D E P A R T M E N T   O F   M A T H E M A T I C S

ON THE RELATIONSHIP BETWEEN THE MOVING FINITE ELEMENT PROCEDURE AND BEST PIECEWISE LINEAR L_2 FITS WITH ADJUSTABLE NODES

M.J. Baines

Numerical Analysis Report 2/91

U N I V E R S I T Y   O F   R E A D I N G

ON THE RELATIONSHIP BETWEEN THE MOVING FINITE ELEMENT PROCEDURE AND BEST PIECEWISE LINEAR L_2 FITS WITH ADJUSTABLE NODES

M.J. Baines

Numerical Analysis Report 2/91

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

Abstract

It is argued that for certain ordinary differential equations the Moving Finite Element (MFE) procedure of Miller closely parallels an algorithm for obtaining best L_2 fits with adjustable nodes, reducing the magnitude of an L_2 norm at each step. By considering other types of projection in the MFE procedure generally, the sensitivity of the nodal speeds to the type of projection is also demonstrated. Two effects are identified. For first order partial differential equations the presence of u_x terms generates approximate characteristic speeds whatever the projection while, for the L_2 projection, there is an additional speed which may reduce an L_2 norm.

§1. Introduction

In [1] a simple procedure (called here MBF) was given for determining the best piecewise linear continuous L_2 fit to a convex function of a single variable, with an extension for obtaining near-best piecewise linear L_2 fits to non-convex functions. The procedure has much in common with the Moving Finite Element (MFE) procedure of Miller [2] as presented by Wathen & Baines [3], Baines & Wathen [4]. In both procedures best L_2 fits to functions are first found amongst piecewise linear discontinuous functions on an element basis. This information is then used to generate updated nodal positions. In the case of MBF this procedure is iterated to convergence. The aim of this report is to bring out the analogy between the two procedures. In the process we discuss the nature of the MFE projection and its consequences.

§2. The MBF Procedure

In [1] the following simple algorithm, based on a two-step minimisation procedure, for finding the best piecewise linear L_2 fit $U(x)$, with adjustable nodes, to a continuous function $f(x)$, is described.

(i) Set up initial node positions X_j $(j = 1,2,\ldots,J)$, perhaps equally spaced.

(ii) For each element k, corresponding to the interval (X_{k-1},X_k) obtain $\mathcal{P}_k f(x)$, the best L_2 fit to $f(x)$ in element k.

(iii) For each node j, common to elements $k-1$ and k, obtain a new nodal position through a displacement δX_j calculated from

$$\{\mathcal{P}_{k-1}f(X_j) + M_{k-1}\delta X_j - f(X_j)\}^2 = \{\mathcal{P}_k f(X_j) + M_k\delta X_j - f(X_j)\}^2 \quad (2.1)$$

where

$$M_{k-1} = (U_x)_{k-1},\ M_k = (U_x)_k \ .$$

There are two cases:

(a) if $f(X_j) - \mathscr{P}_{k-1}f(X_j)$ and $f(X_j) - \mathscr{P}_k f(X_j)$ have the same sign (certainly true if $f(x)$ convex),

$$\delta X_j = - \frac{\{\mathscr{P}_k f(X_j) - \mathscr{P}_{k-1} f(X_j)\}}{M_k - M_{k-1}} \tag{2.2}$$

(b) otherwise (near a point of inflection of $f(x)$, say)

$$\delta X_j = - \frac{\{\mathscr{P}_{k-1} f(X_j) + \mathscr{P}_k f(X_j) - 2f(X_j)\}}{M_k + M_{k-1}} \tag{2.3}$$

(iv) Repeat steps (ii) and (iii) to convergence.

<u>Remarks:</u>

(a) Steps (ii) and (iii) are individual solutions of two constrained minimisation problems.

(b) Step (iv) is needed because the best fit requires the solution of the two minimisation problems simultaneously.

§3. The MFE Procedure

As formulated in [3],[4], the Moving Finite Element (MFE) procedure is also a two step procedure having a similar structure to that of MBF. Briefly described, the procedure is as follows.

For the 1-D partial differential equation

$$u_t = \mathcal{L}(u) \tag{3.1}$$

with all space derivatives contained within the operator $\mathcal{L}$, the following steps are carried out:

(i) Take the current nodal positions X_j and nodal amplitudes U_j^n $(j = 1,2,\ldots,J)$.

(ii) For each element k, corresponding to the interval (X_{k-1},X_k), obtain $\mathcal{P}_k\mathcal{L}(x)$, the best L_2 fit to $\mathcal{L}(U^n)$ in the element k, where U^n is the current piecewise linear approximation.

(iii) For each node j, common to elements $k-1$ and k, obtain a nodal speed $\dot{X}_j$ calculated from

$$\dot{X}_j = -\frac{\{\mathcal{P}_k\mathcal{L}(X_j)-\mathcal{P}_{k-1}\mathcal{L}(X_j)\}}{M_k-M_{k-1}} \tag{3.2}$$

($\dot{U}_j$ can then be deduced.)

(iv) Step forward in time, perhaps using explicit Euler time-stepping, from step (ii).

<u>Remarks</u>:

(a) Unlike the MBF procedure:

1. the MFE procedure is not an iteration procedure but a series of time steps.

2. There is no interaction between steps (ii) and (iii) provided that the time step is explicit and requires only $\mathcal{L}(U)$ at time level n.

(b) In the non-convex inflection case in which $M_k \simeq M_{k-1}$ there is a clear danger of bad numerical conditioning in (3.2). In the MBF case equation (2.3) then applies but if this were used in MFE it would no longer satisfy the finite dimensional form of (3.1). However it would, if solved implicitly, give the same value of

$$(u-f)_k^2 - (u-f)^2{}_{k-1} \tag{3.3}$$

as (2.1).

§4. A Special Case

In order to analyse the connection between the two procedures we consider first the application of the MFE procedure of §3 to the equation

$$u_t = - u + f(x) \tag{4.1}$$

where $f(x)$ is convex. Step (ii) of §3 requires $\mathcal{P}_k\mathcal{L}(x)$, the best L_2 fit to $- U^n + f(x)$ in element k and this is evidently

$$\mathcal{P}_k\mathcal{L}(x) = - U_k^n(X) + \mathcal{P}_k f(x) \tag{4.2}$$

since $U_k^n(x)$, the local restriction of U^n to the element k, is already in the space of linear functions on the element. Step (iii) of

§3 then gives the speed of node j as

$$\dot{X}_j = - \frac{\{-U_k^n(X_j)+\mathscr{F}_k f(X_j)+U_{k-1}^n(X_j)-\mathscr{F}_{k-1} f(X_j)\}}{M_k - M_{k-1}} . \tag{4.3}$$

Since U^n is continuous, $U_k^n(X_j) = U_{k-1}^n(X_j)$ and (4.3) reduces to

$$\dot{X}_j = - \frac{\{\mathscr{F}_k f(X_j)-\mathscr{F}_{k-1} f(X_{j-1})\}}{M_k - M_{k-1}} , \tag{4.4}$$

the same expression as (2.2).

The form of $\dot{U}_j$ corresponding to the $\dot{X}_j$ of (4.3) or (4.4) is

$$\dot{U}_j = - U_j + ½(M_k+M_{k-1})\, \dot{X}_j = - \frac{\{M_k \mathscr{F}_{k-1} f - M_{k-1} \mathscr{F}_k f\}}{M_k - M_{k-1}} \tag{4.5}$$

which, for $\Delta t = 1$, corresponds in the case of explicit Euler time-stepping to

$$U_j^{n+1} = ½(M_k + M_{k-1})\dot{X}_j \tag{4.6}$$

so that the two procedures give identical nodal values also (see Appendix).

Therefore the MBF step (2.2), for convex $f(x)$, corresponds to an explicit MFE step with $\Delta t = 1$. Alternatively, we can say that the MFE displacement corresponds to an MBF step with a relaxation factor Δt inserted into the right hand side of (2.2). Similarly, if $\Delta t < 1$ the new MFE U_j is sandwiched between the previous (common) U_j and the next MBF U_j (see Appendix). In either case the MBF or MFE step reduces the L_2 norm of $-u + f(x)$ (see [1]).

Now consider the steady limit of (4.1), ultimately reached by displacement speeds (4.4) (with corresponding $\dot{U}_j$'s). Analogously, the MFE steps (2.2) follow the MBF iteration to convergence, although the path (depending on Δt) may be different. At the common (steady) limit we find simultaneously the MFE steady solution and the best fit to $f(x)$ amongst piecewise linear functions with variable nodes. Observe that, at convergence,

$$\mathcal{P}_k(f(X_j) = \mathcal{P}_{k-1} f(X_j) \tag{4.7}$$

corresponding to the well-known result that the best discontinuous L_2 fit to a convex function with variable nodes is continuous [5].

For the more general equation

$$u_t = G(x,u) \tag{4.8}$$

similar arguments apply so long as the function $G(x,u)$ satisfies certain conditions, e.g. $\frac{\partial G}{\partial u} < 0$, $\frac{\partial G}{\partial x} \neq 0$, and (4.8) has a steady solution. Here MBF steps arise from a variational principle

$$\int F(x,u)dx \tag{4.9}$$

where $G(x,u) = F_u(x,u)$ (see [1], Appendix A).

Jimack [6] has shown that for convex g_{xx} the MFE method applied to the equation

$$u_t = u_{xx} - g_{xx} \tag{4.10}$$

leads in the steady state limit to a best fit U to the function $g(x)$ with adjustable nodes, but this time in the Dirichlet norm, i.e. U_x is a best piecewise constant fit to g_x in the L_2 norm (see [6],[7]). Moreover, in this case the steady state best fit is exact at the nodes.

We may summarise the results as follows. In the limit as $t \to \infty$ we have, in many cases,

(a) if $\mathscr{L}u = - u + f(x)$ the nodes seek a best piecewise linear fit to $f(x)$ with adjustable nodes in the L_2 norm, generalising to $\mathscr{L}u = G(x,u)$. The MBF iteration and MFE procedure with explicit Euler time-stepping track similar paths, reducing $||u - f(x)||_{L_2}$.

(b) if $Lu = u_{xx} - g_{xx}$ the nodes seek a best piecewise linear fit to $g(x)$ with adjustable nodes in the Dirichlet norm.

(In (a), (b), equidistribution of a power of u_{xx}, u_x, respectively, is obtained asymptotically [1],[7]).

We also have, from [8],

(c) if $\mathscr{L}u = H(u_x)$ the nodes follow characteristics approximately (and move in such a way that U_x in an element remains constant in time (i.e. $\dot{M}_j = 0$)).

In the next section we combine these ideas, but first consider another type of projection.

§5 Other Projections and Nodal Movement

Although the L_2 projection is the one everyone uses (following Miller [2]), it is instructive to replace the L_2 projection by the (continuous) linear interpolant in x (see fig. 1). Then, for the equation (4.1), since both U_k and $\mathcal{P}_k f(X_j)$ are continuous, (4.3) gives

$$\dot{X}_j = 0 \tag{5.1}$$

and the method is a fixed node method. Similarly, (5.1) holds for the equation (4.8) when using a linear interpolant of $G(x,u)$.

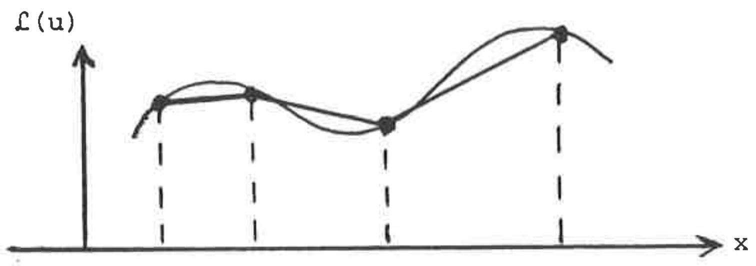

fig. 1

Although it might be argued that in the case of (4.1) or (4.8), with no derivative terms present, zero nodal speeds are natural and expected, the choice of projection makes too much difference for comfort. However, we can make some sense of the situation by considering the equation

$$u_t + H(x,u,u_x) = 0 \ , \tag{5.3}$$

discussed fully in [8], where the L_2 projection is used. Here the nodal speeds are derived as

$$\dot{X}_j = \frac{\mathcal{P}_k H(X,U,M)_j - \mathcal{P}_{k-1} H(X,U,M)_j}{M_k - M_{k-1}} \tag{5.4}$$

where $\mathcal{P}_k H(X,U,M)$ is the L_2 projection of $H(X,U,M)$ into the space of linear functions on element k. Even if the projection is changed so that $\mathcal{P}_k H$ is replaced by the linear interpolant of H in X we still obtain the expression (5.4), which approximates the characteristic speed $\frac{\partial H}{\partial U_x}$. But if u_x had been missing from the function H, as in (4.8), the type of projection would have made a considerable difference to the nodal speed, as we have already seen.

Thus, choosing the linear interpolant projection results in

(a) zero nodal speeds if u_x is absent from $\mathcal{L}u$.

(b) Approximate characteristic nodal speeds if u_x is present.

On the other hand, the L_2 projection in case (a) leads to non-zero nodal speeds which eventually carry U in the steady state into the best fit with adjustable nodes. The L_2 projection in case (b) also leads to a different approximation to the characteristic speed which has a degree of best fitting within it, as argued below.

For the not so special case

$$u_t + \sum_i G_i(x,u) . H_i(u_x) = 0 \tag{5.5}$$

of (5.3), the projection of a typical term of $\mathcal{L}u$ into element k is

$$\mathcal{P}_k\{G(X,U)\ H(U_x)\} = H(U_x)\ \mathcal{P}_k G(X,U) \ , \tag{5.6}$$

whether linear interpolant or L_2 best fit (since U_x is constant in each element).

In the case of the linear interpolant projection we have

$$\mathcal{P}_k\{GH\} - \mathcal{P}_{k-1}(GH) = \{H(M_k) - H(M_{k-1})\}G(X_j,U_j) \tag{5.7}$$

giving a nodal speed

$$\dot{X} = G(X_j,U_j) \cdot \frac{\{H(M_k) - H(M_{k-1})\}}{M_k - M_{k-1}} \ , \tag{5.8}$$

an approximate characteristic speed. For the L_2 projection the corresponding nodal speed may be written

$$\dot{X}_j = \bar{\mathcal{P}}_j\ G(X,U)\ \frac{[H(M_k) - H(M_{k-1})]}{M_k - M_{k-1}} + \bar{H}_j\ \frac{[\mathcal{P}_k G(X,U) - \mathcal{P}_{k-1}G(X,Y)]}{M_k - M_{k-1}} \tag{5.9}$$

where $\bar{\mathcal{P}}_j = ½(\mathcal{P}_{k-1} + \mathcal{P}_k)$ and $\bar{H}_j = ½[H(M_{k-1}) + H(M_k)]$. In this case the speed is made up of two terms, the first being an approximate characteristic speed and the second being a speed associated with nodes moving towards a variable node L_2 best fit in the steady limit (if there is one) (see fig. 2). (Note that the second term vanishes if $G(x,u)$ is independent of x and u or linear in x and/or u.)

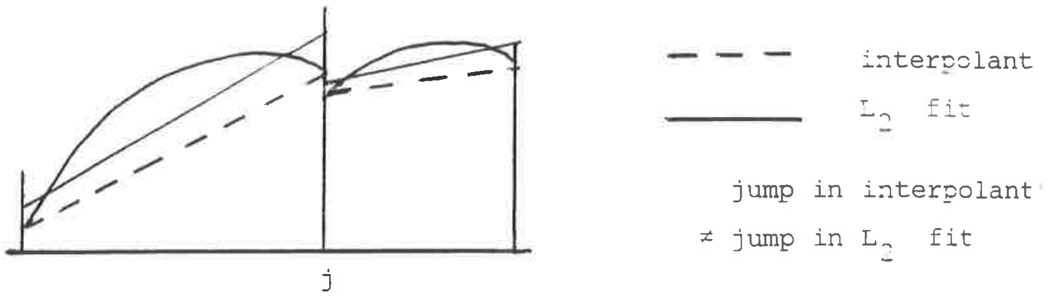

fig. 2.

Thus, unlike the example in §4, the near-characteristic nodal speeds arising from the dependence of $\mathcal{L}u$ on u_x are probably not as sensitive to the type of projection used, the L_2 projection introducing an element of best fitting.

These results complement the result of Morton (see [9]) that, if $H(x,u,u_x) = G(u)u_x$ the MFE procedure for (5.3) carries the best L_2 fit asymptotically.

The choice of projection is even starker in the case of the heat equation

$$u_t = u_{xx} \tag{5.10}$$

where, in contrast to the speeds obtained by the L_2 projection [8], (which is equivalent to recovery of u by local Hermite cubics [10]), the recovery of u by a cubic spline (corresponding to a twice integrated linear interpolant for u_{xx}) leads to zero nodal velocities.

Again the nodal speeds are highly sensitive to the type of projection used.

§6. Conclusions

We have seen that the MFE algorithms and the MBF procedure for the equation

$$u_t = - u + f(x),$$

where $f(x)$ is convex, are identical if explicit time-stepping with $\Delta t = 1$ is used or if a relaxation factor of Δt is introduced into the appropriate step of the MBF algorithm. In each case the key step is an L_2 projection of the function $f(x)$ into the space of piecewise linear discontinuous functions, minimising

$$\int \{\mathcal{P}f - f(x)\}^2 \, dx \qquad (6.1)$$

where $\mathcal{P}f$ is the projection, which may be carried out elementwise. More generally, if (6.1) is replaced by a variational principle, and the minimisation replaced by seeking an extremal of

$$\int F(x,u)dx, \qquad (6.2)$$

where $G(x,u) = \frac{\partial F}{\partial u}(x,u)$ satisfies certain conditions, the "equivalent" MFE algorithm is that applied to the equation

$$u_t = \pm \, G(x,u),$$

where the sign is chosen to facilitate convergence to the steady state solution.

The question of the type of projection used is important in this context. If the L_2 projection is replaced by the continuous linear interpolant in X (where X.U are both piecewise linear continuous functions), there is no jump in $\mathscr{G}G(X,U)$ at a node and hence the nodal speed is zero. In other words, the nodal speed is a consequence of the fact that the projection is L_2.

If on the other hand the operator $\mathscr{L}(u)$ contains u_x the presence of U_x ensures a jump in the projection, whether L_2 or linear interpolant. Both give rise to a nodal speed which approximates the characteristic speed. However, the linear interpolant projection is simpler, avoiding that part of the nodal speed generated purely by the L_2 projection.

Finally we observed that for the linear heat equation recovery of the piecewise linear U by a Hermite cubic in each element leads to the nodal speed obtained from the standard L_2 projection, while recovery by a continuous cubic spline leads to zero nodal speeds. When a convex source term g_{xx} is present the nodes move in the standard method to a steady state in which U gives the best fit to g(x) in the Dirichlet norm with adjustable nodes [6].

We conclude that L_2 projection is not essential to the MFE procedure (for example, linear interpolant projection will give approximate characteristic speeds), but if it is used it has the capability of inducing a best linear fit with adjustable nodes to an associated function in the steady state and will in many cases move towards that best fit as time progresses.

§7 References

[1] Baines M.J. & Carlson N.N. (1989). On Best Piecewise Linear L_2 Fits with Adjustable Nodes. Numerical Analysis Report 6/90. Department of Mathematics, University of Reading.

[2] Miller K. & Miller R.N. (1981). Moving Finite Elements. Part I SIAM J. Num. An. 18, 1019-1032: Miller K, Moving Finite Elements, Part II, SIAM J. Num. An. 18, 1033-1057.

[3] Wathen A.J. & Baines M.J. (1985). On the Structure of the Moving Finite Element Equations. IMAJ. Num. An. 5, 161-182.

[4] Baines M.J. & Wathen A.J. (1988). Moving Finite Element Methods for Evolutionary Problems I. Theory. J. Comput. Phys. 79, 245-269.

[5] Loach P.D. & Wathen A.J. (198). On the Best Least Squares Approximation of Continuous Functions using Linear Splines with Free Knots. IMA J. Num. An. (to appear).

[6] Jimack P. (1990). Large Time Solutions of the Moving Finite Element Equations for Linear Diffusion Problems in One Dimension. Preprint. Department of Mathematics, University of Leeds.

[7] Baines M.J. (1989). On Best Piecewise Constant L_2 Fits with Adjustable Nodes. Numerical Analysis Report 12/90. Department of Mathematics, University of Reading.

[8] Baines M.J. (1991). An Analysis of the Moving Finite Element Method. SIAM J. Num. An. (to appear).

[9] Wathen A.J., Baines M.J. & Morton K.W. (1984). Moving Finite Element Methods for the Solution of Evolutionary Equations in One and Two Dimensions. In MAFELAP V, Proc. of MAFELAP Conference, Brunel 1984. pp. 421-430 Academic Press (1985).

[10] Johnson I.W., Wathen A.J. & Baines M.J. (1988). Moving Finite Element Methods for Evolutionary Problems II Applications. J. Comput. Phys. 79, 270-297.

Appendix

Suppose that at iteration level or time level n the nodal positions and nodal values are X_j^n and U_j^n. MBF seeks

$$\min_{U^{n+1},X^{n+1}} \int \{-U^{n+1} + f(X^{n+1})\}^2 \, dx \tag{A.1}$$

which, as far as the next new nodal value is concerned, reduces in the i'th iteration to

$$\min_{(U^{n+1})^{i+1}} \int \{(-U^{n+1})^{i+1} + f((X^{n+1})^i)\}^2 \, dx \; . \tag{A.2}$$

MFE seeks

$$\min_{U_t} \int \{U_t + U^n - f(X^n)\}^2 \, dx \tag{A.3}$$

which (with X fixed and explicit Euler time-stepping) is equivalent to seeking

$$\min_{U^{n+1}} \int \left\{\frac{U^{n+1} - U^n}{\Delta t} + U^n - f(X^n)\right\}^2 \; . \tag{A.4}$$

With $\Delta t = 1$ (A.4) reduces to (A.2), U^{n+1} becoming (U^{n+1}) and X^n becoming $(X^{n+1})^i$.

If $\Delta t \neq 1$, (A.4) can be written

$$\min_{V^{n+1}} \int \left\{ V^{n+1} - f(X^n) \right\}^2 dx \tag{A.5}$$

where

$$V^{n+1} = \frac{U^{n+1}}{\Delta t} + \left[1 - \frac{1}{\Delta t}\right] U^n \tag{A.6}$$

or, if $V^n = U^n$,

$$U^{n+1} = \Delta t \; V^{n+1} + (1-\Delta t) V^n \tag{A.7}$$

Thus, comparing (A.5) with (A.2), if $\Delta t < 1$ the MFE nodal value in a step Δt corresponds to an averaged (relaxed) MBF iteration step.