly integrated schemes are demonstrated. Section 2 is devoted to the analysis of the effect of quadrature. We prove that quadrature rules including Gauss-Legendre, Gauss-Lobatto, Radau and Newton-Cotes lead to conditionally unstable schemes. In Section 3 the alternative area-weighting technique is presented and its stability and convergence is proved. Section 4 is devoted to numerical examples.

In the following, C will denote a positive generic constant, independent of the discretization parameters.

1. FORMULATION AND PROPERTIES OF THE EXACTLY INTEGRATED LAGRANGE-GALERKIN METHOD

1.1. Derivation of the basic schemes

Consider the Cauchy problem for the scalar, linear advection equation for $u(\underline{x}, t)$:

$$u_t + \underline{a} \cdot \underline{\nabla} u = 0 , \quad \underline{x} \in \mathbb{R}^d , \quad t > 0 , \tag{1.1.1a}$$

$$u(\underline{x}, 0) = u_0(\underline{x}) , \tag{1.1.1b}$$

where u_0 belongs to $L^2(\mathbb{R}^d)$ and the velocity field $\underline{a}(\underline{x}, t)$ is incompressible, i.e.

$$\underline{\nabla} \cdot \underline{a} = 0 \quad \forall \underline{x}, t . \tag{1.1.2}$$

We can define characteristic paths or trajectories, $\underline{X}(\underline{x}, s ; t)$, by

$$\underline{X}(\underline{x}, s ; s) = \underline{x} , \tag{1.1.3a}$$

$$\frac{d}{dt} \underline{X}(\underline{x}, s ; t) = \underline{a}(\underline{X}(\underline{x}, s ; t), t) ; \tag{1.1.3b}$$

or, if desired, (1.1.3) can be replaced by

$$\underline{X}(\underline{x}, s ; t) = \underline{x} + \int_s^t \underline{a}(\underline{X}(\underline{x}, s ; \tau), \tau) \, d\tau . \tag{1.1.4}$$

For $t^{n+1} = t^n + \Delta t$, $\underline{X}(\underline{y}, t^{n+1}; t^n)$ and $\underline{X}(\underline{x}, t^n; t^{n+1})$ will be denoted by $\underline{x}$ and $\underline{y}$ respectively.

A unique (absolutely continuous) solution to (1.1.3) can be guaranteed if it is assumed that $\underline{a}$ belongs to the Bochner space $L^1(0, T; (W^{1,\infty})^d)$. The relation

$$u(\underline{X}(\cdot, t; t+\tau), t+\tau) = u(\cdot, t) \tag{1.1.5}$$

then gives us the solution to (1.1.1).

The most direct formulation (indeed, we shall refer to this approach as the direct Lagrange-Galerkin method) for an approximation at time t^n given in terms of finite element basis functions ϕ_j,

$$U^n = \sum_j U_j^n \phi_j , \tag{1.1.6}$$

uses (1.1.5) directly to obtain U^{n+1} in $L^2(\mathbb{R}^d)$ satisfying

$$\langle U^{n+1}, \phi_i \rangle = \int U^n(\underline{x})\, \phi_i(\underline{y})\, d\underline{y} \tag{1.1.7}$$

where $\langle \cdot, \cdot \rangle$ denotes the L^2 inner product over $\mathbb{R}^n$. Here we have used the Galerkin method applied to an integrated form (1.1.5) of the differential equation (1.1.1). This is the same approach as that used by Bercovier & Pironneau [3], Douglas & Russell [6] and Pironneau [15], for example.

A second, alternative, formulation has been proposed by Benqué *et al.* [2] and this will be referred to as the weak formulation or weak Lagrange-Galerkin method because the adjoint of the differential operator in (1.1.1*a*) is applied to a test function. It introduces new test functions $\psi_t(\cdot, t)$, which now are not only different from the basis functions but depend on time. Multiplying the equation (1.1.1) by this test function and integrating in space and time we get

$$\int_t^{t+\Delta t} \langle u_t + \underline{a} \cdot \underline{\nabla} u, \psi_t \rangle \, dt = 0 . \tag{1.1.8a}$$

Integrating by parts, with respect to either space or time, we obtain

$$\langle u(\cdot, t+\Delta t), \psi_t(\cdot, t+\Delta t) \rangle - \langle u(\cdot, t), \psi_t(\cdot, t) \rangle = \int_t^{t+\Delta t} \langle u, \partial_t \psi_t + \underline{\nabla} \cdot (\underline{a} \psi_t) \rangle \, dt . \tag{1.1.8b}$$

Using (1.1.2), the incompressibility condition, $\underline{\nabla} \cdot (\underline{a}\psi_t)$ can be rewritten as $\underline{a} \cdot \underline{\nabla}\psi_t$ so that this last term vanishes if the test functions satisfy

$$\psi_t(\underline{X}(\cdot, t; t+\tau), t+\tau) = \psi_t(\cdot, t) . \tag{1.1.8c}$$

To solve (1.1.8*c*) a final condition on ψ_i is imposed by setting

$$\psi_i(\cdot\,, t + \Delta t) = \phi_i(\cdot)\,. \tag{1.1.8d}$$

Substituting our finite element approximation (1.1.6) into (1.1.8*b*) together with (1.1.8*c*) and (1.1.8*d*) gives

$$\langle U^{n+1}, \phi_i \rangle = \int U^n(\underline{x})\, \psi_i(\underline{x}, t^n)\, d\underline{x}\,:$$

equivalently,

$$\langle U^{n+1}, \phi_i \rangle = \int U^n(\underline{x})\, \phi_i(\underline{y})\, d\underline{x}\,, \tag{1.1.9a}$$

where

$$\underline{y} = \underline{X}(\underline{x}, t^n\,; t^{n+1})\,. \tag{1.1.9b}$$

If J is the Jacobi matrix of the transformation defined by the mapping $\underline{X}(\cdot\,, t^n\,; t^{n+1})$, then we have, as shown in Chorin and Marsden [5],

$$\frac{d|J|}{dt} = (\underline{\nabla} \cdot \underline{a})|J|\,,$$

where $|J|$ is the determinant of J. For $\underline{\nabla} \cdot \underline{a} = 0$ this means that

$$|J| = \text{constant} = 1\,, \tag{1.1.10}$$

which in turn implies that $d\underline{x} \equiv d\underline{y}$ and so (1.1.9) and (1.1.7) give the same scheme when exact integration is used.

Remark : When $\underline{\nabla} \cdot \underline{a} \neq 0$ the direct and the weak formulations are not the same. With $\underline{x}$ and $\underline{y}$ still related by (1.1.3), the direct form (1.1.7) still approximates (1.1.1), whereas the weak form, with the test functions given by (1.1.8*c, d*), approximates instead the equation $u_t + \underline{\nabla} \cdot (\underline{a}u) = 0$.

Remark : In the context of hyperbolic problems and shock-modelling, a third formulation has been used for what have been called characteristic Galerkin methods but which are equivalent to Lagrange-Galerkin methods for linear advection — see Morton & Sweby [14] and Childs & Morton [4].

1.2. Some properties of the exactly integrated schemes

If finite elements are used for which $\sum_i \phi_i \equiv 1$, which it will be assumed always holds, then conservation for either of the formulations follows

immediately. Suppose also that we denote by $E_{\Delta t}(t)$ the solution operator $u(\cdot\,, t+\Delta t) = E_{\Delta t}(t)\,u(\cdot\,, t)$ and similarly $E_{\Delta t}^{-1}(t)\,u(\cdot\,, t+\Delta t) = u(\cdot\,, t)$ for the equation (1.1.1) over the time-step Δt ; that is from (1.1.5)

$$E_{\Delta t}(t)\,u(\cdot\,, t) = u(\cdot\,, t+\Delta t) = u(\underline{X}(\cdot\,, t+\Delta t\,; t), t)\,. \tag{1.2.1}$$

Now, where $\|\cdot\|$ denotes the usual L^2 norm over $\mathbb{R}^n$, and $\|\cdot\|_*$ is the correspondingly defined operator norm,

$$\|E_{\Delta t}(t)\|_* = \sup_{u \neq 0} \frac{\|E_{\Delta t}(t)\,u(\cdot\,, t)\|}{\|u(\cdot\,, t)\|} = \sup_{u \neq 0} \frac{\|u(\cdot\,, t+\Delta t)\|}{\|u(\cdot\,, t)\|}\,,$$

and from (1.2.1) and (1.1.10) this equals 1 and hence

$$\|E_{\Delta t}(t)\|_* = 1 = \|E_{\Delta t}^{-1}(t)\|_*\,. \tag{1.2.2}$$

The unconditional stability of the direct or weak methods will now follow immediately.

For the direct method we can write

$$\langle U^{n+1}, \phi_i \rangle = \langle E_{\Delta t}(t^n)\,U^n, \phi_i \rangle\,. \tag{1.2.3}$$

Multiplying by U_i^{n+1} and summing over i gives

$$\begin{aligned} \|U^{n+1}\|^2 &= \langle E_{\Delta t}(t^n)\,U^n, U^{n+1} \rangle \\ &\leqslant \|E_{\Delta t}(t^n)\,U^n\| \cdot \|U^{n+1}\|\,. \end{aligned}$$

Hence

$$\|U^{n+1}\| \leqslant \|U^n\|\,.$$

Similarly for the weak method, we can write

$$\langle U^{n+1}, \phi_i \rangle = \langle U^n, E_{\Delta t}^{-1}(t^n)\,\phi_i \rangle\,. \tag{1.2.4}$$

By virtue of (1.2.2), and the Cauchy-Schwarz inequality, we have,

$$\|U^{n+1}\| \leqslant \|U^n\|\,,$$

to again deduce unconditional stability.

THEOREM 1.1 : *The direct Lagrange-Galerkin method with C^0 finite elements of degree k ($k \geqslant 1$) converges with order k in the $\ell^\infty(0, T\,; (L^2)^d)$ norm, provided that $u_0 \in (H^{k+1})^d$, $\underline{a} \in L^\infty(0, T\,; (W^{1,\infty})^d)$ and the corresponding solution u of* (1.1.1) *belongs to the space $H^1(0, T\,; (H^{k+1})^d)$.*

Proof: We introduce $\eta^n = u^n - Iu^n$ and $\xi^n = Iu^n - U^n$, where u^n denotes $u(\cdot\,, n\,\Delta t)$ and Iu^n is a piecewise polynomial interpolant of u^n of degree k. From (1.2.3) we have

$$\begin{aligned}\langle \xi^{n+1} - E_{\Delta t}(t^n)\,\xi^n, \xi^{n+1}\rangle &= \langle Iu^{n+1} - E_{\Delta t}(t^n)\,Iu^n, \xi^{n+1}\rangle \\ &= \langle E_{\Delta t}(t^n)\,\eta^n - \eta^{n+1}, \xi^{n+1}\rangle \\ &= \langle E_{\Delta t}(t^n)\,\eta^n - \eta^n, \xi^{n+1}\rangle + \langle \eta^n - \eta^{n+1}, \xi^{n+1}\rangle .\end{aligned}$$

Thus, by using the Cauchy-Schwarz inequality for all inner products and (1.2.2),

$$\|\xi^{n+1}\| \leq \|\xi^n\| + \|E_{\Delta t}(t^n)\,\eta^n - \eta^n\| + \|\eta^{n+1} - \eta^n\| . \qquad (1.2.5)$$

Now it remains to estimate the right-hand side terms in (1.2.5). Following the convergence proof in Süli [17],

$$\|E_{\Delta t}(t^n)\,\eta^n - \eta^n\| \leq C\,\Delta t\,\|\underline{\nabla}\eta^n\| ,$$

$$\|\eta^{n+1} - \eta^n\| \leq \int_{t_n}^{t_{n+1}} \left\| \frac{d\eta}{dt}\,(t)\right\| dt ,$$

so that

$$\|\xi^m\| \leq \|\xi^0\| + \left\| \frac{d\eta}{dt} \right\|_{L^1(0,T,\,(L^2)^d)} + C\,\|\underline{\nabla}\eta\|_{L^\infty(0,T;\,(L^2)^d)} .$$

Assuming that U^0 has been chosen to be Iu^0, standard interpolation results yield the desired estimate :

$$\|u - U\|_{\ell^\infty(0,T;\,(L^2)^d)} \leq Ch^k \|u\|_{H^1(0,T;\,(H^{k+1})^d)} .$$

Remark : In some special cases, the order of convergence in Theorem 1.1 may be improved. For example, for one-dimensional constant linear advection, the Lagrange-Galerkin method with piecewise linear elements on a uniform grid is third order accurate. Furthermore, when $\underline{a}$ is a smooth function, the method is second order accurate in the $\ell^\infty(0, T\,; \ell^2)$ norm on a non-uniform mesh.

Remark : For a generalization of the flux-corrected transport algorithm of Boris and Book a result similar to Theorem 1.1 has been proved by Lesaint [12].

2. THE EFFECT OF QUADRATURE

2.1. Introduction

Except for some one-dimensional flows, the evaluation of the integrals in either the direct or the weak formulation is going to require some form of approximation and then the two methods will not in general be identical. Even for the relatively simple flow of the rotating cone problem to be described in the last section, the exact evaluation of the integrals (1.1.8) or (1.1.10) is not practicable.

Some analysis of the influence of quadrature on Lagrange-Galerkin methods has been performed by Lesaint [12] and by Russell [16] and this shows that $0(h^k/\Delta t)$ terms are added to the error, where k is the degree of the element. For linear elements on a regular mesh this can be improved to $0(h^2/\Delta t)$, which suggests that if we choose $\Delta t \geqslant C\ \Delta x$ the scheme will still converge, albeit at a reduced rate.

In this section we will present results that show that, for certain quadrature rules, it is not so much the mesh ratio $\frac{\Delta t}{\Delta x}$ which must be bounded away from zero but the CFL (*Courant-Friedrichs-Lewy*) number $|\underline{a}|\ \frac{\Delta t}{\Delta x}$. However, in general, we may not be able to bound this away from zero unless $\underline{a}(\underline{x}, t)$ is a constant. For other important classes of integration rules our results give a severe stability restriction.

For the sake of simplicity, each result on the influence of quadrature will be presented for just one of the Lagrange-Galerkin methods, the proof being similar for the other. The analysis in d dimensions can be greatly simplified by the following result.

LEMMA 2.1 : *For constant linear advection in d dimensions the Lagrange-Galerkin method is just a tensor product of d one-dimensional Lagrange-Galerkin methods, assuming that the basis functions themselves are tensor products of the corresponding one-dimensional basis functions on a grid which is uniform in each coordinate direction.*

Proof: We shall prove this lemma only for $d = 2$. The general case can be handled by induction. In one dimension it is necessary to evaluate, in the direct method (1.1.7) for example, the following integrals :

$$\int \phi_i(x) \sum_j (T_x\, U_j^n\, \phi_j(x))\, dx \quad \forall i\ ,$$

where $T_x\, f(x) = f(x - a\, \Delta t)$. This is rewritten in a more convenient form as

$$I_x\, \phi_i(x)\, T_x\, U^n(x) \quad \forall i\ , \tag{2.1.1}$$

where the I_x just represents the integration in x.

In two dimensions the mass matrix is clearly the tensor product of two one-dimensional mass matrices and so it remains to see if this is also true of the right-hand side. Clearly, (2.1.1) becomes

$$I_{x_1 x_2} \phi_i(x_1)\, \phi_j(x_2)\, T_{x_1 x_2} U^n(x_1, x_2) \quad \forall i, j \tag{2.1.2}$$

where

$$T_{x_1 x_2} f(x_1, x_2) := f(\underline{x} - \underline{a}\, \Delta t)\,,$$

so that

$$T_{x_1 x_2} = T_{x_1} T_{x_2}\,,$$

since a_1 and a_2 are constants. The following relationship also clearly holds, since a uniform rectangular mesh is being employed,

$$I_{x_1 x_2} = I_{x_1} I_{x_2}.$$

Now T_{x_2} and I_{x_2} have no effect on a function solely of x_1 and similarly T_{x_1} and I_{x_1} have no effect on a function solely of x_2 and so (2.1.2) becomes

$$I_{x_1 x_2} \phi_i(x_1)\, \phi_j(x_2)\, T_{x_1 x_2} U^n(\underline{x}) = (I_{x_1} \phi_i(x_1)\, T_{x_1})(I_{x_2} \phi_j(x_2)\, T_{x_2})\, U^n(\underline{x})\,.$$

Therefore the two-dimensional case is just the product of the one-dimensional cases. ■

A second simplification is the following :

LEMMA 2.2 : *For one-dimensional constant linear advection the weak and direct Lagrange-Galerkin methods are equivalent, provided that the basis functions and the quadrature rule are symmetric.*

Proof: See Morton & Priestley [13].

2.2. The instability of quadrature with linear elements

Thus we consider here only one-dimensional constant linear advection, and it is most convenient to work with the weak form. Generally one will want the scheme to reduce to the identity when the advection speed is zero. Thus if the exact mass matrix is to be used for the left-hand side of (1.1.9), the quadrature rule to be used on the right should be exact for quadratics. Such cases are covered in the main theorem below. But there are two simple quadrature rules that we deal with first. In all results the CFL number ν can be replaced by $\nu + m$, where m is an arbitrary integer.

LEMMA 2.3 : *The (weak) Lagrange-Galerkin method with linear elements on a uniform mesh is unconditionally unstable when the mass matrix and the*

right-hand side are both evaluated by centroid quadrature : when the exact mass matrix is used it is unstable for CFL numbers $\nu \in [1/\sqrt{6}, 1 - 1/\sqrt{6}]$; *and when the lumped mass matrix is used it is unconditionally stable.*

Proof : Our scheme is

$$\langle U^{n+1}, \phi_i \rangle = \int U^n(x)\, \phi_i(y)\, dx\,, \tag{2.2.1}$$

where $y = x + a\,\Delta t$. We shall consider only CFL numbers $\nu \in [0, 1/2]$ for which the three elements between $(i-2)\,h$ and $(i+1)\,h$ may contribute to the terms in (2.2.1) ; the interval $[1/2, 1]$ is handled by an analogous argument. We need to compute the right-hand side of (2.2.1) and apply Fourier analysis. Contributions to the right-hand side are tabulated below :

Element centre	Contribution to $\sum wU^n(x)\,\phi_i(y)$
$(i-3/2)\,h$	0
$(i-1/2)\,h$	$0.5\,(U_{i-1} + U_i)(0.5 + \nu)$
$(i+1/2)\,h$	$0.5\,(U_i + U_{i+1})(0.5 - \nu)$

Hence, with the mass matrix obtained by setting $\nu = 0$ here, the scheme (2.2.1) becomes, in operator notation,

$$\left[1 + \frac{\delta^2}{4}\right] U_i^{n+1} = \left[1 + \frac{\delta^2}{4} - \nu\,\Delta_0\right] U_i^n\,. \tag{2.2.2}$$

The standard notation has been used here :

$$\delta^2 U_i := U_{i+1} - 2\,U_i + U_{i-1}\,, \tag{2.2.3a}$$

$$\Delta_0\, U_i := (1/2)(U_{i+1} - U_{i-1})\,, \tag{2.2.3b}$$

$$\Delta_-\, U_i := U_i - U_{i-1}\,. \tag{2.2.3c}$$

Replacing the operators by their Fourier transforms gives

$$[1 - s^2]\,\lambda = 1 - s^2 - 2\,\nu isc\,, \tag{2.2.4}$$

where $s = \sin(\xi h/2)$ and $c = \cos(\xi h/2)$, and λ is the amplification factor. For stability we require $|\lambda|^2 \leqslant 1$ for all ξ in $[-\pi, \pi]$ which can clearly never hold for $\nu > 0$.

If we were to use the exact mass matrix, the left-hand side of (2.2.4) would become $[1 - 2\,s^2/3]\,\lambda$ and stability would require

$$1 - 2\,s^2 + s^4 + 4\,\nu^2\,s^2(1 - s^2) \leqslant 1 - \frac{4\,s^2}{3} + \frac{4\,s^4}{9}\,:$$

it is easily seen that we must have $\nu^2 \leqslant 1/6$ for this to be satisfied.

Finally if the lumped mass matrix is used, so that the term in the brackets on the left of (2.2.4) were removed, it is easily checked that $|\lambda|^2 = 1 - 2s^2(1 - 2\nu^2) + s^4(1 - 4\nu^2) \leqslant 1$ for all $s^2 \in [0, 1]$ if $\nu \in [0, 1/2]$. ■

LEMMA 2.4 : *The Lagrange-Galerkin method, with linear elements on a uniform mesh, is unconditionally stable when both the mass matrix and the right-hand side are evaluated by vertex quadrature : but when the exact mass matrix is used it is unconditionally unstable.*

Proof: Proceeding as with centroid quadrature we get contributions to the i-th component from only two different nodes and obtain the familiar first order upwind scheme :

$$U_i^{n+1} = U_i^n - \nu(U_i^n - U_{i-1}^n), \tag{2.2.5}$$

which is stable for all $\nu \in [0, 1]$.

However, the exact mass matrix introduces the operator $[1 + \delta^2/6]$ on the left to give, after Fourier transforms,

$$\left[1 - \frac{2s^2}{3}\right]\lambda = 1 - 2\nu(s^2 + isc). \tag{2.2.6}$$

Hence

$$\left[1 - \frac{2s^2}{3}\right]^2 |\lambda|^2 = 1 - 4\nu(1 - \nu)s^2$$

which requires for $|\lambda|^2 \leqslant 1$ that

$$\nu(1 - \nu) \geqslant \frac{1}{3} - \frac{s^2}{9} \quad \forall s^2 \in [0, 1]. \tag{2.2.7}$$

Since the maximum value of the left-hand side is only 1/4, (2.2.7) is not satisfied for any $\nu \in [0, 1]$. ■

Remark : Eastwood [7] with his EPIC algorithm, which is identical to the direct method, uses a compound trapezium rule. The same quadrature rule should also be applied to the mass matrix in order to obtain an unconditionally stable scheme. However, to avoid significant loss of accuracy compared with the exactly integrated scheme, the mass matrix has to be approximated quite accurately.

Of all the common types of quadrature, centroid (lowest order Gauss-Legendre) and vertex (lowest order Gauss-Lobatto) are the only two that do not integrate quadratics exactly. Having dealt with these, we can deal with all the higher order quadratures by the following theorem. In what follows, the usual abbreviations of Gauss-Legendre to Gauss or Gaussian and Gauss-Lobatto to Lobatto are made.

THEOREM 2.5 : *If the right-hand side of the (weak) Lagrange-Galerkin method, using piecewise linear elements on a uniform mesh, is approximated by a quadrature of the form*

$$\int_0^1 f(x)\,dx \approx w_0\, f(0) + \sum_1^m w_k\, f(x_k) + w_{m+1}\, f(1)\,, \qquad (2.2.8)$$

where the weights $w_0, \ldots, w_{m+1}$ *and the quadrature points* $0 < x_1 < \cdots < x_m < 1$ *are free to be chosen except that we assume that the quadrature evaluates the integrals of quadratic polynomials exactly, then the method is unstable for* $\nu \in (2\, w_{m+1}, 1 - x_m)$ *if*

$$2\, w_{m+1} < 1 - x_m\,. \qquad (2.2.9)$$

Proof: We just consider $\nu \in [0, 1 - x_m]$ and proceed in the same way as for centroid and vertex quadrature by looking at all the quadrature points that give non-zero contributions to $\phi_i(y)$ in (2.2.1) :

Element centre	Contribution to $\sum wU^n(x)\,\phi_i(y)$
$(i - 3/2)\,h$	$w_{m+1}\, U_{i-1}\, \nu$
$(i - 1/2)\,h$	$\sum_{k=0}^{m} w_k[(1 - x_k)\, U_{i-1} + x_k\, U_i](x_k + \nu) + w_{m+1}\, U_i(1 - \nu)$
$(i - 1/2)\,h$	$\sum_{k=0}^{m} w_k[(1 - x_k)\, U_i + x_k\, U_{i+1}](1 - x_k - \nu)\,.$

This gives for the right-hand side

$$\left[\sum_0^{m+1} w_k(1 - \nu\Delta_- + x_k\,\delta^2 - x_k^2\,\delta^2 - x_k\,\nu\delta^2) + w_{m+1}\,\nu\delta^2\right] U_i^n\,. \qquad (2.2.10)$$

Since the quadrature rule integrates quadratics exactly, we have

$$\sum_0^{m+1} w_k = 1\,, \quad \sum_0^{m+1} w_k\, x_k = 1/2\,, \quad \sum_0^{m+1} w_k\, x_k^2 = 1/3\,. \qquad (2.2.11)$$

Using (2.2.11) to simplify (2.2.10) and replacing the difference operators by their Fourier transforms the scheme gives

$$\left[1 - \frac{2\,s^2}{3}\right]\lambda = 1 - \frac{2\,s^2}{3} - 2\,i\nu sc - 4\,w_{m+1}\,\nu s^2\,. \qquad (2.2.12)$$

For $|\lambda|^2 \leqslant 1$ we require, upon simplification,

$$-2\,w_{m+1} + \frac{4}{3}\,w_{m+1}\,s^2 + 4\,w_{m+1}^2\,\nu s^2 + \nu(1-s^2) \leqslant 0 \quad \forall s^2 \in [0,1]\,.$$

This means that for stability we require

$$\nu \leqslant 2\,w_{m+1} \quad \text{and} \quad w_{m+1}\left(w_{m+1}\,\nu - \frac{1}{6}\right) \leqslant 0\,.$$

In most cases, the first inequality is more restrictive than the second and hence the claimed instability results if

$$2\,w_{m+1} < 1 - x_m\,. \tag{2.2.13}$$ ■

Thus, by combining Lemma 2.3 with Theorem 2.5, we see that any Gauss-Legendre quadrature that integrates quadratics exactly and has no quadrature point at $x = 1$ will lead to a method with a region of instability. The natural choice might therefore be a Lobatto quadrature : but we have the following result.

COROLLARY 2.6 : *All Lobatto quadratures except the vertex quadrature lead to conditionally unstable Lagrange-Galerkin methods.*

We need first the following lemma.

LEMMA 2.7 : *If z_n is the n-th zero of P'_{n+1}, where P_{n+1} is the $(n+1)$-st Legendre polynomial, then*

$$z_n \leqslant 1 - \frac{4}{n(n+3)}\,, \quad n > 0\,. \tag{2.2.14}$$

Proof: Since

$$P'_{n+1}(z) = zP'_n(z) + (n+1)\,P_n(z)$$

and $P_n(1) = 1$, it follows that

$$P'_{n+1}(1) = \frac{(n+1)(n+2)}{2} > 0\,. \tag{2.2.15}$$

We also have

$$(1-z^2)\,P''_{n+1} - 2\,zP'_{n+1} + (n+1)(n+2)\,P_{n+1} = 0\,. \tag{2.2.16}$$

Differentiating (2.2.16) gives

$$(1-z^2)\,P'''_{n+1} - 4\,zP''_{n+1} + \{(n+1)(n+2) - 2\}\,P'_{n+1} = 0\,, \tag{2.2.17}$$

and putting $z = 1$ into (2.2.17) yields

$$P''_{n+1}(1) = \frac{n(n+3)}{4}\,P'_{n+1}(1)\,. \tag{2.2.18}$$

By Rolle's Theorem, $P''_{n+1}(z) \neq 0$ for all z in $[z_n, 1]$ and hence by (2.2.18) $P''_{n+1}(z) > 0$ for all z in $[z_n, 1]$. By a similar argument the third derivative can be shown to be positive in the entire interval $[z_n, 1]$. The error after one step of Newton's method for finding z_n from the initial value 1 is given by

$$\text{error} = -\frac{1}{2}\left(\frac{P'''_{n+1}(\eta)}{P''_{n+1}(1)}\right) e_0^2,$$

where e_0 is the initial error and $\eta \in [z_n, 1]$. As can be seen this error is always negative and so Newton's method gives us the following result after one iteration

$$z_n \leq 1 - \frac{P'_{n+1}(1)}{P''_{n+1}(1)}. \tag{2.2.19}$$

Now using (2.2.15) and (2.2.18) in (2.2.19) we obtain the desired result. ■

Proof of Corollary 2.6 : It is a well known result, Krylov [11], that for Lobatto quadrature $w_0 = w_{m+1} = \dfrac{1}{(m+1)(m+2)}$, and that the quadrature points are the zeroes of $P'_{m+1}(z)$, Abramovich & Stegun [1]. Rearranging (2.2.13), transforming the interval from [0, 1] to [− 1, 1], since this is the region on which the Legendre polynomials are defined, and substituting for w_{m+1} we see that Lobatto quadrature is conditionally unstable if z_m, the m-th zero of $P'_{m+1}(z)$, satisfies

$$z_m < 1 - \frac{4}{(m+1)(m+2)}. \tag{2.2.20}$$

Lemma 2.7 gives the required result, except for vertex quadrature where $m = 0$. ■

Remark : As can be seen from the proof of Theorem 2.5 and Corollary 2.6, the Lagrange-Galerkin method based on Lobatto quadrature with $m > 0$ is stable for $\nu \in [0, 2/(m+1)(m+2)]$, and so its instability can be rectified by choice of time step. However, for higher order quadratures this condition puts a severe restriction on the time step.

For the Newton-Cotes formulae, w_{m+1} can be shown to satisfy (2.2.9) asymptotically for $m \to \infty$ since $1 - x_m = 1/(m+1)$ for all m and as $m \to \infty$

$$w_{m+1} = \frac{1}{(m+1)\ln(m+1)}\left[1 + 0\left(\frac{1}{(m+1)\ln(m+1)}\right)\right].$$

Moreover, for at least $m \leq 25$ (which includes all rules of any practical interest), one can easily check (see, Kopal [10]) that the formulae satisfy this inequality.

As with centroid and vertex quadrature, lumping the mass matrix stabilizes the scheme but drastically reduces the accuracy of the method. This is particularly true in more than one dimension, where the diagonal elements of the mass matrix are less dominant, making the expensive calculation of the right-hand side unjustifiable.

2.3. Stability with quadratic and constant basis functions

To gain accuracy some authors, e.g. Benqué *et al.* [2], have used quadratic elements. This, however, does not improve the stability.

THEOREM 2.8 : *The (weak) Lagrange-Galerkin method, using piecewise quadratic elements has regions of instability if the right-hand side is evaluated by using an interior quadrature that evaluates the integrals of quadratics exactly and the exact mass matrix is used.*

Proof: There are now two types of basis functions to consider. For that centred at an element boundary, the relevant row of the mass matrix is

$$\frac{1}{15}\,[-U_{i-2}+2\,U_{i-1}+8\,U_i+2\,U_{i+1}-U_{i+2}]^{n+1}\,, \qquad (2.3.1)$$

which has the Fourier transform

$$\frac{1}{15}\,[10+8\,s^2-16\,s^4]\left(=\frac{2}{3}\quad\text{at}\quad s^2=0\right)\,. \qquad (2.3.2)$$

We shall use a quadrature of the form

$$\int_0^2 f(x)\,dx=\sum_1^m w_k\,f(x_k)\,,\quad 0<x_1<\cdots<x_m<2\,, \qquad (2.3.3)$$

and assume that the relations (2.2.11) hold with right-hand sides 2, 2 and 8/3, respectively, and k ranging from 1 to m. As before we just consider $\nu\in[0,2-x_m]$.

It will be sufficient to study only the case $s^2=0$, that is we can take $U^n(x)\equiv 1$. Then the contributions to the right-hand side of (2.2.1) from ϕ_i are as follows for each quadrature point :

Element centre	Contribution to $\sum w\phi_i(y)$
$i-1$	$\frac{1}{2}\,[x_k^2+2\,x_k\,\nu+\nu^2-x_k-\nu]$
$i+1$	$\frac{1}{2}\,[x_k^2+2\,x_k\,\nu+\nu^2-3\,x_k-3\,\nu+2]$

So the right-hand side of the Lagrange-Galerkin method becomes

$$\sum_{1}^{m} w_k \{x_k^2 + 2 x_k \nu + \nu^2 - 2 x_k - 2 \nu + 1\} . \tag{2.3.4}$$

Using the assumptions about the accuracy of the quadrature to simplify (2.3.4), we obtain that the Fourier transform of the i-th component of the right-hand side evaluated at $s^2 = 0$ is $\frac{2}{3} + 2\nu^2$. Combining with (2.3.2) implies that for $|\lambda|^2|_{s^2=0} \leqslant 1$ we need $\left[\frac{2}{3} + 2\nu^2\right]^2 \leqslant \left[\frac{2}{3}\right]^2$, which is clearly only satisfied if $\nu = 0$, and hence the scheme is unstable for $\nu \in (0, 2 - x_m)$. ■

On the other hand we have the following result for constant elements.

THEOREM 2.9 : *The Lagrange-Galerkin method, using piecewise constant elements, is unconditionally stable provided that the weights, w_k, used in the quadrature are all positive and that the sum of the weights is* 1.

Proof: We may write the scheme for $0 \leqslant \nu \leqslant 1$ as

$$U_i^{n+1} = \left[\sum_{1}^{j} w_k\right] U_i^n + \left[\sum_{j+1}^{m} w_k\right] U_{i-1}^n ,$$

where $j \leqslant m$, so with the given hypotheses the unconditional stability follows from the maximum principle. ■

3. A STABLE IMPLEMENTATION OF THE METHOD

3.1. Introduction

We have seen that in most cases quadrature leads to unstable schemes. The range of instability, in terms of the CFL number, is $(2 w_{m+1}, 1 - x_m)$ and so clearly Lobatto quadratures are better placed than Gaussian quadratures or any quadrature with purely interior abscissae. Indeed Douglas *et al.* [6] do use Lobatto quadratures. Unfortunately, these are still conditionally unstable. This is a serious deficiency compared with the unconditional stability demonstrated with exact integration for simple problems in Section 1 and for the Navier-Stokes equations by Süli [17]. The unconditional stability of the exactly integrated scheme with linear elements can be recovered with the use of a compound trapezium rule on both sides of the equation (2.2.1). However, to achieve comparable accuracy by this means, many quadrature points are needed and hence the evaluation of the right-hand side becomes very expensive.

An efficient technique developed here to recover the stability properties of the exactly integrated Lagrange-Galerkin schemes is derived from a device employed by users of particle methods and is referred to in their literature as *area-weighting* (see Harlow [8] and Hockney & Eastwood [9]). The basic idea is as follows: in the conventional use of quadrature considered above it is supposed that the trajectories $\underline{x} \to \underline{y}$ are calculated exactly (or very accurately) for each quadrature point and then the consequential inner products (1.1.7) or (1.1.9*a*) evaluated approximately; in area-weighting we instead calculate the trajectories only approximately in such a way that the resultant inner products can be evaluated exactly. In the simplest case only the centroid of each element is tracked and the whole element deemed to move without distortion and rotation.

3.2. Area-weighting for the Lagrange-Galerkin method

In the (direct) Lagrange-Galerkin method we need to evaluate integrals like (1.1.7)

$$\int U^n(\underline{x})\, \phi_{ij}(\underline{y})\, d\underline{y} \quad \forall i, j\,,$$

where $\underline{x} = \underline{x}(\underline{y})$ is as defined in Section 1. This is assembled from a sum of integrals over the elements E_{kl},

$$\sum_{k,l} \int_{E_{kl}} U^n(\underline{x}(\underline{y}))\, \phi_{ij}(\underline{y})\, d\underline{y} \quad \forall i, j\,. \tag{3.2.1}$$

If a square mesh is assumed and $\underline{y}_c$ is used to denote the centroid of element (k, l) then (3.2.1) becomes

$$\sum_{k,l} \int_{-h/2}^{h/2} \int_{-h/2}^{h/2} U^n(\underline{x}(\underline{y}_c) + \underline{s})\, \phi_{ij}(\underline{y}_c + \underline{s})\, d\underline{s} \quad \forall i, j\,. \tag{3.2.2}$$

To apply area-weighting to this problem, $\underline{x}(\underline{y}_c + \underline{s})$ is approximated by $\underline{x}(\underline{y}_c) + \underline{s}$ so that in place of (3.2.2) we calculate

$$\sum_{k,l} \int_{-h/2}^{h/2} \int_{-h/2}^{h/2} U^n(\underline{x}(\underline{y}_c) + \underline{s})\, \phi_{ij}(\underline{y}_c + \underline{s})\, d\underline{s}\,. \tag{3.2.3}$$

for all the ϕ_{ij}'s. These integrals can then be calculated exactly.

It is not necessary to restrict ourselves to a square mesh as the algorithm clearly works equally well on a rectangular grid. Also we have assumed no

distortion of the square/rectangle about its centre : but this assumption is not necessary and for domains within fixed boundaries one has to relax it. However, we do only allow the square to distort to a rectangle so that the intersections with the underlying grid remain rectangles and hence exact integrations remain possible. This also places a limitation on the theoretical results that can be obtained for the scheme.

Finally, it is clearly a simple matter to subdivide each element into subrectangles and treat each of these by area-weighting. In this way there is no limit to the accuracy that can be achieved stably by the area-weighting technique.

3.3. Theoretical results for area-weighting

For constant linear advection it is clear that area-weighting is exact so the scheme is then unconditionally stable. Thus to study the effect of the scheme we must consider the more general problem (1.1.1)-(1.1.2), for which we cannot use Fourier analysis. This we do for the direct method ; the weak method can be handled in a similar way.

Before proceeding with the stability and convergence results some notation is defined : u^n is the true solution, U^n is the exactly integrated direct method solution, U^n_{aw} is the area-weighted solution ; $\underline{y}$ is the end point of a trajectory, $\underline{x} = \underline{x}(\underline{y}) = \underline{X}(\underline{y}, t^{n+1}; t^n)$ is the true foot of the trajectory, $\underline{x}_h = \underline{x}_h(\underline{y}) = \underline{y} - \underline{y}_c + \underline{X}(\underline{y}_c, t^{n+1}; t^n)$ is the area-weighted approximation to the foot of the trajectory with end point $\underline{y}$ in the element whose centroid is $\underline{y}_c$. Also $\mathscr{C}_h$ will denote a tensor product grid in $\mathbb{R}^n$, uniform in each coordinate direction.

The area-weighted direct Lagrange-Galerkin approximation U^{n+1}_{aw} can then be defined by

$$\langle U^{n+1}_{aw}, \phi_i \rangle = \int U^n_{aw}(\underline{x}_h)\, \phi_i(\underline{y})\, d\underline{y} \tag{3.3.1}$$

which is to be compared with (1.1.7).

LEMMA 3.1 : *If the velocity field $\underline{a}$ belongs to $L^\infty(0, T ; (W^{1,\infty})^d)$, then the distance between the true foot of the trajectory and the area-weighted approximate foot is of the order $h\,\Delta t$.*

Proof: From the definition of the trajectories (1.1.4) we have

$$\underline{x}(\underline{y}) = \underline{y} - \int_0^{\Delta t} \underline{a}(\underline{X}(\underline{y}, t^{n+1}; t^n + \tau), t^n + \tau)\, d\tau$$

and therefore

$$\begin{aligned}\left|\underline{x}(\underline{y}_c + \underline{s}) - (\underline{x}(\underline{y}_c) + \underline{s})\right| &= \left| \int_0^{\Delta t} [\underline{a}(\underline{X}(\underline{y}_c + \underline{s}, t^{n+1}; t^n + \tau), t^n + \tau) - \right. \\ &\quad \left. - \underline{a}(\underline{X}(\underline{y}_c, t^{n+1}; t^n + \tau), t^n + \tau)]\, d\tau \right| \\ &\leq \int_0^{\Delta t} L \left| \underline{X}(\underline{y}_c + \underline{s}, t^{n+1}; t^n + \tau) - \underline{X}(\underline{y}_c, t^{n+1}; t^n + \tau) \right| d\tau \\ &\leq L\,|\underline{s}| \int_0^{\Delta t} e^{L(\Delta t - \tau)} d\tau = (e^{L\,\Delta t} - 1)\,|\underline{s}| = 0(h\,\Delta t)\,,\end{aligned}$$

where $L = \|a\|_{L^\infty(0,T;\,(W^{1,\infty})^d)}$.

For piecewise constant elements, the following lemma leads to a very simple stability proof.

LEMMA 3.2 : *If* $\underline{a}$ *belongs to* $L^\infty(0, T; (W^{1,\infty})^d)$ *then, for every function* V *which is piecewise constant on the mesh* $\mathscr{C}_h$,

$$\left| \|V(\underline{x}_h)\|^2 - \|V(\underline{x})\|^2 \right| \leq C\,\Delta t\,\|V\|^2\,.$$

Proof:

$$\|V(\underline{x}_h)\|^2 - \|V(\underline{x})\|^2 = \sum_{E \in \mathscr{C}_h} \int_E \{V(\underline{x}_h)^2 - V(\underline{x})^2\}\, d\underline{y}\,.$$

Whenever $\underline{x}_h(\underline{y})$ and $\underline{x}(\underline{y})$ fall in the same element the integrand is zero, and so it is only necessary to consider the situations where this is not the case. As can be seen from Figure 3.1, where the hashed area is where $V(\underline{x}_h)$ may

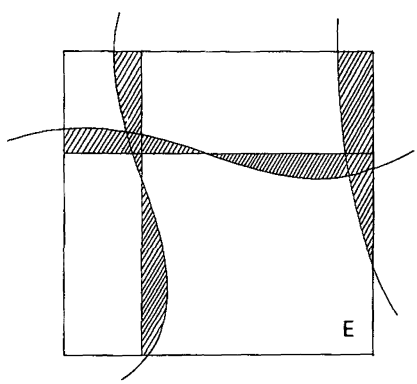

Figure 3.1. — Incidence of error in area-weighting technique. The curved (resp. straight) lines inside the square E indicate the inverse image of the intersection of $x_h(E)$ with the mesh lines under the transformation $x^{-1}(y)$(resp. $x_h^{-1}(y)$).

possibly differ from $V(\underline{x})$, the area within an element where the integrand may be non-zero is at most $0(h^{d-1})\,0(h\,\Delta t)$ or $0(h^d\,\Delta t)$. A summation of non-zero contributions over all elements, yields the desired estimate. ■

THEOREM 3.3 : *If $\underline{a}$ belongs to $L^\infty(0, T\,;\,(W^{1,\infty})^d)$, then the direct area-weighted Lagrange-Galerkin method with piecewise constant elements is stable in the $\ell^\infty(0, T\,;\,(L^2)^d)$ norm.*

Proof: For any basis functions the scheme may be written in the obvious notation as

$$\langle U_{aw}^{n+1}, \phi_i \rangle = \langle U_{aw}^n(\underline{x}_h), \phi_i \rangle \quad \forall i \,. \tag{3.3.2}$$

Multiplying (3.3.2) by the i-th nodal parameter of U_{aw}^{n+1} and summing over i yields

$$\| U_{aw}^{n+1} \|^2 = \langle U_{aw}^n(\underline{x}_h), U_{aw}^{n+1} \rangle \leqslant \| U_{aw}^n(\underline{x}_h) \| \, \| U_{aw}^{n+1} \| \,.$$

This implies

$$\begin{aligned} \| U_{aw}^{n+1} \|^2 \leqslant \| U_{aw}^n(\underline{x}_h) \|^2 &\leqslant \| U_{aw}^n(\underline{x}_h) \|^2 - \| U_{aw}^n(\underline{x}) \|^2 + \| U_{aw}^n \|^2 \\ &\leqslant C\ \Delta t \, \| U_{aw}^n \|^2 + \| U_{aw}^n \|^2 \end{aligned}$$

by Lemma 3.2. This means that

$$\| U_{aw}^{n+1} \|^2 \leqslant (1 + C\ \Delta t) \| U_{aw}^n \|^2$$

and so stability. ■

For higher order elements, the same result requires a different proof.

THEOREM 3.4 : *The direct Lagrange-Galerkin method evaluated with the area-weighted approximation is stable in the $\ell^\infty(0, T\,;\,(L^2)^d)$ norm for piecewise linear (and higher order) elements provided that $\underline{a}$ belongs to the space $L^\infty(0, T\,;\,(W^{1,\infty})^d)$.*

Proof: Let E_j denote an arbitrary element of $\mathscr{C}_h$ and let $\underline{c}_j$ be its centroid. For $\underline{y}$ in E_j and $0 \leqslant \theta \leqslant 1$, we define

$$\underline{H}_\theta(\underline{y}) := \theta \underline{x}(\underline{y}) + (1 - \theta)\underline{x}_h(\underline{y}) \,,$$

where $\underline{x}_h(\underline{y}) = \underline{x}(\underline{c}_j) + \underline{y} - \underline{c}_j$ is the area weighted approximation to the foot of the trajectory ending at $\underline{y}$. It is easy to show that there exists a positive constant C_0, independent of j and the discretization parameters, such that for $\Delta t \leqslant C_0$,

$$\frac{1}{2} \left| \underline{y}_1 - \underline{y}_2 \right| \leqslant \left| \underline{H}_\theta(\underline{y}_1) - \underline{H}_\theta(\underline{y}_2) \right| \leqslant 2 \left| \underline{y}_1 - \underline{y}_2 \right| \quad \forall\, \underline{y}_1, \underline{y}_2 \in E_j \ \forall \theta \in [0, 1].$$

Hence $\underline{H}_\theta$ is a quasi-isometry of E_j onto $\underline{H}_\theta(E_j)$ and the Jacobian of the mapping $\underline{H}_\theta : \underline{y} \in E_j \to \underline{H}_\theta(E_j)$ is not less then $(1/2)^d$ for $\Delta t \leqslant C_0$. The relation

$$\|U_{aw}^{n+1}\|^2 - \langle U_{aw}^n(\underline{x}), U_{aw}^{n+1}\rangle = \langle U_{aw}^n(\underline{x}_h) - U_{aw}^n(\underline{x}), U_{aw}^{n+1}\rangle$$

implies that

$$\|U_{aw}^{n+1}\| \leqslant \|U_{aw}^n\| + \|U_{aw}^n(\underline{x}_h) - U_{aw}^n(\underline{x})\| . \tag{3.3.3}$$

It remains to estimate the second term on the right-hand side.

$$\begin{aligned}
&\leqslant \sum_j \int_{E_j} |\underline{x}(\underline{y}) - \underline{x}_h(\underline{y})|^2 \int_0^1 |\nabla U_{aw}^n(\underline{H}_\theta(\underline{y}))|^2 \, d\theta \, d\underline{y} \\
&\leqslant \sum_j \int_{E_j} |\underline{x}(\underline{y}) - \underline{x}_h(\underline{y})|^2 \int_0^1 |\nabla U_{aw}^n(\underline{H}_\theta(\underline{y}))|^2 \, d\theta \, d\underline{y} \\
&\leqslant Ch^2 \Delta t^2 \sum_j \int_{E_j} \int_0^1 |\nabla U_{aw}^n(\underline{H}_\theta(\underline{y}))|^2 \, d\theta \, d\underline{y} ,
\end{aligned}$$

by virtue of Lemma 3.1. Using the properties of the mapping $\underline{H}_\theta$ to change variables,

$$\|U_{aw}^n(\underline{x}) - U_{aw}^n(\underline{x}_h)\|^2 \leqslant Ch^2 \Delta t^2 \int_0^1 \left(\sum_j \int_{\underline{H}_\theta(E_j)} |\nabla U_{aw}^n(\underline{y})|^2 \, d\underline{y} \right) d\theta .$$

Since there is only a finite amount of overlapping,

$$\|U_{aw}^n(\underline{x}) - U_{aw}^n(\underline{x}_h)\|^2 \leqslant Ch^2 \Delta t^2 \int_0^1 \|\nabla U_{aw}^n\|^2 \, d\theta .$$

As the integrand is independent of θ, we have

$$\|U_{aw}^n(\underline{x}) - U_{aw}^n(\underline{x}_h)\| \leqslant Ch \, \Delta t \, \|\nabla U_{aw}^n\| . \tag{3.3.4}$$

Finally, (3.3.3), (3.3.4) and the inverse property

$$\|\nabla U_{aw}^n\| \leqslant Ch^{-1} \|U_{aw}^n\|$$

imply

$$\|U_{aw}^{n+1}\| \leqslant (1 + C \, \Delta t) \|U_{aw}^n\| ,$$

and hence stability. ■

LEMMA 3.5 : *If* u *belongs to* $H^1(0, T ; (W^{1,\infty} \cap H^{k+1})^d)$, $k \geqslant 1$, *and* $u_0 \in (W^{1,\infty} \cap H^{k+1})^d$ *then*

$$\| U^n(\underline{x}) - U^n(\underline{x}_h) \| \leqslant Ch\, \Delta t ,$$

where U^n *denotes the exactly integrated direct method solution with* C^0 *finite elements of degree* k.

Proof: Dropping the superscripts we define $w(\cdot\,) = U(\cdot\,) - u(\cdot\,)$. Then

$$\| U(\underline{x}) - U(\underline{x}_h) \| \leqslant \| u(\underline{x}) - u(\underline{x}_h) \| + \| w(\underline{x}) - w(\underline{x}_h) \| . \quad (3.3.5)$$

For the first term on the right-hand side,

$$\| u(\underline{x}) - u(\underline{x}_h) \| \leqslant \| \underline{x} - \underline{x}_h \| \cdot \| u \|_{1,\infty} \leqslant Ch\, \Delta t \, \| u \|_{1,\infty} . \quad (3.3.6)$$

For the second term, following the stability proof, we have

$$\| w(\underline{x}) - w(\underline{x}_h) \| \leqslant Ch\, \Delta t \, \| \underline{\nabla} w \| . \quad (3.3.7)$$

It now remains to show that $\| \underline{\nabla} w \|$ is bounded by a constant, independent of the discretization parameters. To begin with,

$$\| \underline{\nabla} w \| = | w |_1 = | U - u |_1 \leqslant | U - Iu |_1 + | u - Iu |_1 ,$$

where Iu is the interpolant of u. Hence

$$\begin{aligned} \| \underline{\nabla} w \| &\leqslant \frac{C}{h} \| U - Iu \| + | u - Iu |_1 \\ &\leqslant \frac{C}{h} \{ \| U - u \| + \| u - Iu \| \} + | u - Iu |_1 . \end{aligned} \quad (3.3.8)$$

But $\| U - u \| = 0(h^k)$ from Theorem 1.1, and $\| u - Iu \| = 0(h^{k+1})$, $| u - Iu |_1 = 0(h^k)$ from standard interpolation results. Hence $\| \underline{\nabla} w \| = 0(h^{k-1})$ which in the worst case of $k = 1$ means that $\| \underline{\nabla} w \|$ is bounded by a constant, independent of the discretization parameters, and we have from (3.3.7) the estimate

$$\| w(\underline{x}) - w(\underline{x}_h) \| \leqslant Ch\, \Delta t . \quad (3.3.9)$$

The lemma follows from (3.3.5), (3.3.6) and (3.3.9). ■

THEOREM 3.6 : *The direct Lagrange-Galerkin method with* C^0 *finite elements of degree* k, $k \geqslant 1$, *is convergent in the* $\ell^\infty(0, T ; (L^2)^d)$ *norm with error* $0(h)$ *when the integrals are approximated using area-weighting, provided that the function* $\underline{a}$ *belongs to the space* $L^\infty(0, T ; (W^{1,\infty})^d)$, $u_0 \in (W^{1,\infty} \cap H^{k+1})^d$ *and the corresponding solution of problem* (1.1.1) *belongs to* $H^1(0, T ; (W^{1,\infty} \cap H^{k+1})^d)$.

Proof: By the triangle inequality,

$$\left\| U_{aw}^{n+1} - u^{n+1} \right\| \leq \left\| U_{aw}^{n+1} - U^{n+1} \right\| + \left\| U^{n+1} - u^{n+1} \right\| .$$

From Theorem 1.1 we know that the exactly integrated direct method converges at a rate $0(h^k)$, where k is the order of the element being used. It now remains to put a bound on $\left\| U_{aw}^{n+1} - U^{n+1} \right\|$.

Recalling the definitions (2.2.1) and (3.3.1) of the exactly integrated and area-weighted direct methods, forming their difference, multiplying by the i-th component of $U^{n+1} - U_{aw}^{n+1}$ and summing over i gives, after applying the Cauchy-Schwarz inequality,

$$\begin{aligned} \left\| U^{n+1} - U_{aw}^{n+1} \right\| &\leq \left\| U^n(\underline{x}) - U_{aw}^n(\underline{x}_h) \right\| \\ &\leq \left\| U^n(\underline{x}) - U^n(\underline{x}_h) \right\| + \left\| U^n(\underline{x}_h) - U_{aw}^n(\underline{x}_h) \right\| \\ &\leq Ch\,\Delta t + \left\| U^n(\underline{x}_h) - U_{aw}^n(\underline{x}_h) \right\| , \end{aligned}$$

by Lemma 3.5. Using the same result from the proof of stability that was used to change variables in Lemma 3.5 we can write

$$\left\| U^n(\underline{x}_h) - U_{aw}^n(\underline{x}_h) \right\| \leq (1 + C\,\Delta t) \left\| U^n(\underline{x}) - U_{aw}^n(\underline{x}) \right\| .$$

Another change of variables can now be made to obtain

$$\left\| U^{n+1} - U_{aw}^{n+1} \right\| \leq Ch\,\Delta t + (1 + C\,\Delta t) \left\| U^n - U_{aw}^n \right\| . \qquad (3.3.10)$$

Assuming that both methods are started with the same initial data, i.e. $U^0 = U_{aw}^0$, then successive substitution into (3.3.10) gives

$$\left\| U^{n+1} - U_{aw}^{n+1} \right\| \leq Ch\, e^{CT} ,$$

and hence the result. ■

Remark : Numerical evidence indicates that the area-weighted Lagrange-Galerkin method for one-dimensional linear advection, with $\underline{a}$ smooth and piecewise linear elements on a smoothly varying mesh, is second order accurate.

4. NUMERICAL EXPERIMENTS FOR THE LAGRANGE-GALERKIN METHOD WITH NON-EXACT INTEGRATION

4.1. Numerical testing of the method with quadrature

It was shown in Section 1 that the Lagrange-Galerkin schemes with exact integration are unconditionally stable, but in Section 2 that by approximating these methods using standard quadrature rules this property is lost. Here

we show that, for a small number of time-steps, despite the predicted instability of a scheme, accurate results may be obtained. When the number of time-steps is increased, instability manifests itself, with this occurring sooner the fewer the number of quadrature points.

To demonstrate these effects we consider the model problem of a cone in a fixed, rotating velocity field governed by the equation :

$$u_t + 2\pi(-y, x) \cdot \underline{\nabla} u = 0 \tag{4.1.1}$$

on the domain $\Omega = (-1, 1) \times (-1, 1)$, with periodic boundary conditions. The initial data consists of a $\cos^2$ cone centred at $(-1/2, 0)$ and of radius 1/4. If $r^2 = (x + 1/2)^2 + y^2$ then

$$u = \begin{cases} \cos^2 2\pi r & \text{for} \quad r \leqslant 1/4 \\ 0 & \text{otherwise} . \end{cases} \tag{4.1.2}$$

This initial data on a 40×40 square mesh is shown in Figure 4.1 and is interpolated rather than L^2 fitted.

Although the theoretical results obtained earlier are strictly speaking only for bilinear elements on rectangles, no qualitative difference has been found between the behaviour of such schemes on a 40×40 grid with a tensor product Gaussian quadrature and the behaviour of linear elements on triangles with various quadrature rules. Also, for this test problem, there was little qualitative difference between the behaviour of the weak and the direct method. Indeed, although in the following we may just give results for one of the methods on one of the meshes, similar results can be obtained with the other method or on the other mesh.

The results in Table 4.1 have been obtained on a $32 \times 32 \times 2$ triangular grid after 25 time-steps with $\Delta t = 0.02$, i.e. after one half-rotation.

The results confirm that the weak formulation conserves $\int U \, d\underline{x}$ more accurately, although the direct method is, in the cases shown here, more accurate in the ℓ_2^2 norm. However, when the vertex scheme is taken to five revolutions (250 time steps) $\int U \, d\underline{x}$ is still 15.32 for the weak method and the maximum shrunk to 0.123, while for the direct method $\int U \, d\underline{x}$ is 17.6 and the maximum has shrunk to 0.005. Without considerable element subdivision seven Gauss points are needed for results which reliably reflect the accuracy of the exactly integrated scheme, with the weak and direct formulations agreeing well.

The CFL number, with these standard grids, is almost one at the centre of the cone. The Fourier analysis predicts instability for the Lagrange-Galerkin

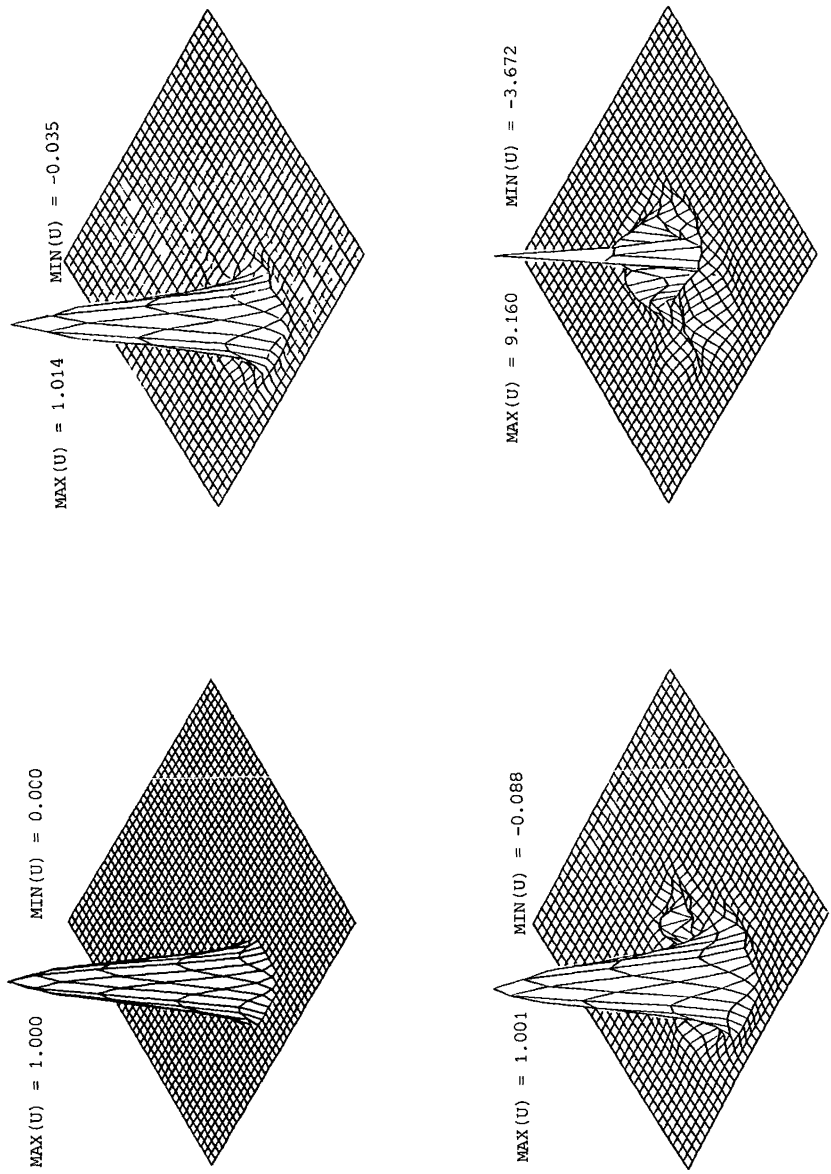

Figure 4.1. — Rotating cone problem with three interior point quadrature on triangles after 0, 1, 5 and 20 revolutions.

TABLE 4.1

Quadrature	Formulation	$\int U\, d\underline{x}$	Max	Min	ℓ_2^2 error
Vertex	Weak	14.970	0.422	0	1.6
Vertex	Direct	14.980	0.424	0	1.6
Centroid	Weak	14.965	0.611	− 0.014	1.2
Centroid	Direct	14.938	0.608	− 0.007	1.1
Mid edge	Weak	14.967	0.930	− 0.047	1.6×10^{-1}
Mid edge	Direct	14.919	0.933	− 0.043	7.9×10^{-2}
3 interior	Weak	14.961	1.034	− 0.025	4.8×10^{-2}
3 interior	Direct	15.006	1.010	− 0.024	1.5×10^{-2}
7 Gauss	Weak	14.964	0.996	− 0.014	2.2×10^{-2}
7 Gauss	Direct	14.965	0.991	− 0.013	2.0×10^{-2}

method with small CFL numbers, which for this problem occur at the centre of the region irrespective of the choice of Δt, and hence the region that will give us difficulties is well away from the cone. Figure 4.1 shows the direct method on the triangular grid with linear elements, evaluated using a three interior-point quadrature, after 1, 5 and 20 complete revolutions. Clearly, after 20 revolutions, or 1 000 time-steps, we see that the calculated solution is completely dominated by the oscillations at the centre, the region of low CFL numbers. A similar situation occurs for the weak method on the 40×40 bilinear grid evaluated with 2×2 Gaussian quadrature.

With the 4×4 Gaussian quadrature on rectangles, and seven point quadrature for the triangles, instability is difficult to demonstrate because for most CFL numbers they are extremely accurate. However for a similar one-dimensional linear advection problem with $\Delta x = 0.05$ and $\Delta t = 0.003$, the weak method with 4 point Gauss quadrature was unstable after 1 600 time-steps.

Figure 4.2 shows the solution of the same problem as in Figure 4.1 calculated by Lobatto quadrature after 2 and 20 revolutions. In contrast with Lagrange-Galerkin methods based on Gaussian quadrature, this scheme can be stabilized by choosing a sufficiently small time-step.

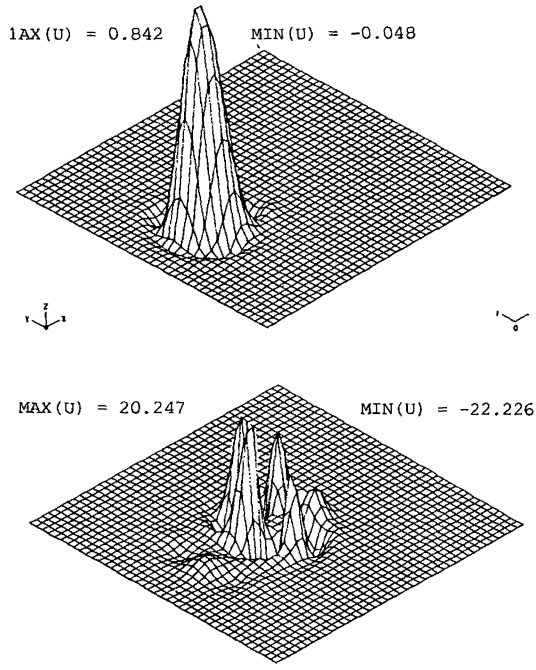

Figure 4.2. — Rotating cone problem with Lobatto quadrature on rectangles after 2 and 20 revolutions. The CFL number at the peak of the cone is 1.265.

4.2. Numerical testing of the method with area-weighting

The area-weighted Lagrange-Galerkin method has also been tested on the rotating cone problem described above. We use the 40×40 rectangular mesh with bilinear elements, $\Delta t = 0.02$ and perform 25 time steps, as for Table 4.1. The results are given in Table 4.2 where they are compared with 4×4 Gaussian quadrature : timings are for a VAX 11/785. This demonstra-

TABLE 4.2

Method	Max	Min	ℓ_2^2 error	CPU secs.
Weak, 4×4 Gauss	0.991	− 0.009	4.64×10^{-3}	419
Direct, 4×4 Gauss	0.993	− 0.008	4.47×10^{-3}	440
Weak, Area Weighting	0.987	− 0.008	6.17×10^{-3}	159

tes that for comparable accuracy the area-weighted Lagrange-Galerkin method is significantly faster. Greater accuracy can be achieved by using area-weighting with subdivided elements.

To demonstrate stability, the area-weighted scheme was run for 500 time-steps (10 revolutions) for the direct method and 1 000 time-steps (20 revolutions) for both the direct and the weak methods. The solutions are given in Figure 4.3. In this quite severe test the methods maintain their

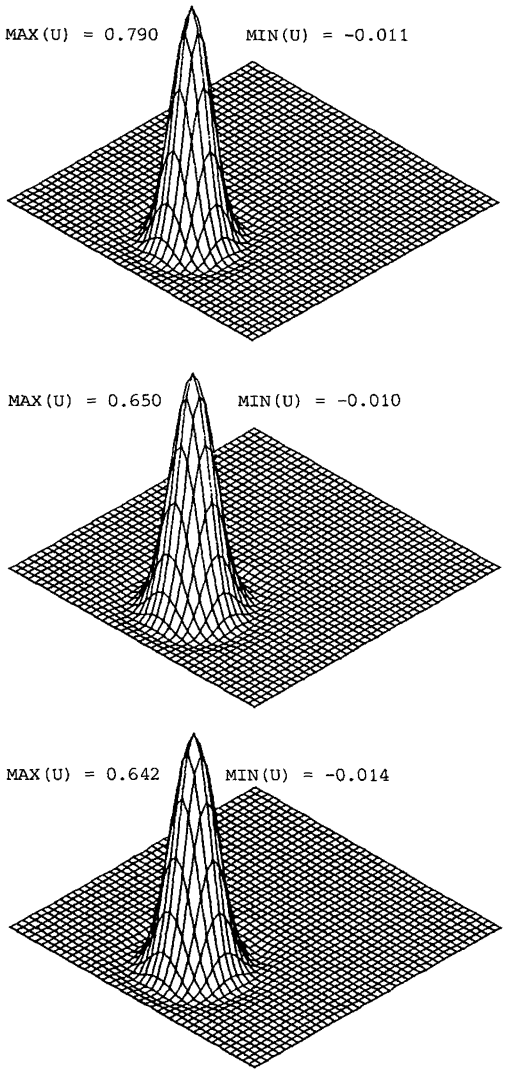

Figure 4.3. — Rotating cone problem with direct area-weighting after 10 and 20 revolutions and weak area-weighting after 20 revolutions. The CFL number at the peak of the cone is 1.265.

accuracy and show no inclination to become unstable. As with the quadrature experiments of subsection 4.1 there is no qualitative difference between the weak and the direct method.

5. CONCLUSIONS

In practice the inner products peculiar to the Lagrange-Galerkin method must be approximated by using non-exact integration. This can be performed in two ways : the first, and most natural, is to approximate the integral by numerical quadrature and evaluate the trajectories of the quadrature points exactly (to all intents and purposes) and hence calculate the integrand exactly at these points ; the second option is to approximate the trajectories of all points so that the integrals may be performed exactly.

It has been shown here that with Gauss-Legendre, Gauss-Lobatto, Radau and Newton-Cotes quadrature formulae the unconditional stability of the exactly integrated Lagrange-Galerkin method is lost. Lumping the mass-matrix stabilizes the scheme in all these cases but the resulting method is so diffusive as to make the expensive calculation of the right-hand side unjustifiable. One stable implementation of quadrature that has been proposed is to use a compound trapezium rule for both sides of the equation, but again this introduces too much diffusion unless many points are used, making it very expensive.

This leads us to the second of the two options which, with area-weighting, yields an efficient, stable and convergent scheme with the same order of convergence as the exactly integrated method, for both linear and constant elements.

As a conclusion then, the Lagrange-Galerkin methods have shown themselves to be a very powerful class of methods although subject to stability problems with non-exact integration. The technique developed here has been proved theoretically and shown experimentally to retrieve the desirable properties of the exactly integrated Lagrange-Galerkin method.

REFERENCES

[1] M. ABRAMOWITZ & I. A. STEGUN, *Handbook of Mathematical Functions.* Dover Publications Inc., New York, 1965.

[2] J. P. BENQUÉ, G. LABADIE & J. RONAT, *A new finite element method for the Navier-Stokes equations coupled with a temperature equation.* Proc. 4th Int. Symp. on Finite Element Methods in Flow Problems (Ed. T. Kawai), North-Holland, Amsterdam, Oxford, New York, 1982, pp. 295-301.

[3] M. BERCOVIER & O. PIRONNEAU, *Characteristics and the finite element method.* Proc. 4th Int. Symp. on Finite Element Methods in Flow Problems (Ed. T. Kawai), North-Holland, Amsterdam, Oxford, New York, 1982, pp. 67-73.

[4] P. N. CHILDS & K. W. MORTON, *Characteristic Galerkin methods for scalar conservation laws in one dimension.* Oxford University Computing Laboratory Report No. 86/5, 1986. To appear in SIAM J. Numerical Analysis.

[5] A. J. CHORIN & J. E. MARSDEN, *A Mathematical Introduction to Fluid Mechanics (Universitext).* Springer-Verlag, New York, Berlin, Heidelberg, Tokyo, 1984.

[6] J. DOUGLAS Jr & T. F. RUSSELL, *Numerical methods for convection-dominated diffusion problems based on combining the method of characteristics with finite element or finite difference procedures.* SIAM J. Numer. Anal., 19 (1982), pp. 871-885.

[7] J. W. EASTWOOD, Private communication.

[8] F. H. HARLOW, *The particle in cell computing method for fluid dynamics.* Methods in Computational Physics (Ed. B. Adler, S. Fernbach & M. Rotenberg), Vol. 3, Academic Press, New York, 1964.

[9] R. W. HOCKNEY & J. W. EASTWOOD, *Computer Simulation Using Particles.* McGraw-Hill, New York, 1981.

[10] Z. KOPAL, *Numerical Analysis.* Chapman & Hall Ltd., London, 1961.

[11] I. V. KRYLOV, *Approximate Calculation of Integrals.* Mac Millan, New York, 1962.

[12] P. LESAINT, *Numerical solution of the equation of continuity.* Topics in Numerical Analysis III (Ed. J. J. H. Miller), Academic Press, London, New York, San Francisco, 1977, pp. 199-222.

[13] K. W. MORTON & A. PRIESTLEY, *On characteristic and Lagrange-Galerkin methods.* Pitman Research Notes in Mathematics Series (Ed. D. F. Griffiths & G. A. Watson), Longman Scientific and Technical, Harlow, 1986.

[14] K. W. MORTON & P. SWEBY, *A comparison of flux limited difference methods and characteristic Galerkin methods for shock modelling.* To appear in J. Comput. Phys.

[15] O. PIRONNEAU, *On the transport diffusion algorithm and its application to the Navier-Stokes equations,* Numer. Math., 38 (1982), pp. 309-332.

[16] T. F. RUSSELL, *Time stepping along characteristics with incomplete iteration for a Galerkin approximation of miscible displacement in porous media.* Ph. D. Thesis, University of Chicago, 1980.

[17] E. SÜLI, *Convergence and nonlinear stability of the Lagrange-Galerkin method for the Navier-Stokes equations,* Numer. Math., 53 (1988), pp. 459-483.