MOVING FINITE ELEMENTS: REGULARISATION TECHNIQUES

M.J. Baines

Dept. of Mathematics
University of Reading

Numerical Analysis Report 19/86

Keywords: Moving Finite Elements, Regularisation.

The work reported here forms part of the research programme of the Institute for Computational Fluid Dynamics at Oxford and Reading.

C O N T E N T S

§1 Introduction

In recent years progress in the use of Moving Finite Elements, as introduced by Miller [1], has taken place along different paths. In Berkeley, Miller and his collaborators have seen penalty functions and implicit methods as integral parts of the method, thus avoiding singularity at the expense of stiffness. In Reading a number of workers have pursued an explicit approach to the problem, avoiding penalty functions and making explicit provision for singularities. In general the Miller approach has been able to solve harder problems, albeit at more expense, while the Reading approach has produced very fast but perhaps less robust algorithms.

Penalty functions, or regularisation, are implemented in the approach of Miller & Carlson [2] element by element in an ingenious way. Indeed the MFE matrix itself is constructed in the same way, as summarised in §2 below. There is a relation here with the MFE-plus-constraints approach of Baines [3]. The analysis reveals a 5 matrix decomposition of the (unregularised) MFE matrix, which shows the link between the assemblies of Miller & Carlson [2] and Wathen & Baines [4].

Some kind of special procedure is needed to avoid singularities in the MFE matrix. If nodes are collinear or coplanar the MFE matrix becomes singular, but the root cause of the trouble lies in the representation by linear elements. For very smooth solutions, representation by a piecewise linear function with free nodes automatically overspecifies the solution. Some restriction is needed in the specification (e.g. constrained node movement) to realise a well-conditioned problem. Miller's penalty functions do precisely this in a continuous manner, their invention depending on somewhat

mechanical analogies [1]. The Reading group on the other hand has opted for a cut-off below which a reduced problem is solved [4].

Again, node overtaking destroys the basis of the MFE equations. This is avoided using "spring function" penalties by Miller and via preconditioning and restricting the time step by the Reading group.

In this report we analyse the need for regularisation (in §3) and discuss possible regularisations in relation to the fast methods originating at Reading [5]. The aim is to invent a regularisation technique which can be used with explicit time stepping, whose MFE matrix, when approximately preconditioned [6], is well behaved. This is done by regularising with the "tangential" velocity, which produces a regularising term which is in some sense orthogonal to the unregularised MFE matrix. Moreover, it has a decomposition which allows the good conditioning of the MFE matrix to extend to the singular case. This is discussed in §4. By penalising variations in the tangential velocity from the previous time step (i.e. the tangential acceleration), a smoother more robust solution is obtained from the explicit method. This is described in §5. The extension of the analysis to two dimensions is given in §6, where the results of the full analysis are replaced by a simplified regularisation which takes account of matrix properties rather than by following rigorous regularisation via the tangential velocities.

Section 7 refers briefly to scaling matters and to a local formulation. In §8 we summarise the results and discuss a further regularisation associated with small node spacing.

I particularly wish to thank Colin Please for many valuable

discussions. Thanks are also due to Neil Carlson and Andy Wathen for illuminating comments on assembly and to Robin Dixon for programming assistance.

There are two Appendixes. The first substitutes tangential regularisation by "horizontal" regularisation, i.e. regularisation of the nodal speeds, on the grounds that this gives a simple direct attack on the problem of near-indeterminacy on node overtaking. The second describes a family of conservative moving element methods using the null space of the so-called β equation, gives a geometric construction in the case of the local method and points out that all methods described in this report (and some others) fall into this framework.

§2 Element Minimisation

In the Miller & Carlson approach an approximate solution of the equation

$$u_t = \mathcal{L}(u) \tag{2.1}$$

is sought by minimising the L_2 norm

$$\| v_t - \mathcal{L}(v) \|_2 \tag{2.2}$$

in the space of moving finite element approximations (in one dimension)

$$v = \sum_j a_j \alpha_j \qquad v_t = \sum_j \dot{a}_j \alpha_j + \dot{s}_j \beta_j \tag{2.3}$$

(see fig. 1 and refs. [1], [4]), as follows. In each element k write

$$u_t = \dot{a}_{k1}\phi_{k1} - m_k \dot{s}_{k1}\phi_{k1} + \dot{a}_{k2}\phi_{k2} - m_k \dot{s}_{k2}\phi_{k2} \tag{2.4}$$

(see fig. 2), as if the nodal speeds at nodes common to two elements were independent (i.e. ignoring continuity of $\dot{a}$, $\dot{s}$). Then, in each element k, the square of the L_2 norm (2.2) may be written (apart from a term independent of $\underline{\dot{Y}}_k$) as

$$\underline{\dot{Y}}_k \, E_k \, \underline{\dot{Y}}_k - 2\underline{\dot{Y}}_k^T \underline{G}_k \tag{2.5}$$

$$\text{where} \quad \underline{\dot{y}}_k = \begin{bmatrix} \dot{a}_{k1} \\ \dot{s}_{k1} \\ \dot{a}_{k2} \\ \dot{s}_{k2} \end{bmatrix}, \quad E_k = \frac{1}{6} \Delta S_k \begin{bmatrix} 2\,\underline{m}\,\underline{m}^T & \underline{m}\,\underline{m}^T \\ \underline{m}\,\underline{m}^T & 2\,\underline{m}\,\underline{m}^T \end{bmatrix}, \quad \underline{m}^T = [1 - m_k] \tag{2.6}$$

$$\text{and} \quad \underline{G}_k^T = \{<\phi_{k1}, \mathcal{L}(u)>, -m_k<\phi_{k1}, \mathcal{L}(n)>, <\phi_{k2}, \mathcal{L}(u)>, -m_k<\phi_{k2}, \mathcal{L}(u)>\} \tag{2.7}$$

with $\Delta s_k = s_{k+1} - s_k$. The sum of the element contributions (2.5) to the norm (2,2) is

$$\sum_k (\underline{\dot{y}}_k E_k \underline{\dot{y}}_k - 2\underline{\dot{y}}_k \underline{G}_k). \tag{2.8}$$

Miniminisation of the square of the element norm over $\dot{a}_{k1}$, $\dot{s}_{k1}$, $\dot{a}_{k2}$, $\dot{s}_{k2}$ leads to the (singular) system

$$E_k \underline{\dot{y}}_k = \underline{G}_k \tag{2.9}$$

for each k. If, however, we minimise the total L_2 norm, with the constraints

$$\dot{a}_{k1} = \dot{a}_{k-1,2} \quad , \quad \dot{s}_{k1} = \dot{s}_{k-1,2} \quad , \tag{2.10}$$

we obtain the standard MFE system, non-singular in general.

The technique is as described in [3]. Because of the constraints the number of free unknowns is reduced by roughly half. Write

$$\begin{bmatrix} \dot{a}_{k-1,2} \\ \dot{s}_{k-1,2} \\ \dot{a}_{k1} \\ \dot{s}_{k1} \end{bmatrix} = R_j \begin{bmatrix} \dot{a}_{k1} \\ \dot{s}_{k1} \end{bmatrix} \quad \text{where} \quad R_j = \begin{bmatrix} I_2 \\ I_2 \end{bmatrix}, \; I_2 = \begin{bmatrix} 10 \\ 01 \end{bmatrix} \tag{2.11}$$

With the usual definitions $\underline{\dot{y}} = \{\ldots, \dot{a}_j, \dot{s}_j, \ldots\}^T$ and

$\underline{\dot{Y}}^T = \left\{ \underline{\dot{Y}}_j^T \right\}$, we may write

$$\underline{\dot{Y}} = R\, \underline{\dot{y}} \tag{2.12}$$

where

$$R^T = \begin{bmatrix} \ddots & & & \\ & I_2 I_2 & 0 & 0 \\ & & & \\ & 0 & 0 & I_2 I_2 \\ & & & \ddots \end{bmatrix} \tag{2.13}$$

Then the total L_2 norm (2.8) (squared) can be written (apart from a term independent of $\underline{\dot{y}}$) as

$$\underline{\dot{y}}^T R^T E R \underline{\dot{y}} - 2\, \underline{\dot{y}}^T \underline{g} \tag{2.14}$$

where $E = \text{diag}\ \{E_k\}$, $\underline{g} = R^T \underline{G}$. Minimisation of (2.14) over $\underline{\dot{y}}$ yields the system

$$R^T E R\, \underline{\dot{y}} = \underline{g}\ . \tag{2.15}$$

The system (2.15) is the standard MFE system and shows a decomposition of the MFE matrix A (see [4], [5]) into

$$A = R^T E R. \tag{2.16}$$

This may be compared with the decomposition described by Wathen & Baines [4],

$$A = M^T C M, \tag{2.17}$$

where
$$C = \text{diag}\{C_k\}, \quad M = \text{diag}\{M_j\}, \tag{2.18}$$

$$C_k = \begin{bmatrix} \langle \phi_{k1}, \phi_{k1} \rangle & \langle \phi_{k1}, \phi_{k2} \rangle \\ \langle \phi_{k2}, \phi_{k1} \rangle & \langle \phi_{k2} \phi_{k2} \rangle \end{bmatrix}, \quad M_j = \begin{bmatrix} 1 & - & m_{j-\frac{1}{2}} \\ 1 & - & m_{j+\frac{1}{2}} \end{bmatrix}. \tag{2.19}$$

Since it is possible to decompose E_k further into

$$E_k = \tilde{M}_k^T C_k \tilde{M}_k \tag{2.20}$$

where
$$\tilde{M}_k = \begin{bmatrix} 1 & -m_k & 0 & 0 \\ 0 & 0 & 1 & -m_k \end{bmatrix}, \tag{2.21}$$

we may write

$$E = \tilde{M}^T C \tilde{M} \tag{2.22}$$

where $\tilde{M} = \text{diag } \{\tilde{M}_k\}$. This yields a five matrix decomposition of A, namely,

$$A = R^T \tilde{M}^T C \tilde{M} R. \tag{2.23}$$

Coupling $\tilde{M}R$ gives M and hence (2.17). Coupling $\tilde{M}^T C \tilde{M}$ gives E and hence (2.16). Note that M is node based but $\tilde{M}$ is element based. The decomposition (2.16) allows the slope m_k of the solution to be brought into the equations elementwise, unlike in the decomposition (2.17). Moreover regularisation terms may also be introduced elementwise, as we shall see below.

§3 Need for Regularisation

To exhibit the need for regularisation we transform to a rotated system of coordinates (see fig. 3) which approximately aligns with the tangential and normal directions of the underlying solution curve at a node. Let m_j be an estimate of the slope of the tangent to the solution at a node j, for example the average of the slopes in adjacent elements. Then

$$u_t = \sum_j \dot{a}_j \alpha_j + \dot{s}_j \beta_j = \sum_j \dot{N}_j \bar{\alpha}_j + \dot{T}_j \bar{\beta}_j \tag{3.1}$$

(fig. 3), where

$$\left.\begin{aligned} \bar{\alpha}_j &= \alpha_j \cos\theta_j - \beta_j \sin\theta_j \\ \bar{\beta}_j &= \alpha_j \sin\theta_j + \beta_j \cos\theta_j \end{aligned}\right\}, \tag{3.2}$$

$\tan\ \theta_j = m_j$. The resulting form of the L_2 norm (2.2) is (apart from a constant term)

$$\dot{\underline{Y}}^T\ \bar{M}^T\ C\bar{M}^T\ \dot{\underline{Y}}\ -\ 2\dot{\underline{Y}}^T\ \bar{\underline{g}} \tag{3.3}$$

where $\dot{\underline{Y}} = \{\ldots, \dot{N}_j, \dot{T}_j, \ldots\}$,

$$\left.\begin{aligned} \bar{M} = \text{diag}\ \{\bar{M}_j\}, \quad \bar{M}_j &= \begin{bmatrix} \mu_{j-\frac{1}{2}} & \nu_{j-\frac{1}{2}} \\ \mu_{j+\frac{1}{2}} & \nu_{j+\frac{1}{2}} \end{bmatrix}, \\ \bar{\underline{g}} = \{\bar{\underline{g}}_j\}\ , \quad \bar{\underline{g}}_j &= \begin{bmatrix} <\bar{\alpha}_j\ ,\ \mathcal{L}(u)> \\ <\bar{\beta}_j\ ,\ \mathcal{L}(u)> \end{bmatrix}, \end{aligned}\right\} \tag{3.4}$$

and C is as in (2.19). The entries in $\bar{M}_j$ are given by

$$\left.\begin{aligned} \mu_{j\mp\frac{1}{2}} &= \cos\theta_j + m_{j\mp\frac{1}{2}}\ \sin\theta_j \\ \nu_{j\mp\frac{1}{2}} &= \sin\theta_j - m_{j\mp\frac{1}{2}}\ \cos\theta_j \end{aligned}\right\} \tag{3.5}$$

which correspond to $\bar{\alpha}_j$ and $\bar{\beta}_j$, respectively, in the sense that

$$\bar{\alpha}_j = \mu_j\ \alpha_j\ , \quad \bar{\beta}_j = \nu_j\ \alpha_j\ . \tag{3.6}$$

The corresponding MFE system is

$$\bar{M}^T \bar{C}\bar{M} \dot{\underline{y}} = \underline{\bar{g}} \ . \tag{3.7}$$

It is clear that if $m_{j-\frac{1}{2}} \cong m_{j+\frac{1}{2}} \cong m_j$ (near to parallelism), then $\nu_{j+\frac{1}{2}}$ are both nearly zero and M_j has an almost zero second column. As a result the value of $\dot{T}_j$ obtained by minimising (3.3) is arbitrary. Computationally, a large spurous value of $\dot{T}_j$ may be produced.

This may also be seen in the elementwise formulation of Miller and Carlson [2], prior to assembly and without regularisation. In each element k the contribution to u_t is

$$\mu_{k1} \dot{N}_{k1} \phi_{k1} + \nu_{k1} \dot{T}_{k1} \phi_{k1} + \mu_{k2} \dot{N}_{k2} \phi_{k2} + \nu_{k2} \dot{T}_{k2} \phi_{k2} \tag{3.8}$$

(c.f. (2.4)), and the square of the L_2 norm (2.2) is (apart from a constant term)

$$\dot{\underline{y}}_k \bar{E}_k \dot{\underline{y}}_k - 2 \dot{\underline{y}}^T \underline{\bar{G}}_k \tag{3.9}$$

where $\dot{\underline{y}}_k = \begin{bmatrix} \dot{N}_{k1} \\ \dot{T}_{k1} \\ \dot{N}_{k2} \\ \dot{T}_{k2} \end{bmatrix}$, $\bar{E}_k = \frac{1}{6} \Delta s_k \begin{bmatrix} 2\mu_{k1}^2 & 2\mu_{k1}\nu_{k1} & \mu_{k1}\mu_{k2} & \mu_{k1}\nu_{k2} \\ 2\nu_{k1}\mu_{k1} & 2\nu_{k1}^2 & \nu_{k1}\mu_{k2} & \nu_{k1}\nu_{k2} \\ \mu_{k2}\mu_{k1} & \mu_{k2}\nu_{k1} & 2\mu_{k2}^2 & 2\mu_{k2}\nu_{k2} \\ \nu_{k2}\mu_{k1} & \nu_{k2}\mu_{k1} & 2\nu_{k2}\mu_{k2} & 2\nu_{k2}^2 \end{bmatrix}$ (3.10)

and

$$\underline{\bar{G}}_k^T = \{\mu_{k1}\langle\phi_{k1}, \mathcal{L}(u)\rangle, \nu_{k1}\langle\phi_{k1}, \mathcal{L}(u)\rangle, \mu_{k2}\langle\phi_{k2}, \mathcal{L}(u)\rangle, \nu_{k2}\langle\phi_{k2}, \mathcal{L}(u)\rangle\} \tag{3.11}$$

(c.f.(2.6)). Near to parallelism $\nu_{k1} \cong \nu_{k2} \cong 0$ and all coefficients multiplying $\dot{T}_{k1}$, $\dot{T}_{k2}$ are very small, allowing arbitrary $\dot{T}$'s as before. The situation is unaltered by constraints, since it occurs at each end of each element.

§4 Regularisation by the Tangential Velocity

To avoid the ill-conditioning described in §3 it is proposed to add regularisation terms. The terms needed in the regularisation are those which add into the coefficients of the $\dot{T}$'s to prevent such coefficients going to zero in the event of parallelism. We therefore add $\dot{T}^2$ terms to the quadratic form coming from (2.2), following Miller [1]. We describe this in the elementwise formulation of Miller and Carlson [2].

Consider first addition of the regularisation

$$\varepsilon^2 \, \| V_T \|^2 , \tag{4.1}$$

where V_T is the (notional) velocity of points along the tangent to the solution curve. In the finite element representation we may approximate this by the piecewise linear function

$$V_T = \sum_j \dot{T}_j \, \alpha_j , \tag{4.2}$$

and then the elementwise quadratic form (3.9) is augmented to become

$$\underline{\dot{y}}^T \, \underline{\underline{E}}_k \, \underline{\dot{y}}_k - 2\underline{\dot{y}}^T \, \underline{\underline{G}} + \varepsilon^2 \, \underline{\dot{y}}_k^T \, F_k \, \underline{\dot{y}}_k \tag{4.3}$$

where
$$F_k = \frac{1}{6}\,\Delta s_k \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 2 \end{bmatrix}. \tag{4.4}$$

In the case of parallelism the coefficients of $\dot{T}$ do not vanish, and the $\dot{T}$'s are now determined by the minimisation. However, for this regularisation, since $\bar{\underline{G}}$ is unaffected by the minimisation, the $\dot{T}$'s come out to be zero. This point is taken up again in §5.

We note that in the non-parallel case the quadratic form (4.3), when constrained as in §2, becomes

$$\dot{\underline{y}}^T \left(R^T \bar{E} R + \varepsilon^2 R^T F R\right) \dot{\underline{y}} - 2\, \dot{\underline{y}}\, \bar{\underline{g}} \tag{4.5}$$

where

$$\bar{E} = \text{diag}\,\{\bar{E}_k\}\ , \quad F = \text{diag}\,\{F_k\}\ . \tag{4.6}$$

This may be rewritten as

$$\dot{\underline{y}}\, R^T \left(\bar{E} + \varepsilon^2 F\right) R\, \dot{\underline{y}} - 2\dot{\underline{y}}\, \bar{\underline{g}}\ . \tag{4.7}$$

Now recall that we may decompose $\bar{\underline{E}}_k$ (c.f. (2.20)) into

$$\bar{E}_k = \tilde{\underline{M}}_k^T\, C_k\, \tilde{\underline{M}}, \tag{4.8}$$

where $$\tilde{\tilde{M}}_k = \begin{bmatrix} \mu_{k1} & \nu_{k1} & 0 & 0 \\ 0 & 0 & \mu_{k1} & \nu_{k2} \end{bmatrix}, \tag{4.9}$$

while F_k trivially decomposes into

$$F_k = \tilde{\tilde{W}}_k^T \, C_k \, \tilde{\tilde{W}}_k , \tag{4.10}$$

where $$\tilde{\tilde{W}}_k = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} . \tag{4.11}$$

Then (see (4.7))

$$R^T(\bar{E} + \varepsilon^2 F)R = R^T\tilde{\tilde{M}}^T C\tilde{\tilde{M}}R + \varepsilon^2 R^T\tilde{\tilde{W}}^T C\tilde{\tilde{W}}R, \tag{4.12}$$

where $$\tilde{\tilde{M}} = \text{diag}\,\{\tilde{\tilde{M}}_k\} , \quad \tilde{\tilde{W}} = \text{diag}\,\{\tilde{\tilde{W}}_k\} . \tag{4.13}$$

Using the result $\tilde{\tilde{M}}R = \bar{M}$ and putting

$$\bar{W} = WR = \begin{bmatrix} \ddots & & & \\ & \begin{matrix} 01 \\ 01 \end{matrix} & & \\ & & \begin{matrix} 01 \\ 01 \end{matrix} & \\ & & & \ddots \end{bmatrix} \tag{4.14}$$

we obtain a new form for (4.12), namely

$$R^T (\bar{E} + \varepsilon^2 F) R = \bar{M}^T C \bar{M} + \varepsilon^2 \bar{W}^T C \bar{W} , \tag{4.15}$$

and (4.7) becomes

$$\dot{\underline{y}}^T (\bar{M}^T C \bar{M} + \varepsilon^2 \bar{W}^T C \bar{W}) \dot{\underline{y}} - 2 \dot{\underline{y}} \bar{\underline{g}}. \tag{4.16}$$

The corresponding MFE equations are (c.f.(3.7))

$$(\bar{M}^T C \bar{M} + \varepsilon^2 \bar{W}^T C \bar{W}) \dot{\underline{y}} = \bar{\underline{g}} . \tag{4.17}$$

Finally, reverting to the $\dot{a}$, $\dot{s}$ frame, $\bar{M}^T C \bar{M} \dot{\underline{y}}$ becomes $M^T C M \dot{\underline{y}}$ and $\varepsilon^2 \bar{W}^T C \bar{W} \dot{\underline{y}}$ becomes

$$\varepsilon^2 W^T C W, \tag{4.18}$$

where $W = \text{diag } \{W_j\}$, $W_j = \begin{bmatrix} \sin\theta_j & \cos\theta_j \\ \sin\theta_j & \cos\theta_j \end{bmatrix} = \cos\theta_j \begin{bmatrix} \mu_j & 1 \\ \mu_j & 1 \end{bmatrix}$. (4.19)

The form of the regularised equations (4.17) then becomes

$$(M^T C M + \varepsilon^2 W^T C W) \dot{\underline{y}} = \underline{g} , \tag{4.20}$$

or

$$(A + \varepsilon^2 \beta) \dot{\underline{y}} = \underline{g} , \tag{4.21}$$

where

$$\beta = W^T C W \qquad \text{(c.f.[4])}. \tag{4.22}$$

In practice, since the term $\dot{\underline{y}}^T \beta \dot{\underline{y}}$ has an effect only in the near parallel case, we can replace the matrix W without loss of generality by the matrix

$$W^* = \text{diag}\{W_j^*\}, \quad W_j^* = \cos\theta_j \begin{bmatrix} \mu_{j-\frac{1}{2}} & 1 \\ \mu_{j+\frac{1}{2}} & 1 \end{bmatrix} \tag{4.23}$$

which has the same form as the W of (4.19) in the near parallel limit.

Equation (4.21) can now be written

$$(A + \varepsilon^2 B^*)\, \dot{\underline{y}} = \underline{g} \,. \tag{4.24}$$

The matrix $B^* = W^{*T} C W^*$ is the mass matrix with inner products

$$\begin{bmatrix} < -\beta_i^*, -\beta_j^* > & < -\beta_i^*, \alpha_j^* > \\ < \alpha_i^*, -\beta_j^* > & < \alpha_i^*, \alpha_j^* > \end{bmatrix} \tag{4.25}$$

where

$$\beta_i^* = \frac{u_x}{\sqrt{1 + u_x^2}}\, \alpha_i \,, \quad \alpha_i^* = \frac{1}{\sqrt{1 + u_x^2}}\, \alpha_i \tag{4.26}$$

(c.f. the gradient weighted MFE method of Miller [7], who uses

$$\frac{\alpha_i}{\sqrt{1 + u_x^2}} \quad , \quad \frac{\beta_i}{\sqrt{1 + u_x^2}}$$

instead of α_i , β_i to generate the unregularised MFE matrix).

The regularisation described in this section contrasts with that of Miller [8] who in recent work regularises with a term proportional to

$$(\dot{T}_{k1} - \dot{T}_{k2})^2 . \tag{4.27}$$

This leads, by an argument similar to that used earlier in this section, to a regularisation term which is proportional to the matrix product (c.f. (4.23))

$$W^{*T} K W^{*} \tag{4.28}$$

where $K = \text{diag}\{K_k\}$, $K_k = \frac{2}{\Delta s_k} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$, (4.29)

the local stiffness matrix. In this way Miller penalises the relative velocity (4.27) of nodes at the ends of the element k.

Note that the matrix B or B^* of the regularising term of (4.21) or (4.23) is an assembly of element mass matrices (contained in C). As pointed out by Wathen [6], the diagonal of B is

$$M^T C_D M + \varepsilon^2 W^T C_C W \tag{4.30}$$

where C_C is the diagonal of C. Moreover, the definiteness (w.r.t. λ) of the quadratic form

$$\dot{\underline{y}}^T [M^T C M + \varepsilon^2 W^T C W - \lambda (M^T C_D M + \varepsilon^2 W^T C_D W)] \dot{\underline{y}} \tag{4.31}$$

$$= \underline{x}^T (C - \lambda C_D) \underline{x} + \varepsilon^2 \underline{z}^T (C - \lambda C_D) \underline{z} \tag{4.32}$$

where $\underline{x} = M\dot{\underline{y}}$, $\underline{z} = W\dot{\underline{y}}$, depends only on the definiteness of the matrix $C - \lambda C_D$. Hence preconditioning of the matrix

$$A + \varepsilon^2 B = M^T C M + \varepsilon^2 W^T C W \tag{4.33}$$

by the inverse of (4.31), a diagonal matrix, clusters the eigenvalues, in the same way as for $A = M^T C M$ (in the non-parallel case).

The argument above can be used to indicate the form of the regularising parameter ε^2. As pointed out by Miller [8], Wathen's result [6] essentially says that the modified Rayleigh Quotient

$$\frac{\dot{\underline{y}}^T (M^T C M + \varepsilon^2 W^T C W) \dot{\underline{y}}}{\dot{\underline{y}}^T (M^T C_D M + \varepsilon^2 W^T C_D W) \dot{\underline{y}}} \tag{4.34}$$

lies within tight bounds ($\frac{1}{2}$ and $\frac{3}{2}$ in the one-dimensional case). It follows that the ill-conditioning inherent in the MFE matrix may be studied by considering only the quadratic form

$$\dot{\underline{y}}^T \left(M^T C_D M + \varepsilon^2 W^T C_D W\right) \dot{\underline{y}} \tag{4.35}$$

or, equivalently,

$$\dot{\bar{\underline{y}}}^T \left(\bar{M} C_D \bar{M} + \varepsilon^2 \bar{W} C_D \bar{W}\right) \dot{\bar{\underline{y}}} \tag{4.36}$$

(c.f. (4.16)). But since C_D is diagonal this quadratic form is a sum over j of purely nodal contributions

$$\dot{\bar{\underline{y}}}_j \left\{\bar{M}_j \frac{1}{6} \begin{bmatrix} \Delta s_{j-\frac{1}{2}} & 0 \\ 0 & \Delta s_{j+\frac{1}{2}} \end{bmatrix} \bar{M}_j + \varepsilon^2 \bar{W}_j \frac{1}{6} \begin{bmatrix} \Delta s_{j-\frac{1}{2}} & 0 \\ 0 & \Delta s_{j+\frac{1}{2}} \end{bmatrix} \bar{W}_j \right\} \dot{\bar{\underline{y}}}_j \tag{4.37}$$

The matrix of this quadratic form is (see (3.4), (4.14))

$$\frac{1}{6} \begin{bmatrix} \mu_{j-\frac{1}{2}}^2 \Delta s_{j-\frac{1}{2}} + \mu_{j+\frac{1}{2}}^2 \Delta s_{j+\frac{1}{2}} & \mu_{j-\frac{1}{2}} \nu_{j-\frac{1}{2}} \Delta s_{j-\frac{1}{2}} + \mu_{j+\frac{1}{2}} \nu_{j+\frac{1}{2}} \Delta s_{j+\frac{1}{2}} \\ \nu_{j-\frac{1}{2}} \mu_{j-\frac{1}{2}} \Delta s_{j-\frac{1}{2}} + \nu_{j+\frac{1}{2}} \mu_{j+\frac{1}{2}} \Delta s_{j+\frac{1}{2}} & \nu_{j-\frac{1}{2}}^2 \Delta s_{j-\frac{1}{2}} + \nu_{j+\frac{1}{2}}^2 \Delta s_{j+\frac{1}{2}} + \varepsilon^{2\prime} (\Delta s_{j-\frac{1}{2}} + \Delta s_{j+\frac{1}{2}}) \end{bmatrix} \tag{4.38}$$

Close to parallelism this reduces to

$$\frac{1}{6} \begin{bmatrix} (\cos \theta_j + m_j \sin \theta_j)^2 \left(\Delta s_{j-\frac{1}{2}} + \Delta s_{j+\frac{1}{2}}\right) & 0 \\ 0 & \varepsilon^2 \left(\Delta s_{j-\frac{1}{2}} + \Delta s_{j+\frac{1}{2}}\right) \end{bmatrix} \tag{4.39}$$

and the elements of this matrix are of the same order if we choose

$$\varepsilon^2 = \varepsilon_j^2 \sim (\cos\theta_j + m_j \sin\theta_j)^2 = \sec^2\theta_j \tag{4.40}$$

This ensures a well-behaved minimisation process.

There is no difficulty in allowing ε^2 to be variable in this approach. Indeed the result (4.40) may be conveniently implemented by deleting the $\cos\theta_j$ in (4.19) or (4.23) and replacing ε^2 by ε_o^2 (a small constant) in the regularised equations (4.20) or (4.21). We thus obtain

$$(A + \varepsilon_o^2 B)\,\dot{\underline{y}} = (M^T C M + \varepsilon_o^2 W^T C W)\,\dot{\underline{y}} = \underline{g} \tag{4.41}$$

or

$$(A + \varepsilon_o^2 B^*)\,\dot{\underline{y}} = (M^T C M + \varepsilon_o^2 W^{*T} C W^*)\,\dot{\underline{y}} = \underline{g} \tag{4.42}$$

with $$W_j^* = \begin{bmatrix} m_{j-\frac{1}{2}} & 1 \\ m_{j+\frac{1}{2}} & 1 \end{bmatrix} \quad \text{and fixed } \varepsilon_o^2 \text{ (small).} \tag{4.43}$$

As a result the elements (4.24) of B^* become

$$\begin{bmatrix} <-\beta_i, -\beta_j> & <-\beta_i, \alpha_j> \\ <\alpha, -\beta_j> & <\alpha_i, \alpha_j> \end{bmatrix} \tag{4.44}$$

instead of (4.25).

The above regularisation also copes with the possible singularity in A when one of the $\Delta s_{j-\frac{1}{2}}$, $\Delta s_{j+\frac{1}{2}}$ is vanishingly small. If both $\Delta s_{j-\frac{1}{2}}$ and $\Delta s_{j+\frac{1}{2}}$ become vanishingly small, however, $A + \varepsilon^2 B^*$ still becomes singular. Note that in this case the other term $2 \dot{\underline{y}}^T \underline{g}$ in the norm squared (4.16) also goes to zero and the problem may be regarded as one of scaling. To some extent the technique of preconditioning covers this point by automatically scaling the Δs's (see [6]).
Another strategy is mentioned in the Conclusion (§9).

§5 Other Regularisations

It is undesirable to always force the nodes to have zero velocity in the event of near-parallelism or small node spacing. For example, if the nodes are travelling with the appropriate wave speed in a hyperbolic problem, say, an unnatural stop will damage the solution. In this section we give two alternative regularisations. The first penalises the acceleration rather than the velocity, the second penalises relative velocities rather than absolute velocities but in a way somewhat different from that of Miller (see §4).

(a) Regularisation by the Tangential Acceleration.

Replace the regularisation term (4.1) by

$$\varepsilon^2 \| V_T - V_T^{n-1} \|^2 \qquad (5.1)$$

where V_T^{n-1} is the tangential velocity (or an estimate) at the previous time step. Thus apart from a quantity Δt (absorbed into ε^2) the regularising term is an approximation to the tangential acceleration immediately prior to the current time. The result of this form of regularisation is to leave the left hand side of the MFE system (4.17) unaltered but to add terms to the right hand side which ensure that in the event of singularity nodes move with a tangential velocity similar to that at the previous time step.

Taking

$$V_T - V_T^{n-1} = \sum_j (\dot{T}_j - \dot{T}_j^{n+1})\ \alpha_j \tag{5.2}$$

the right hand side vector $\underline{\bar{g}}$ in (4.20) now becomes $\underline{\bar{g}}^*$ where

$$\underline{\bar{g}}^* = \underline{\bar{g}} + \varepsilon^2\ \bar{W}^T\, C\, \bar{W}\, \dot{\underline{\bar{y}}}^{n-1} \qquad (\dot{\underline{\bar{y}}}^{n-1} = (0, \dot{T}^{n-1})^T)$$

while $\underline{g}$ in (4.21) is replaced by

$$\underline{g}^* = \underline{g} + \varepsilon^2\ W^T\, C\, W\, \dot{\underline{y}}^{n-1} \tag{5.3}$$

This is a reasonable regularisation if nodes are not attempting to change speed along the tangent.

(b) Even the regularisation in (a) is not sufficiently general to cover all important situations, and in many ways the approach of Miller, as described in §4, in regularising with a term proportional

to (4.7) is desirable, since nodes will then tend to travel with the speed of their neighbours. The disadvantage of this approach is the appearance of the matrix K in (4.28) which allows the eigenvalue spectrum implicit in (4.34) to escape from the bounds $(\frac{1}{2}, \frac{3}{2})$ inherent in (4.20) (4.23) (in fact as far as (0,2), see Wathen [6]).

There is however another possible approach to penalising the relative speed of neighbouring nodes, which is to add a regularising term

$$\varepsilon^2 \| \delta^2 V_T \|^2 \tag{5.4}$$

where $\delta^2 V_T = (V_T)_{j+1} - 2(V_T)_j + (V_T)_{j-1}$ and V_T is given by (4.2).

As a result (4.16) becomes

$$\dot{\underline{y}}^T (\bar{M}^T C \bar{M} + \varepsilon^2 \Delta^T \bar{W}^T C \bar{W} \Delta) \dot{\underline{y}} - 2 \dot{\underline{y}} \underline{g} \tag{5.5}$$

where

$$\Delta = \begin{bmatrix} \ddots & & & & \\ & 0\ 0 & 0\ 0 & 0\ 0 & \\ & 0\ 1 & 0\text{-}2 & 0\ 1 & \\ & & 0\ 0 & 0\ 0 & 0\ 0 \\ & & 0\ 1 & 0\text{-}2 & 0\ 1 \\ & & & & \ddots \end{bmatrix} \tag{5.6}$$

with the corresponding form in the $\dot{a}$, $\dot{s}$ system

$$\dot{\underline{y}}^T (M^T C M + \varepsilon^2 \tilde{\Theta}^T \Delta^T \bar{W}^T C \bar{W} \Delta \tilde{\Theta}) \dot{\underline{y}} - 2 \dot{\underline{y}}^T \underline{g} \tag{5.7}$$

where $\Theta = \text{diag}\{\Theta_j\}$, $\Theta_j = \begin{bmatrix} \cos\theta_j & -\sin\theta_j \\ \sin\theta_j & \cos\theta_j \end{bmatrix}$. (5.8)

With this form of regularisation we preserve the result that the modified Rayleigh Quotient

$$\frac{\dot{\underline{y}}^T \left(M^T C M + \varepsilon^2 \Theta^T \Delta^T \bar{W}^T C \bar{W} \Delta \Theta\right) \dot{\underline{y}}}{\dot{\underline{y}}^T \left(M^T C_D M + \varepsilon^2 \Theta^T \Delta^T \bar{W}^T C_D W \Delta \Theta\right) \dot{\underline{y}}} \tag{5.9}$$

lies within the tight bounds given after (4.34), but unfortunately since the matrix Δ is block tridiagonal, rather than diagonal, the matrix of the denominator of (5.9) is no longer the block diagonal submatrix of the matrix of the numerator and is harder to invert. Thus the evaluation of the preconditioner itself becomes a problem.

§6 Two Dimensions

We now generalise the preceding theory to higher dimensions, illustrating the ideas throughout in the two-dimensional case.

Section 2 holds verbatim with the following changes only. In (2.3) we now have

$$\nu = \sum_j a_j \alpha_j \qquad \nu_t = \sum_j (\dot{a}_j \alpha_j + \dot{x}_j \beta_j + \dot{y}_j \gamma_j) \tag{6.1}$$

(see fig. 4 and refs. [1], [4]), and in each (triangular) element we

now write

$$u_t = \sum_{i=1}^{3} (\dot{a}_{ki}\ \phi_{ki} - m_k\ \dot{x}_{ki}\ \phi_{ki} - n_k\ \dot{y}_{ki}\ \phi_{ki}) \tag{6.2}$$

where the suffix i runs over the element vertices. The quadratic form to be minimised is again (2.5) with

$$\underline{\dot{y}}_H^T = (\dot{a}_{k1},\ \dot{s}_{k1},\ \dot{a}_{k2},\ \dot{s}_{k2},\ \dot{a}_{k3},\ \dot{s}_{k3}), \tag{6.3}$$

$$E_k = \frac{1}{12}\ \Delta_k \begin{bmatrix} 2\ \underline{m}\ \underline{m}^T & \underline{m}\ \underline{m}^T & \underline{m}\ \underline{m}^T \\ \underline{m}\ \underline{m}^T & 2\ \underline{m}\ \underline{m}^T & \underline{m}\ \underline{m}^T \\ \underline{m}\ \underline{m}^T & \underline{m}\ \underline{m}^T & 2\ \underline{m}\ \underline{m}^T \end{bmatrix}, \quad \underline{m} = \begin{bmatrix} 1 \\ -m_k \\ -n_k \end{bmatrix}, \tag{6.4}$$

Δ_k being the area of element k, and

$$\underline{G}_k^t = \{<\phi_{ki},\ \mathcal{L}(v)>,\ -m_k<\phi_{ki},\ \mathcal{L}(v)>,\ -n_k<\phi_{ki},\ \mathcal{L}(v)>\}\ i = 1,2,3 \tag{6.5}$$

Minimisation over $\dot{a}_{ki}$, $\dot{s}_{ki}$ leads to (2.9) as before.

Minimising now the <u>total</u> L_2 norm squared (2.8), <u>subject to the constraints</u>

$$\dot{a}_{j\mu} = \dot{a}_{j\Omega}\ ,\ \dot{x}_{j\mu} = \dot{x}_{j\Omega},\ \dot{y}_{j\mu} = \dot{y}_{j\Omega} \tag{6.6}$$

where μ,Ω refer to elements around the node j, we reduce the number of unknowns by roughly one sixth. Write

$$\begin{bmatrix} \dot{a}_{j\mu} \\ \dot{x}_{j\mu} \\ \dot{y}_{j\mu} \\ \vdots \end{bmatrix}_{\forall\mu} = R_j \begin{bmatrix} \dot{a}_j \\ \dot{x}_j \\ \dot{y}_j \end{bmatrix} \tag{6.7}$$

where $R_j^T = [I_3, I_3, \ldots, I_3]$, $I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$. (6.8)

With $\underline{\dot{y}}^T = \{\ldots, \dot{a}_j, \dot{x}_j, \dot{y}_j, \ldots\}$ and $\underline{\dot{Y}}^T = \{\dot{Y}_j^T\}$ we have

$$\underline{\dot{Y}} = R\underline{\dot{y}} \tag{6.9}$$

as before, (2.12), where R now contains elements I_3 in appropriate positions (but not in blocks since the numbering on the left hand side of (6.7) is elementwise). However there exists a permutation matrix Q such that Q^TRQ is in block form

$$Q^TRQ = \begin{bmatrix} \ddots & & \\ & [I_3 \; I_3 \; \ldots \; I_3] & \\ & & \ddots \end{bmatrix}^T \tag{6.10}$$

With these definitions we obtain (2.14) and (2.15), as before.

Furthermore, there exists a decomposition of E_k (2.20) as before,

where

$$\tilde{M}_k = \begin{bmatrix} 1 & -m_k & -n_k & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -m_k & -n_k & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -m_k & -n_k \end{bmatrix} , \tag{6.11}$$

which leads to the five matrix decomposition (2.23).

Note that $M = \tilde{M}R$ is now a rectangular matrix with no simple block structure, although there exist permutation matrices Q_1, Q_2, Q_3 such that

$$Q_1^T M Q_2 = (Q_1^T \tilde{M} Q_3)(Q_3^T R Q_2) \tag{6.12}$$

does have block (rectangular) diagonal structure.

These results extend to higher dimensions.

Turning now to the need for regularisation (§3) we make a transformation from $\dot{a}$, $\dot{x}$, $\dot{y}$ coordinates to $\dot{N}$, $\dot{T}$, $\dot{S}$ coordinates as follows. Let $\hat{\underline{t}}_{jk}$ and $\hat{\underline{s}}_j$ be unit vectors in the tangent plane to the solution at the point j which lie in the $\dot{s}_j$, $\dot{a}_j$ and $\dot{y}_j$, $\dot{a}_j$ planes, respectively.
We may write

$$\hat{t}_j = (\cos\theta_j, 0, \sin\theta_j) \quad , \quad \hat{\underline{s}}_j = (0, \cos\theta_j, \sin\theta_j) \tag{6.13}$$

(see fig. 5). The unit normal to the tangent plane is

$$\hat{\underline{n}}_j = \hat{\underline{t}}_j \times \hat{\underline{s}}_j = (-\sin\theta_j \cos\phi_j, -\cos\theta_j \sin\phi_j, \cos\theta_j, \cos\phi_j). \tag{6.14}$$

Then, if $\dot{N}_j$, $\dot{T}_j$, $\dot{S}_j$ are the components of $(\dot{a}_j, \dot{x}_j, \dot{y}_j)$ in the directions $\underline{\hat{n}}_j$, $\underline{\hat{t}}_j$, $\underline{\hat{s}}_j$,

$$\left.\begin{aligned} \dot{a}_j &= \dot{N}_j \cos\theta_j \cos\phi_j + \dot{T}_j \sin\theta_j + \dot{S}_j \sin\phi_j \\ \dot{x}_j &= \dot{T}_j \cos\theta_j - \dot{N}_j \sin\theta_j \cos\phi_j \\ \dot{y}_j &= \dot{S}_j \cos\theta_j - \dot{N}_j \cos\theta_j \sin\phi_j \end{aligned}\right\} \qquad (6.15)$$

Thus

$$u_t = \sum_j (\dot{a}_j \alpha_j + \dot{x}_j \beta_j \gamma_j) = \sum_j (\dot{N}_j \bar{\alpha}_j + \dot{T}_j \bar{\beta}_j + \dot{S}_j \bar{\gamma}_j) \qquad (6.16)$$

where

$$\left.\begin{aligned} \bar{\alpha}_j &= \alpha_j \cos\theta \cos\phi - \beta_j \sin\theta \cos\phi - \gamma_j \cos\theta \sin\phi \\ \bar{\beta}_j &= \alpha_j \sin\theta + \beta_j \cos\theta \\ \bar{\gamma}_j &= \alpha_j \cos\phi + \gamma_j \cos\theta \end{aligned}\right\} \qquad (6.17)$$

In the element k the contribution to u_t is

$$\sum_{i=1}^{3} (\mu_{ki} \dot{N}_k \phi_{ki} + \nu_{ki} \dot{T}_k \phi_{ki} + \Omega_{ki} \dot{S}_k \phi_{ki}) \qquad (6.18)$$

where

$$\left.\begin{aligned} \mu_{ki} &= \cos\theta_{ki}\cos\phi_{ki}\,(1 + m_k \tan\theta_{ki} + n_k \tan\phi_{ki}) \\ \nu_{ki} &= \sin\theta_{ki}\,(\tan\theta_{ki} - m_k) \\ \Omega_{ki} &= \sin\phi_{ki}\,(\tan\phi_{ki}\, n_k) \end{aligned}\right\} \qquad (6.19)$$

Then the contribution to the corresponding quadratic form from the element is (3.9) with

$$\dot{\underline{y}}_k^T = \left\{ \dot{N}_{ki},\ \dot{T}_{ki},\ \dot{S}_{ki} \right\}_{i=1,2,3} \qquad (6.20)$$

$$\bar{E}_k = \frac{1}{12}\,\Delta\alpha_k \begin{bmatrix} 2\,\underline{\bar{m}}\,\underline{\bar{m}}^T & \underline{\bar{m}}\,\underline{\bar{m}}^T & \underline{\bar{m}}\,\underline{\bar{m}}^T \\ \underline{\bar{m}}\,\underline{\bar{m}}^T & 2\,\underline{\bar{m}}\,\underline{\bar{m}}^T & \underline{\bar{m}}\,\underline{\bar{m}}^T \\ \underline{\bar{m}}\,\underline{\bar{m}}^T & \underline{\bar{m}}\,\underline{\bar{m}}^T & 2\,\underline{\bar{m}}\,\underline{\bar{m}}^T \end{bmatrix}, \quad \underline{\bar{m}}^T = [\mu,\ \nu,\ \omega] \qquad (6.21)$$

and $\underline{\bar{G}}_k^T$ is (3.11) extended from μ,ν to $\mu,\ \nu,\omega$ with basis functions ϕ_{ki} $(i = 1,2,3)$.

Parallelism occurs when the rank of $\underline{m}\,\underline{m}^T$ is reduced below 3, in which case the coefficient of a tangential velocity component vanishes.

As in §4, §5, we regularise the norm squared of the residual by adding terms

$$\varepsilon_T^2 \left\| V_T - V_T^{n-1} \right\|^2 + \varepsilon_S^2 \left\| V_S - V_S^{n-1} \right\|^2 \qquad (6.22)$$

where

$$V_T - V_T^{n-1} = \sum_j (\dot{T}_j - \dot{T}_j^{n-1})\alpha_j \quad , \quad V_S - V_S^{n-1} = \sum_j (\dot{S}_j - \dot{S}_j^{n-1})\alpha_j \tag{6.23}$$

which ensures that near to parallelism the node j is moved with (an estimate of) the tangential speeds V_T^{n-1}, V_S^{n-1} at the previous time step. As a result (4.3) becomes (when $\varepsilon_T = \varepsilon_S$)

$$\dot{\underline{y}}_k^T \bar{E}_k \dot{\underline{y}}_k - 2\dot{\underline{y}}_k^T \bar{\underline{G}}_k + \varepsilon^2 \dot{\underline{y}}_k^T F_k (\dot{\underline{y}}_k - \dot{\underline{y}}_k^{n-1}) \tag{6.24}$$

where

$$F_k = \frac{1}{12} \Delta_k \begin{bmatrix} 2\,\underline{e}\,\underline{e}^T & \underline{e}\,\underline{e}^T & \underline{e}\,\underline{e}^T \\ \underline{e}\,\underline{e}^T & 2\,\underline{e}\,\underline{e}^T & \underline{e}\,\underline{e}^T \\ \underline{e}\,\underline{e}^T & \underline{e}\,\underline{e}^T & 2\,\underline{e}\,\underline{e}^T \end{bmatrix}, \quad \underline{e} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \tag{6.25}$$

This leads to the global form (after the imposition of constraints)

$$\dot{\underline{y}}^T (R^T \bar{E} R + \varepsilon^2 R^T F R)\, \dot{\underline{y}} - 2\dot{\underline{y}}^T (\bar{\underline{g}} - \bar{\underline{g}}^{n-1}) \tag{6.26}$$

(c.f. (4.5)), where $\bar{E} = \text{diag}\,\{\bar{E}_k\}$, $F = \text{diag}\,\{F_k\}$, $\bar{\underline{g}}^{n-1} = \varepsilon^2 R^T F R\, \dot{\underline{y}}^{n-1}$,

or

$$\dot{\underline{y}}^T R^T (\bar{E} + \varepsilon^2 F) R \dot{\underline{y}} - 2\dot{\underline{y}} (\bar{\underline{g}} - \bar{\underline{g}}^{n-1}) \tag{6.27}$$

Following through the arguments of §4, §5, leads to

$$\dot{\underline{y}}^T \left(\bar{M}^T C \bar{M} + \varepsilon^2 \bar{W}^T C \bar{W}\right) \dot{\underline{y}} - 2\dot{\underline{y}} \left(\underline{g} - \underline{\bar{g}}^{n-1}\right) \tag{6.28}$$

where $$\underline{\bar{g}}^{n-1} = \varepsilon^2 \bar{W}^T C \bar{W}\, \dot{\underline{y}}^{n-1} \quad , \quad \bar{W} = \tilde{\bar{W}}\, R \tag{6.29}$$

and $$\tilde{\bar{W}} = \text{diag}\,\{\tilde{\bar{W}}_k\} \quad , \quad \tilde{\bar{W}}_k = \begin{bmatrix} 0\,1\,1 & 0\,0\,0 & 0\,0\,0 \\ 0\,0\,0 & 0\,1\,1 & 0\,0\,0 \\ 0\,0\,0 & 0\,0\,0 & 0\,1\,1 \end{bmatrix} , \forall_k \tag{6.30}$$

and, reverting back to $\dot{a}$, $\dot{x}$, $\dot{y}$ coordinates we obtain the regularised MFE equations (with the ε's absorbed into the W)

$$(M^T C M + \varepsilon^2 W^T C W)\dot{\underline{y}} = \underline{g} + \varepsilon^2 W^T C W \dot{\underline{y}}^{n-1}. \tag{6.31}$$

Here $W = \bar{W}\, P$ where $P = \text{diag}\,\{P_j\}$, (6.32)

$$P_j = \begin{bmatrix} \cos\theta_j\cos\phi_j & \sin\theta_j & \sin\phi_j \\ -\sin\theta_j\cos\phi_j & \cos\theta_j & 0 \\ -\cos\theta_j\sin\phi_j & 0 & \cos\phi_j \end{bmatrix}^{-1} = \frac{\begin{bmatrix} \cos\theta_j\cos\phi_j & \sin\theta_j\cos^2\phi_j & \cos^2\theta_j\sin\phi_j \\ \sin\theta_j\cos\phi_j & \cos\theta_j - \cos\theta_j & \sin\theta_j\sin\phi_j \\ \cos\theta_j\sin\phi_j & -\sin\theta_j\cos\phi_j & \sin\phi_j\cos\phi_j \end{bmatrix}}{\left(\cos^2\theta_j\cos^2\phi_j + \sin^2\theta_j\cos^2\phi_j + \cos^2\theta_j\sin^2\phi_j\right)} \tag{6.33}$$

or $$\left(M^T C M + \varepsilon^2\, P^T \bar{W}^T C\, \bar{W}\, P\right) \dot{\underline{y}} = \underline{g} + \varepsilon^2\, P^T \bar{W}^T C\, \bar{W}\, P\, \dot{\underline{y}}^{n-1}. \tag{6.34}$$

It remains to discuss the form of ε. As in §4 we may argue that the conditioning of the quadratic form

$$\dot{\underline{y}}^T \left(M^T C M + \varepsilon^2 \, W^T C W\right) \dot{\underline{y}} \tag{6.35}$$

is essentially that of

$$\dot{\underline{y}}^T \left(M^T C_D M + \varepsilon^2 W^T C_D W\right) \dot{\underline{y}} \tag{6.36a}$$

or

$$\dot{\bar{\underline{y}}}^T \left(\bar{M}^T C_D \bar{M} + \varepsilon^2 \bar{W}^T C_D \bar{W}\right) \dot{\bar{\underline{y}}} \tag{6.36b}$$

where C_D is the diagonal of C. Then using nodewise numbering, the matrix of (6.36b) is block diagonal with (nodal) blocks

$$\begin{bmatrix} \sum \Delta_k \, \mu^2 & \sum \Delta_k \, \mu\nu & \sum \Delta_k \, \mu_\Omega \\ \sum \Delta_k \, \nu\mu & \sum \Delta_k \, \nu^2 + \varepsilon^2 \sum \Delta_k & \sum \Delta_k \, \nu_\Omega \\ \sum \Delta_k \, \Omega\mu & \sum \Delta_k \, \Omega\nu & \sum \Delta_k \, \Omega^2 + \varepsilon^2 \sum \Delta_k \end{bmatrix} \tag{6.37}$$

In the case of parallelism the rank of $\underline{m}\,\underline{m}^T$ is less than 3. Again, as in §4, to keep good conditioning we choose (c.f. (4.38))

$$\varepsilon^2 = \varepsilon_j^2 \sim \mu_j^2 = \cos^2\theta_j \cos^2\phi_j \, (1 + m_j \tan\theta_j + n_j \tan\phi_j)^2$$

$$= \sec^2\theta_j \sec^2\phi_j \, (\cos^2\theta_j \cos^2\phi_j + \sin^2\theta_j \cos^2\phi_j + \cos^2\theta_j \sin^2\phi_j)^2 \tag{6.38}$$

With this form of ε^2 we can rewrite (6.34) with ε^2 replaced by ε_o^2 (small) and P_j replaced by

$$P_j = \begin{bmatrix} \cos\theta_j \cos\phi_j & \sin\theta_j \cos^2\phi_j & \cos^2\theta_j \sin\phi_j \\ \tan\theta_j \cos\phi_j & 1 & -\sin\theta_j \sin\phi_j \\ \cos\theta_j \tan\phi_j & -\sin\theta_j \sin\phi_j & 1 \end{bmatrix} \tag{6.39}$$

In one dimension $\bar{W}$ picks out the $\dot{T}$ component of P : here it will pick out the $\dot{T}$ and $\dot{S}$ components of P. Hence we expect $\bar{W}\,P$ (in the nodal numbering) to take the form

$$(\bar{W}\,P)_j = W_j = \begin{bmatrix} \tan\theta_{j\ell}\cos\phi_{j\ell} & 1 & -\sin\theta_{j\ell}\sin\phi_{j\ell} \\ \cos\theta_{j\ell}\tan\phi_{j\ell} & -\sin\theta_{j\ell}\sin\phi_{j\ell} & 1 \\ \vdots & \vdots & \vdots \end{bmatrix} \quad \forall\ \ell \tag{6.40}$$

$$= \begin{bmatrix} m_{j\ell}\cos\phi_{j\ell} & 1 & -\sin\theta_{j\ell}\sin\phi_{j\ell} \\ n_{j\ell}\cos\phi_{j\ell} & -\sin\theta_{j\ell}\sin\phi_{j\ell} & 1 \\ \vdots & \vdots & \vdots \end{bmatrix} \quad \forall\ \ell \tag{6.41}$$

where ℓ runs round the elements adjacent to node j.

A simpler strategy, based on an analogy with the one-dimensional case, is to split the term $\varepsilon^2\ W^T\ C\ W$ in (6.31) into two terms

$$\varepsilon_1^2\ W_T^T\ C\ W_T \quad \text{and} \quad \varepsilon_2^2\ W_S^T\ C\ W_S, \tag{6.42}$$

$$\text{where} \quad (W_T)_j = \begin{bmatrix} m_{j\ell} & 1 & 0 \\ 0 & 0 & 0 \\ \vdots & \vdots & \vdots \end{bmatrix}_{\forall\ell}, \quad (W_S)_j = \begin{bmatrix} 0 & 0 & 0 \\ n_{j\ell} & 0 & 1 \\ \vdots & \vdots & \vdots \end{bmatrix}_{\forall\ell}, \tag{6.43}$$

again using the nodal numbering. The two terms (6.42) regularise the $\dot{T}$ and $\dot{S}$ terms, respectively.

With this approach we arrive at the regularised MFE equations (c.f. (4.24), (4.25), (5.3)).

$$(A + \varepsilon_1^2 B_1^* + \varepsilon_2^2 B_2^*)\,\dot{\underline{y}} = \underline{g} + \varepsilon_1^2\, B_1^*\, \dot{\underline{y}}_1^{n-1} + \varepsilon_2^2\, B_2^*\, \dot{\underline{y}}_2^{n-1} \tag{6.44}$$

where A contains inner products

$$\begin{bmatrix} \langle \alpha_i, \alpha_j \rangle & \langle \alpha_i, \beta_j \rangle & \langle \alpha_i, \gamma_j \rangle \\ \langle \beta_i, \alpha_j \rangle & \langle \beta_i, \beta_j \rangle & \langle \beta_i, \gamma_j \rangle \\ \langle \gamma_i, \alpha_j \rangle & \langle \gamma_i, \beta_j \rangle & \langle \gamma_i, \gamma_j \rangle \end{bmatrix} \tag{6.45}$$

B_1^* contains inner products

$$\begin{bmatrix} \langle -\beta_i, -\beta_j \rangle & \langle -\beta_i, \alpha_j \rangle & 0 \\ \langle \alpha_i, -\beta_j \rangle & \langle \alpha_i, \alpha_j \rangle & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{6.46}$$

B_2^* contains inner products

$$\begin{bmatrix} \langle -\gamma_i, -\gamma_j \rangle & 0 & \langle -\gamma_i, \alpha_j \rangle \\ 0 & 0 & 0 \\ \langle \alpha_i, -\gamma_j \rangle & 0 & \langle \alpha_i, \alpha_j \rangle \end{bmatrix} \tag{6.47}$$

in which various orthogonality properties are obvious. When A becomes singular as a result of loss of rank, the addition of $\varepsilon_1^2 B_1^*$ or $\varepsilon_2^2 B_2^*$ or both will restore the rank. As in §5 the presence of the right hand side of (6.44) ensures that dominance of the ε^2 terms force the tangential velocity to be the same as its value $\dot{\underline{y}}^{n-1}$ (or an estimate) at the previous time step. Typically we might expect to choose ε_1^2 and ε_2^2 to be on the level of a tolerance, perhaps close to truncation error.

Although this simpler strategy relies on regularisation of the

matrix system by adding orthogonal matrices rather than on penalising the residual norm, it has the required effect and is equivalent to the penalty approach in one dimension.

§7 Scaling and Local Variants

One result of the regularisation above (seen most clearly in one dimension) is the presence of terms such as

$$M\dot{\underline{y}} \quad \text{and} \quad W\dot{\underline{y}} \tag{7.1}$$

in the same equation. The first of these leads to expressions of the type

$$\dot{a} - m\dot{s} \tag{7.2}$$

the terms of which are one-dimensionally consistent, but the second gives

$$m\dot{a} + \dot{s}, \tag{7.3}$$

which are not dimensionally consistent.

One way of clearing up this difficulty is to work with the variables $\dot{a}$ and $\bar{m}\dot{s}$ rather than $\dot{a}$ and $\dot{s}$, where $\bar{m}$ is (for example) an average slope at the node. We must then redefine β as $-m\alpha/\bar{m}$ (and γ similarly). Singularity and regularisation occur as before, but (7.2) and (7.3) now become

$$\left.\begin{array}{l} \dot{a} - \dfrac{m}{\bar{m}}\,(\bar{m}\dot{s}) \\ \\ \dfrac{m}{\bar{m}}\,\dot{a} + (\bar{m}\dot{s}) \end{array}\right\} \qquad (7.4)$$

Dimensional consistency is restored and the only additional task is to calculate $\dot{s}$ from $(\bar{m}\dot{s})$. (We can take $\bar{m} = \frac{1}{2}(|m_{j-\frac{1}{2}}| + |m_{j+\frac{1}{2}}|)$.

The elementwise construction of Miller & Carlson in §2 is affected by this procedure because $\bar{m}$ is an average over all elements adjacent to a node (although it appears that this may be postponed to the assembly stage). All matrices must be recalculated using the new β's and γ's and the rotation of §3 now becomes one of 45°. Apart from these differences, the essential results are unchanged.

Turning now to a variant of the method (the local method) we first write the equation

$$(A + \varepsilon_0^2\beta)\dot{\underline{y}} = (M^TCM + \varepsilon_0^2 W^TCW)\dot{\underline{y}} = \underline{g} + \varepsilon_0^2 W^TCW\dot{\underline{y}}^{u-1} \qquad (7.5)$$

as two steps. In the first step we minimise

$$\| v_t - \mathcal{L}(v) \|^2 \qquad (7.6)$$

over v's of the form

$$v = \sum_k \sum_\nu w_{k\nu}\, \phi_{k\nu}$$

as in [4]. This leads to the equation

$$C\underline{w} = \underline{b} \qquad (7.7)$$

where $\underline{w} = \{w_{k\nu}\}$ and $\underline{b} = \{b_{k\nu}\}$, $b_{k\nu} = \langle \phi_{k\nu}, \mathcal{L}(\nu) \rangle$ and we note that $M^T \underline{b} = \underline{g}$. No regularisation is needed in this step.

In the second step, standard MFE minimises

$$\| C^{\frac{1}{2}} (M\dot{\underline{y}} - \underline{w}) \|^2 \tag{7.8}$$

over $\dot{\underline{y}}$ to yield the equation $M^T C M \dot{\underline{y}} = M^T \underline{w} = \underline{g}$. In the regularised form we minimise

$$\| C^{\frac{1}{2}} (M\dot{\underline{y}} - \underline{w}) \|^2 + \varepsilon_0^2 \| C^{\frac{1}{2}} (W (\dot{\underline{y}} - \dot{\underline{y}}^{n-1})) \|^2 \tag{7.9}$$

to yield (7.5).

The local form arises from replacing C in (7.9) by Δ, where $\Delta = \text{diag } \{\Delta_k\}$ and Δ_k is the size of the element k. The method may be viewed as arising from a Petrov-Galerkin approach [9] or minimisation in an alternative norm [7]. The minimisation (7.9) then becomes

$$\| \Delta^{\frac{1}{2}} (M\dot{\underline{y}} - \underline{w}) \|^2 + \varepsilon_0^2 \| \Delta^{\frac{1}{2}} (W (\dot{\underline{y}} - \dot{\underline{y}}^{n-1})) \|^2 \tag{7.10}$$

to yield

$$(M^T \Delta M + \varepsilon_0^2 W^T \Delta W) \dot{\underline{y}} = \underline{g}^L + \varepsilon_0^2 W^T \Delta W \dot{\underline{y}}^{n-1} \tag{7.11}$$

where

$$\underline{g}^L = M^T \Delta \underline{w}$$

and, by the nature of the left hand side matrix (proportional to the $M^T C_D M + \varepsilon_0^2 W^T C_D W$ of (4.34)) we have a local nodewise method

$$(M_j^T \Delta_j M_j + \varepsilon_0^2 W_j^T \Delta_j W_j)\dot{\underline{y}}_j = \underline{g}_j^L + \varepsilon_0^2 W_j^T \Delta_j W_j \dot{\underline{y}}_j^{n-1} \qquad (7.12)$$

$\forall_j$ where $\Delta_j = \text{diag}[\Delta_{j-\frac{1}{2}}, \Delta_{j+\frac{1}{2}}]$, regularised to avoid the parallelism singularity.

We conclude with the local MEM form (see Edwards [10]), which can be constructed in the same way as the local form above with M_j replaced by $[-m_{j-\frac{1}{2}} \quad -m_{j+\frac{1}{2}}]^T$ and $\dot{\underline{y}}_j$ replaced by $\dot{s}_j$, the nodal speed. A singularity occurs when $m_{j-\frac{1}{2}} = m_{j+\frac{1}{2}} = 0$, and in this case W_j may be chosen to be an arbitrary non-zero vector, say [1 - 1]. Since the resulting form of (7.12) is scalar, both W_j^T and W_j may be absorbed into the ε_0^2, leading to the form

$$\left[[-m_{j-\frac{1}{2}} \quad -m_{j+\frac{1}{2}}] \Delta_j \begin{bmatrix} -m_{j-\frac{1}{2}} \\ \\ -m_{j+\frac{1}{2}} \end{bmatrix} + \varepsilon_0^2(\Delta_{j-\frac{1}{2}} + \Delta_{j+\frac{1}{2}}) \right] \dot{s}_j =$$

$$g_j^L + \varepsilon_0^2(\Delta_{j-\frac{1}{2}} + \Delta_{j+\frac{1}{2}})\, \dot{s}_j^{n-1} \qquad (7.13)$$

or

$$\dot{s}_j = \frac{g_j^L + \varepsilon_0^2\,(\Delta_{j-\frac{1}{2}} + \Delta_{j+\frac{1}{2}})\, \dot{s}_j^{n-1}}{m_{j-\frac{1}{2}}^2\, \Delta_{j-\frac{1}{2}} + m_{j+\frac{1}{2}}^2\, \Delta_{j+\frac{1}{2}} + \varepsilon_0^2(\Delta_{j-\frac{1}{2}} + \Delta_{j+\frac{1}{2}})} \qquad (7.14)$$

where

$$g_j^L = [-m_{j-\frac{1}{2}} \; -m_{j+\frac{1}{2}}] \Delta \underline{w} = -(m_{j-\frac{1}{2}} \; \Delta_{j-\frac{1}{2}} \; w_{j-\frac{1}{2}} + m_{j+\frac{1}{2}} \; \Delta_{j+\frac{1}{2}} \; w_{j+\frac{1}{2}}) \tag{7.15}$$

(c.f. [10], [11]).

From the form of (7.14) we see that if $m_{j-\frac{1}{2}} = m_{j+\frac{1}{2}} = 0$ the value of $\dot{s}_j$ remains determinate equal to an estimate of its value at the previous time step.

§8 Conclusions.

The main conclusion of this report is that simple regularisations of the MFE method exist which preserve the fast solution techniques associated with the explicit approach.

In summary the MFE equations

$$A \, \underline{\dot{y}} = \underline{g} \tag{8.1}$$

may be replaced (in one dimension) by the equations

$$(A + \varepsilon^2 B) \, \underline{\dot{y}} = \underline{g} + \varepsilon_0^2 B \, \underline{\dot{y}}^{n-1} \tag{8.2}$$

where $\underline{\dot{y}}^{n-1}$ is an estimate of $\underline{\dot{y}}$ at the previous time step.

Inversion of the l.h.s. matrix may be carried out easily using the preconditioned conjugate gradient method, as for A above (see below).

In the global form, if A consists of blocks

$$\begin{bmatrix} \langle \alpha_i, \alpha_j \rangle & \langle \alpha_i, \beta_j \rangle \\ \langle \beta_i, \alpha_j \rangle & \langle \beta_i, \beta_j \rangle \end{bmatrix} \tag{8.3}$$

then B consists of blocks

$$\begin{bmatrix} \langle -\beta_i, -\beta_j \rangle & \langle -\beta_i, \alpha_j \rangle \\ \langle \alpha_i, -\beta_j \rangle & \langle \alpha_i, \alpha_j \rangle \end{bmatrix} \tag{8.4}$$

A similar regularisation in two dimensions leads to

$$(A + \varepsilon_1^2 B_1 + \varepsilon_2^2 B_2)\dot{\underline{y}} = \underline{g} + \varepsilon_1^2 B_1 \dot{\underline{y}}^{n-1} + \varepsilon_2^2 B_2 \dot{\underline{y}}^{n-1} \tag{8.5}$$

where the suffices 1 and 2 refer to x and y component terms.

For B_0 and B_1 see equations (6.46) and (6.47).

Another useful result is that the B of (8.2) has the decomposition

$$B = \overset{*T}{W} C \overset{*}{W}, \tag{8.6}$$

See (4.42), which contrasts with the Miller regularisation which takes the form

$$W^T K W, \tag{8.7}$$

(see (4.28)). As a result the MFE matrix, when preconditioned by the inverse of its diagonal (in blocks), has eigenvalues located in a small finite interval of the real axis bounded below by $\frac{1}{2}$, whereas the Miller regularisation permits the corresponding eigenvalues to occur close to the origin.

The only respect in which (8.2) (or (8.5)) may lead to an ill-conditioned system is if all the elements surrounding a node become vanishingly small. A strategy for overcoming even this case is as follows. Let

$$C = \{C_k\} \quad , \quad C_k = \Delta_k \, T_k \tag{8.8}$$

where Δ_k is the size of the element k and $T_k = \{T_{kij}\}$, $T_{kij} = \langle \phi_{ki}, \phi_{kj} \rangle$ (8.9)

Then

$$A + \varepsilon^2 B = M^T C M + \varepsilon_0^2 W^T C W \tag{8.10}$$

$$= M^T \Delta^T T \Delta M + \varepsilon_0^2 W^T \Delta^T T \Delta W \tag{8.11}$$

where $T = \text{diag}\,\{T_k\}$ and $\Delta = \text{diag}\,\{\Delta_k\}$. Now let

$$\Delta_\delta = \text{diag}\,\{\max(\Delta_k, \Delta_0)\} \tag{8.12}$$

where Δ_0 is some small element size, and define

$$\left.\begin{aligned} A &= M^T(\Delta^T)^{\frac{1}{2}T}\,\Delta^{\frac{1}{2}}M \quad , \quad A_\delta = M^T(\Delta_\delta^T)^{\frac{1}{2}T}\,\Delta_\delta^{\frac{1}{2}}\,M \\ B &= W^T(\Delta^T)^{\frac{1}{2}T}\,\Delta^{\frac{1}{2}}W \quad , \quad B_\delta = W^T(\Delta_\delta^T)^{\frac{1}{2}T}\,\Delta_\delta^{\frac{1}{2}}\,W \end{aligned}\right\} \tag{8.13}$$

We now solve the (always) well-conditioned system

$$[A + A_\delta + \varepsilon_0^2\,(B + B_\delta)]\dot{\underline{y}} = \underline{g} + [A_\delta + \varepsilon_0^2(B + B_\delta)]\dot{\underline{y}}^{n-1}, \tag{8.14}$$

constructed on the same principle as in §4.

The corresponding local method comes out of the decomposed form (c.f. (7.10))

$$\begin{aligned} &[M^TCM + M^TC_\delta M + \varepsilon_0^2(W^TCW + W^TC_\delta W)]\ \dot{\underline{y}} \\ &\qquad = \underline{g} + [M^TC_\delta M + \varepsilon_0^2(W^TCW + W^TC_\delta W)]\ \dot{\underline{y}}^{n-1} \end{aligned} \tag{8.15}$$

where $C = \text{diag}\,\{C_k\}$, $C_k = \Delta_\delta T_k$. We replace C and C_δ by Δ and Δ_δ throughout, giving

$$\begin{aligned} &[M^T\Delta M + M^T\Delta_\delta M + \varepsilon_0^2\,(W^T\Delta W + W^TC_\delta W)]\ \dot{\underline{y}} \\ &\qquad = \underline{g} + [M^T\Delta_\delta M + \varepsilon_0^2(W^T\Delta W + W^T\Delta_\delta W)]\ \dot{\underline{y}}^{n-1} \quad , \end{aligned} \tag{8.16}$$

all matrices being diagonal in 2 x 2 blocks, so a local method.

Finally the local MEM form is obtained by replacing M_j^T by $[-m_{j-\frac{1}{2}} - m_{j+\frac{1}{2}}]$ and $\dot{\underline{y}}_j$ by $\dot{s}_j$, giving eventually

$$\dot{s}_j = \frac{g_j + \{\varepsilon_0^2(\Delta s_{j-\frac{1}{2}} + \Delta s_0 + \Delta s_{j+\frac{1}{2}} + \Delta s_0) + \Delta s_0 (m_{j-\frac{1}{2}}^2 + m_{j+\frac{1}{2}}^2)\ \dot{s}_j^{n-1}}{(\Delta s_{j-\frac{1}{2}} + \Delta s_0)(m_{j-\frac{1}{2}} + \varepsilon_0^2) + (\Delta s_{j+\frac{1}{2}} + \Delta s_0)(m_{j+\frac{1}{2}} + \varepsilon_0^2)}, \tag{8.17}$$

c.f. (7.13). In this method the eigenvalue structure is preserved.

We end the discussion with a remark about $\underline{\dot{y}}^{n-1}$. In view of possible irregularities in individual adjacent velocities, it seems advisable to average them in some way to obtain $\underline{\dot{y}}^{n-1}$. A possible formula is

$$\underline{\dot{y}}_j^{n-1} = \frac{w_{-1}\ \underline{y}_{j-1}^{n-1} + w_o \underline{y}_j^{n-1} + w_1\ \underline{y}_{j+1}^{n-1}}{w_{-1} + w_o + w_1} \tag{8.18}$$

where w_{-1}, w_o, w, are weights. An obvious choice is $w_o = 2$, $w_{-1} = w_1 = 1$, but we may also weight by length in the manner

$$w_{-1} = \Delta s_{j+\frac{1}{2}}\ ,\ w_o = 0,\quad w_1 = \Delta s_{j-\frac{1}{2}}. \tag{8.19}$$

This weighting has the advantage that when one Δs spacing becomes very small, the velocity $\underline{\dot{y}}_j^{n-1}$ approximates the $\underline{\dot{y}}$ on the same side. This applies equally to $\dot{s}_j^{n-1}$.

References

1. K. Miller & R. Miller (1981), SIAM J.N.A., 18, 1019 - 1032
K. Miller (1981), SIAM J.N.A., 18, 1033 - 1056.

2. K. Miller & N. Carlson (1986), private communication.

3. M.J. Baines (1986), Numerical Analysis report 2/86, Dept. of Maths., University of Reading.

4. A.J. Wathen & M.J. Baines (1985), IMA Journal of Numerical Analysis 5, 161 - 182.

5. M.J. Baines (1985), Numerical Analysis report 1/85, Dept. of Mathematics, University of Reading.

6. A.J. Wathen (1985), Attainable Eigenvalue Bounds for the Galerkin Mass Matrix, to appear.

7. K. Miller (1984), Proceedings of ARFEC Conference, Lisbon, ed. Zienkiewicz (Wiley).

8. K. Miller (1986), private communication.

9. M.J. Baines & A.J. Wathen (1986), submitted to J. Comput. Phys.

10. M.G. Edwards (1985), Numerical Analysis report 20/85, Dept. of Maths., University of Reading.

11. M.G. Edwards & M.J. Baines (1986), submitted to J. Comput. Phys.

Appendix A.

Although regularisation of the tangential velocity is exactly what is needed theoretically (see also Appendix B), the main practical difficulty in the use of MFE is the problem of elevent folding and this may be attacked directly by regularising with the (horizontal) nodal velocity rather than the tangential velocity. Instead of (4.1) we have

$$\varepsilon^2 \| \dot{s} \|^2 \quad , \quad \dot{s} = \sum \dot{s}_j \alpha_j \tag{A.1}$$

and we then obtain (4.20) with

$$W = \text{diag } \{W_j\}, \; W_j = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix} , \tag{A.2}$$

or (4.21) with

$$B = \text{diag } \{B_j\}, \; B_j = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} . \tag{A.3}$$

In two dimensions the corresponding regularisation is

$$\varepsilon_0^2 \| \dot{x} \|^2 + \varepsilon^2 \| \dot{y}^2 \| \tag{A.4}$$

where $$\dot{x} = \sum \dot{x}_j \alpha_j \qquad \dot{y} = \sum \dot{y}_j \alpha_j \tag{A.5}$$

and then we obtain (6.31), (6.44) (with just $\underline{g}$ on the r.h.s.), with

$$\begin{aligned} W_1 &= \text{diag } \{W_{1j}\} \quad , \quad W_2 = \text{diag } \{W_{2j}\} \\ B_1 &= \text{diag } \{B_{1j}\} \quad , \quad B_2 = \text{diag } \{B_{2j}\} \end{aligned} \tag{A.6}$$

$$W_{1j} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad B_{1j} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad , \quad B_{2j} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad B_{2j} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{A.7}$$

In effect we have set β and γ equal to 0 in these matrices, consistent with the tangent to the solution being horizontal. Extra r.h.s. terms may be added as in (5.3) and (6.44).

The local form of these equations is (7.11) with W replaced by W_j (see (A.2)) in one dimension and by W_{1j} or W_{2j} (see (A.7) in two dimensions. We repeat these forms below for convenience. In one dimension the MFE equations are

$$(A_j^L + \varepsilon^2 B_j^L)\, \dot{\underline{y}}_j = \underline{g}_j^L + \varepsilon_0^2 B_j^L \dot{\underline{y}}_j^{n-1} \tag{A.8}$$

where

$$A_j^L = M_j^T \Delta_j M_j$$
$$B_j^L = W_j^T \Delta_j W_j \tag{A.9}$$

and W_j is as in (A.2). A corresponding two-dimensional form is

$$(A_j^L + \varepsilon_1^2 B_{1j}^L + \varepsilon_2^2 B_{2j}^L)\dot{\underline{y}}_j = \underline{g}_j^L + \varepsilon_1^2 B_{1j}^L \dot{\underline{y}}_j^{n-1} + \varepsilon_2^2 B_{2j}^L \dot{\underline{y}}_j^{n-1} \tag{A.10}$$

where

$$A_j^L = M_j^T \Delta_j M_j \quad , \quad B_{1j}^i = W_{1j}^T \Delta_j W_{1j} \quad , \quad B_{2j}^L = W_{2j}^T \Delta_j W_{2j} \tag{A.11}$$

and W_{1j}, W_{2j} are as in (A.7). Equation (A.8) is a 2 x 2 system and (A.10) is a 3 x 3 system.

These regularisation terms are the same as in the local MEM approach (7.14). In relation to the minimisation structure given in §7, every statement there is correct provided that W is redefined as in (A.2).

We can also add A_δ and B_δ matrices as in §8 to overcome the problem of simultaneous folding.

Appendix B

In this appendix we look particularly at the local MFE method and devise a geometrical construction for a family of moving element schemes which are all conservative.

Leaving aside regularisation and returning to the theme of §3 we write the local MFE method in the form

$$M^T \Delta M \dot{\underline{y}} = \underline{g}^L = M^T \Delta \underline{w} \tag{B.1}$$

(see (7.7) and (7.11) with $\varepsilon_0 = 0$). The first equation of each pair of equations of (B.1) comprises the system

$$L^T \Delta M \dot{\underline{y}} = L^T \Delta \underline{w} \tag{B.2}$$

where L consists of rows of 0's and 1's.

Let $C = T\Delta$ where T is the purely numerical matrix $\mathrm{diag}\{T_j\}$,

$$T_j = \frac{1}{6}\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \quad \text{(in one dimension)} \; : \; T_j = \frac{1}{12}\begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix} \quad \text{(in two dimensions)} \tag{B.3}$$

Note that the rows of L are eigenvectors of T with common eigenvalue $\frac{1}{N}$ (N = 1,2 being the number of dimensions).

Multiplying (B.2) by T gives

$$L^T C M \dot{\underline{y}} = L^T C \underline{w} \tag{B.4}$$

which corresponds to the equation obtained by minimising (2.2) over the α_j of (2.3), namely

$$<\alpha_j, v_t - \mathcal{L}(v)> = 0 \tag{B.5}$$

Since the α_j are a partition of unity, summation of the equations (B.5) gives

$$\int [v_t - \mathcal{L}(v)] d\Omega = 0 \tag{B.6}$$

which shows that any scheme which includes (B.5) is conservative.

Now (B.4) or (B.2) is a rectangular system and the matrix $L^T \Delta M$ possesses a null space. At the node j in one dimension we have

$$(L^T \Delta M)_j = [\Delta s_{j-\frac{1}{2}} + \Delta s_{j+\frac{1}{2}} \quad - m_{j-\frac{1}{2}} \Delta s_{j-\frac{1}{2}} - m_{j+\frac{1}{2}} \Delta s_{j+\frac{1}{2}}], \tag{B.7}$$

whose null space is clearly spanned by

$$\underline{\dot{y}}_{oj} = \begin{bmatrix} m_{j-\frac{1}{2}} \Delta s_{j-\frac{1}{2}} + m_{j+\frac{1}{2}} \Delta s_{j+\frac{1}{2}} \\ \\ \Delta s_{j-\frac{1}{2}} \quad + \quad \Delta s_{j+\frac{1}{2}} \end{bmatrix} = \begin{bmatrix} \Delta a_{j-\frac{1}{2}} + \Delta a_{j+\frac{1}{2}} \\ \\ \Delta s_{j-\frac{1}{2}} + \Delta s_{j+\frac{1}{2}} \end{bmatrix} \tag{B.8}$$

So

$$L^T \Delta M \underline{\dot{y}}_{oj} = 0 \tag{B.9}$$

and the direction of $\underline{\dot{y}}_{oj}$ is along the line joining the nodes $j-1$ and $j+1$ (see fig.5).

Suppose that $\dot{\underline{Y}}_j$ is the solution of (B.1). Then

$$\dot{\underline{z}}_j = \dot{\underline{y}}_j + k_j \dot{\underline{y}}_{oj} \tag{B.10}$$

is a solution of (B.2), by virtue of (B.9). By (B.6) this solution is conservative for any k_j. Moreover there is a geometric construction for $\dot{\underline{z}}_j$ which is also shown in fig. 5. The vectors $\dot{\underline{z}}_j$ are obtained by drawing a line (in one dimension) through the end of the $\dot{\underline{y}}_j$ vector parallel to the line joining nodes $j-1$ and $j+1$.

Where this line intersects the vertical line $x = s_j$ there must be the conservative method on a fixed grid (the local fixed finite element approximation). In this case

$$k_j = -\dot{s}_j/(\Delta s_{j-\frac{1}{2}} + \Delta s_{j+\frac{1}{2}}) \quad ; \quad (\dot{a}_j)_{\text{fixed}} = (\dot{a}_j)_{\text{moving}} - \dot{s}_j \frac{(\Delta a_{j-\frac{1}{2}} + \Delta a_{j+\frac{1}{2}})}{(\Delta s_{j-\frac{1}{2}} + \Delta s_{j+\frac{1}{2}})} \tag{B.11}$$

but for any value of k_j a conservative method is obtained (in a semi-discrete sense).

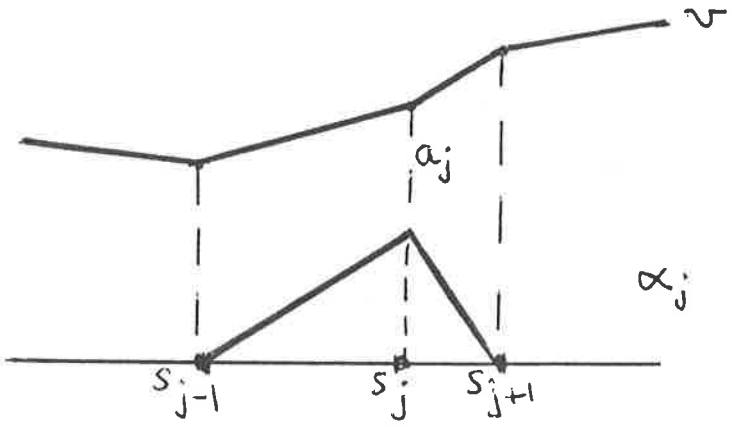
It is easy to see how overshoots occur with the fixed method. In fig. 6 which shows data convected by the scalar wave equation

$$u_t + a u_x = 0 \qquad a > 0 \tag{B.12}$$

the construction of $(\dot{\underline{y}}_j)_{\text{fixed}}$ uses $(\dot{\underline{y}}_j)_{\text{moving}}$ and the "tangent" to v at the point j (being in fact the chord joining $j-1$ to $j+1$). The result is a clear overshoot.

The connection of this section with the need for tangential regularisation in §3 is clear. If the β equation does not lead to singularity then $\dot{\underline{y}}_{MFE}$ is the undisputed solution. But where singularity arises there is reason to select a special nodal speed which should be one of the conservative choices above.

All the regularisations mentioned in this report fall into the framework described here.



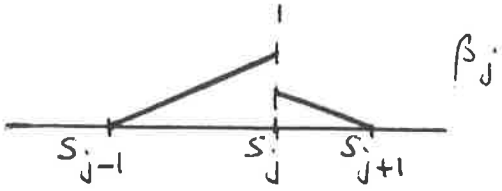


fig. 1

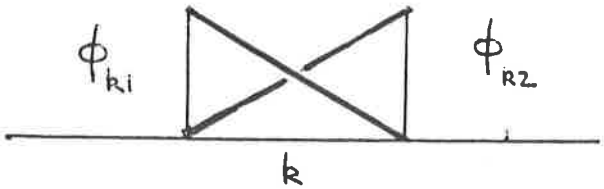


fig. 2.

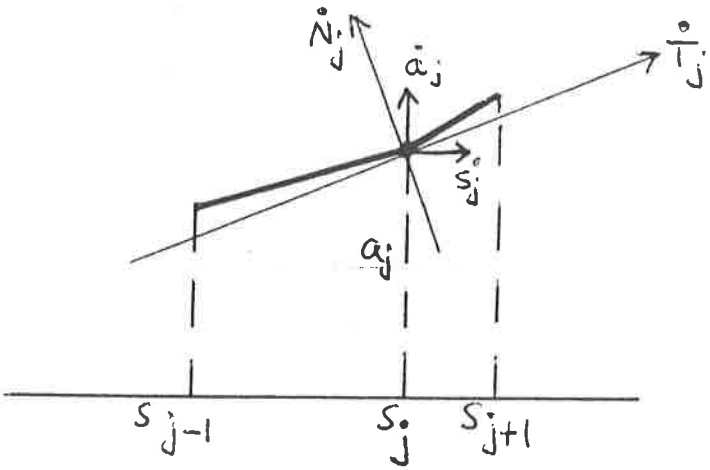
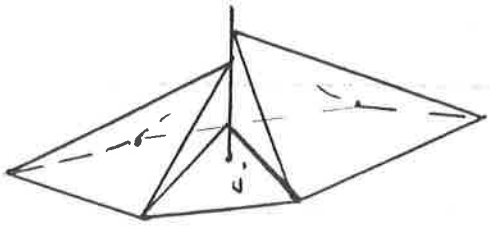


fig. 3



α_j β_j γ_j

fig. 4

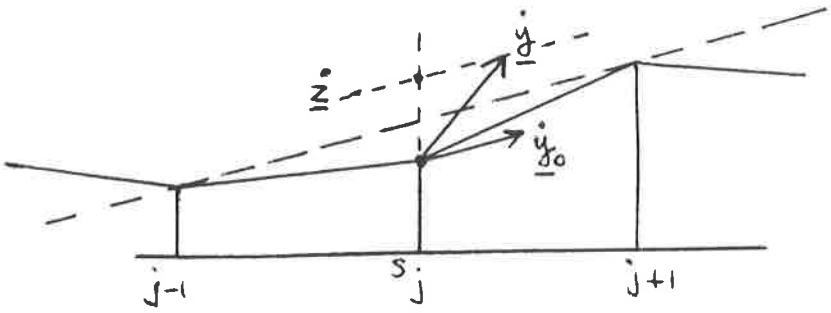

$\underline{\dot{y}}$
$\underline{\dot{z}}$
$\underline{\dot{y}}_0$
j-1
s
j
j+1

fig. 5

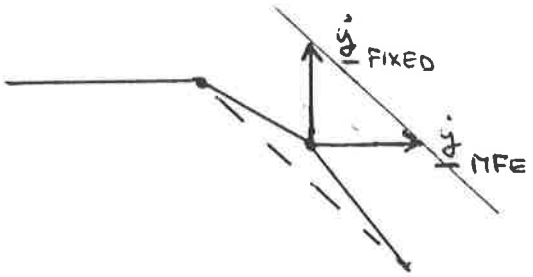

$\underline{\dot{y}}$ FIXED
$\underline{\dot{y}}$ MFE

fig. 6