THE UNIVERSITY OF READING

AN MFE-LIKE METHOD FOR BEST APPROXIMATION

by

M J Baines

Numerical Analysis Report 7/95

DEPARTMENT OF MATHEMATICS

AN MFE-LIKE METHOD FOR BEST APPROXIMATION

by

M J Baines

Numerical Analysis Report 7/95

The University of Reading
Department of Mathematics
P O Box 220,
Reading RG6 2AX
Berkshire, UK

Abstract

Two approaches to node movement are contrasted. In the Moving Finite Element (MFE) method for the approximate solution of PDEs, the equation residual is minimised over a continuous Lagrangian solution and nodal speeds. In the Moving Best Fits (MBF) method the error residual is minimised over a discontinuous Eulerian solution and nodal positions. A comparison is drawn through an intermediate method (called here Adjustable Finite Elements (AFE)). It is shown how to modify the AFE method to avoid singularities and how to apply the method to time dependent problems.

1. The MFE and MBF methods

The key constituents of the Moving Finite Element (MFE) procedure for the approximate solution of PDEs of the form

$$u_t = \mathcal{L}u \tag{1.1}$$

are (in 2-D)

(i) a continuous piecewise linear (CPL) representation U of u

(ii) a search for nodal velocities $\dot{U}, \dot{X}, \dot{Y}$, for which

$$\dot{U} = U_t + U_X \dot{X} + U_Y \dot{Y}, \tag{1.2}$$

where $\dot{X}, \dot{Y}$ are also CPL functions

(iii) minimisation of the L_2 norm of the residual $U_t - \mathcal{L}U$ (where the definition of $\mathcal{L}U$ may involve some recovery of smoothness) over $\dot{U}, \dot{X}$ and $\dot{Y}$.

Note particularly that

(a) U is continuous

(b) $\mathcal{L}U$ may be discontinuous (even after recovery)

(c) the goal is to find the speeds $\dot{U}, \dot{X}, \dot{Y}$

(d) $\dot{U}, \dot{X}, \dot{Y}$ are found in one combined minimisation step.

(Grid displacement is achieved subsequently by a finite difference time-stepping scheme.)

By contrast, the key elements of the MBF (Moving Best Fits) method to find best L_2 fits U with adjustable nodes (free knots) to a continuous function $f(x,y)$ are (again in 2-D)

(i) a piecewise linear discontinuous (DPL) representation of the best fit U

(ii) a search for U and the grid, defined by CPL functions X,Y

(iii) minimisation of the L_2 norm of the residual $U - f(x,y)$ over the unknowns U,X,Y.

Note that in this method

(a) U is discontinuous (although X,Y are continuous)

(b) $f(X,Y)$ is continuous

(c) the goal is to find the solution and grid U,X,Y

(d) U,X,Y are found by an iteration procedure, the equations to be solved being nonlinear.

Here the obvious practical disadvantage of using a discontinuous U is offset by the capability of avoiding the singularities inherent in the MFE method using a continuous $\dot{U}$. The discontinuities are in any case non-zero or small a.e. and easily smoothed to give a continuous U.

In this report we discuss these two methods, in particular comparing them with an intermediate method (AFE) which uses the MFE technique to generate adaptive best fits.

2. Adjustable Finite Elements (AFE)

Consider again the problem of finding the best fit U with adjustable nodes (X,Y) to the continuous function $f(x,y)$, but this time using an MFE-like approach. The ingredients are

(i) a CPL representation of U

(ii) use of displacements $\delta U, \delta X, \delta Y$ for which

$$\delta U = \Delta U + U_X \delta X + U_Y \delta Y \tag{2.1}$$

(iii) minimisation of the norm of a residual which involves ΔU, the natural choice being the norm of $U + \Delta U - f(X,Y)$,

i.e. $$\|\Delta U - [f(X,Y) - U]\|. \tag{2.2}$$

The goal here is to find the displacement ΔU that approximates $f(X,Y) - U$, which involves U. The method is therefore to be regarded as an updating procedure, giving $\delta U, \delta X, \delta Y$ from known functions U, X, Y, and it thus has the same iterative character as the MBF method. By running the iteration to convergence, so that the updates $\delta U, \delta X, \delta Y$ are driven to zero, a best fit U to $f(X,Y)$ is found.

Each iterative step is a single combined minimsation step for $\delta U, \delta X$ and δY. Note particularly that, in contrast to MFE, the function $f(X,Y) - U$ to be approximated is continuous.

The AFE method therefore holds an intermediate position between MFE and the MBF method. It has the same equation structure as MFE (for each iteration step) but also suffers from the same drawbacks of MFE, namely the singularities of parallelism and node overtaking, and therefore needs either regularisation or singularity removal. On the other hand it gives a useful prescription for obtaining best fits in two and higher dimensions (of which more later). If it does not blow up, this MFE-like method iterates towards a best fit to $f(x,y)$ (with $\dot{U}, \dot{X}$ and $\dot{Y}$ in MFE replaced by $f(X,Y) - U, \delta U, \delta X$, and δY, respectively).

It is instructive to compare the AFE method with MFE and its finite difference time-stepping mechanism. The validity of (2.1) limits the size of $\delta U, \delta X, \delta Y$ in the same way that (1.1) is only accurate for small time steps. For larger steps, accuracy is lost in both the AFE and the MFE methods. However, in the AFE method this does not matter since it is only the limit that is required.

In an implicit implementation of MFE the driving term $\mathcal{L}U$ of (1.1) is evaluated at the forward time (thus depending on $\dot{U}, \dot{X}$ and $\dot{Y}$) and minimisation of the norm of the equation residual will involve the variation of this term. This is usually ignored in implicit MFE, however.

3. The Two Step Form of MFE and AFE

Let us now look at the two methods MFE and AFE in more detail on an algebraic level.

As is well-known [1], the MFE method may be written as a two step method in which the first step is a projection of $\mathcal{L}U$ into the space of DPL functions. We shall express this step as

$$\dot{W} = \mathcal{P}_{DPL}\mathcal{L}u \tag{3.1}$$

The second step is to convert the DPL function $\dot{W}$ into nodal and solution speeds using

$$\dot{U} - U_X\dot{X} - U_Y\dot{Y} = \dot{W}. \tag{3.2}$$

c.f.(1.2). Applying (3.2) at each node j gives the set of equations

$$\dot{U}_j - (U_X)_k\dot{X}_j - (U_Y)_k\dot{Y}_j = \dot{W}_k, \tag{3.3}$$

one for each element k surrounding the node j. In the MFE method the values $\dot{U}_j, \dot{X}_j, \dot{Y}_j$ are obtained by a weighted least squares solution of (3.3). There is no iteration in the MFE method, although, once again, we recall that if the resulting time-dependent ODEs are to be solved implicitly then $\mathcal{L}U$ in (3.1) should be evaluated at the advanced time and this can only be done through iterating on (3.1) and (3.3).

In the AFE method the corresponding form of (3.1) is

$$\delta W = \mathcal{P}_{DPL}f(X,Y) - U = V - U, \text{ say} \tag{3.4}$$

for the DPL function δW replacing $\dot{W}$. Here V is used to denote the projection of $f(X,Y)$ and we note that the corresponding form of (3.3) is

$$\delta U_j - (U_x)_k \delta X_j - (U_y)_k \delta Y_j = \delta W_k \tag{3.5}$$

for each element k surrounding node j, which may also be solved for $\delta U_j, \delta X_j, \delta Y_j$ by a weighted least squares procedure. In the AFE algorithm this two step form is iterated to convergence.

To clarify presentation we shall write M and N for U_x and U_y in future. Then (3.5) becomes

$$\delta U_j - M_k \delta X_j - N_k \delta Y_j = \delta W_k . \tag{3.6}$$

The MFE approach to solving the sets of equations (3.6) is to seek a least squares solution with a matrix weight (for details see [1]). A simplified version which uses only area A_k weighting gives a local MFE method. When applied to (3.6) this procedure gives the matrix equation

$$\begin{pmatrix} \Sigma A_k & -\Sigma A_k M_k & -\Sigma A_k N_k \\ -\Sigma A_k M_k & \Sigma A_k M_k^2 & \Sigma A_k M_k N_k \\ -\Sigma A_k N_k & \Sigma A_k M_k N_k & \Sigma A_k N_k^2 \end{pmatrix} \begin{pmatrix} \delta U_j \\ \delta X_j \\ \delta Y_j \end{pmatrix} = \begin{pmatrix} -\Sigma A_k \delta W_k \\ -\Sigma A_k M_k \delta W_k \\ -\Sigma A_k N_k \delta W_k \end{pmatrix}. \tag{3.7}$$

Provided that the left-hand-side matrix is non-singular, the unknowns $\delta U_j, \delta X_j, \delta Y_j$ may be obtained quite simply, by for example Cramer's Rule.

Likewise in the ABF method there are also two distinct steps to each iteration. In the first step the function $f(X,Y)$ is projected into the space of DPL functions, giving V (c.f. (3.4)). The second step is distinct from (3.5), however, and may be written

$$\left| V_{jk} + (V_X)_k \delta X_j + (V_Y)_k \delta Y_j - f(X_j, Y_j) \right| = C_j , \tag{3.8}$$

where C_j is a constant independent of k and V_{jk} is the value of the discontinuous

function V at node j in element k. The new value of U_j is then given by

$$U_{jk} = V_{jk} + (V_X)_k \delta X_j + (V_Y)_k \delta Y_j \tag{3.9}$$

which may be multi-valued at node j. In practice a simple average of the U_{jk} can be used to give a single valued U_j.

We see that (3.8) has the same general form as (3.5) but with at least two important differences, the multi-valued nature of V and the presence of the modulus signs.

There is also a DPC version of the MBF method which differs from the DPL case in two respects. Firstly, the projection (3.4) becomes

$$\delta W = \mathcal{P}_{DPC} f(X,Y) - U = V - U \tag{3.10}$$

and is a DPC function. Secondly, the quantities V_X and V_Y disappear from (3.8), leaving

$$\left| V_{jk} - f(X_j, Y_j) \right| = C_j . \tag{3.11}$$

In future we shall write P,Q for V_x, V_y. Then (3.8) becomes

$$\left| V_{jk} + P_k \delta X_j + Q_k \delta Y_j - f(X_j, Y_j) \right| = C_j \tag{3.12}$$

4. A Strategy for Choosing Between MBF Solutions in 1-D and 2-D

Solutions of (3.12) depend on which sign is taken for the modulus.

It is instructive to consider first the 1-D case for which (3.12) becomes

$$\left| V_{jL} + P_L \delta X_j - f(X_j + \delta X_j) \right| = C_j \tag{4.1}$$

$$\left| V_{jR} + P_R \delta X_j - f(X_j + \delta X_j) \right| = C_j \tag{4.2}$$

since k is either L (left) or R (right). There are only two possibilities for the signs, since either the arguments inside the two moduli have the same sign or opposite signs. In the former case, removing the modulus signs and subtracting gives

$$-[P]\delta X_j = [V_j], \tag{4.3}$$

where

$$[.] = (.)_R - (.)_L, \tag{4.4}$$

irrespective of f, from which U_{jL} and U_{jR} may be calculated from (3.9) (omitting δY_j). The procedure breaks down when $[P] = 0$. If $[P] \neq 0$ it follows from (4.3) and (3.9) that $U_{jL} = U_{jR}$.

In the other case we remove the modulus signs but introduce a negative sign multiplying the left hand side of (4.2). Then, by subtraction,

$$V_{jL} + V_{jR} + (P_L + P_R)\delta X_j - 2f(X_j + \delta X_j) = 0, \tag{4.5}$$

to be solved for X_j. In the this case $U_{jL} \neq U_{jR}$ in general but, from (3.9) and (4.1-4.2),

$$U_{jL} - f(X_j + \delta X_j) = -(U_{jR} - f(X_j + \delta X_j)) \tag{4.6}$$

so that a simple average of the U_{jk}'s gives the new sampled value $f(X_j + \delta X_j)$ itself.

In practice it is often convenient to freeze $f(X_j)$ in solving (4.5) for δX_j. Moreover, a useful approximate guide as to whether to choose the solution of (4.3) or (4.5) for δX_j is also obtained by freezing, namely the signs of the frozen (or lagged) values of $V_{jL} - f(X_j)$ and $V_{jR} - f(X_j)$ in (4.1) and (4.2). This choice discriminates between solutions in virtually the same way as in [2], where both node overtaking and parallelism are avoided. If we adopt this strategy, equations (4.1) and (4.2) may be written

$$s_{jL}(V_{jL} + P_L \delta X_j - f(X_j)) = C_j \tag{4.7}$$

$$s_{jL}(V_{jR} + P_R \delta X_j - f(X_j)) = C_j \tag{4.8}$$

where

$$s_{jk} = \text{sgn}(V_{jk} - f(X_j)). \tag{4.9}$$

Another form of (4.7), (4.8) is

$$C_j - s_{jL} P_L \delta X_j = \left| V_{jL} - f(X_j) \right| \tag{4.10}$$

$$C_j - s_{jR} P_R \delta X_j = \left| V_{jR} - f(X_j) \right| \tag{4.11}$$

from which

$$-[s_j P] \delta X_j = [\left| V_{jk} - f(X_j) \right|]. \tag{4.12}$$

c.f. (4.3).

A similar argument in 2-D yields the equations

$$C_j - s_{jk} P_k \delta X_j - s_{jk} Q_k \delta Y_j = \left| V_{jk} - f(X_j, Y_j) \right| \tag{4.13}$$

for all elements k surrounding node j, where

$$s_{jk} = \text{sgn}\{V_{jk} - f(X_j, Y_j)\}. \tag{4.14}$$

Equations (4.13) may be solved for an averaged solution in the least squares manner of (3.7), giving

$$\begin{pmatrix} \Sigma A_k & -\Sigma s_{jk} P_k & -\Sigma A_k s_{jk} Q_k \\ -\Sigma A_k s_{jk} P_k & \Sigma A_k s_{jk}^2 P_k^2 & \Sigma A_k s_{jk}^2 P_k Q_k \\ -\Sigma A_k s_{jk} Q_k & \Sigma A_k s_{jk}^2 P_k Q_k & \Sigma A_k s_{jk}^2 Q_k^2 \end{pmatrix} \begin{pmatrix} C_j \\ \delta X_j \\ \delta Y_j \end{pmatrix} =$$

$$\begin{pmatrix} \Sigma A_k s_{jk} \left| V_{jk} - f(X_j, Y_j) \right| \\ -\Sigma A_k s_{jk} P_k \left| V_{jk} - f(X_j, Y_j) \right| \\ -\Sigma A_k s_{jk} Q_k \left| V_{jk} - f(X_j, Y_j) \right| \end{pmatrix}. \tag{4.15}$$

When δX_j and δY_j have been found, U_{jk} is given by (3.9) and C_j is dispensible.

5. Modification of AFE and MFE Methods

In the same way, the AFE method may be modified to remove its susceptibility to the parallelism singularity. Writing (3.6) as

$$\delta U_j - M_k \delta X_j - N_k \delta Y_j = V_{jk} - U_j \tag{5.1}$$

or

$$V_{jk} + M_k \delta X_j + N_k \delta Y_j - f(X_j, Y_j) = U_j + \delta U_j - f(X_j, Y_j), \tag{5.2}$$

c.f. (4.1)-(4.2), the same argument as in §4 leads to modification of (5.2) to

$$C_j - s_{jk} M_k \delta X_j - s_{jk} N_k \delta Y_j = \left| V_{jk} - f(X_j, Y_j) \right| \tag{5.3}$$

where, as in (4.14),

$$s_{jk} = \operatorname{sgn}\{V_{jk} - f(X_j, Y_j)\} \tag{5.4}$$

and

$$C_j = U_j + \delta U_j - f(X_j, Y_j). \tag{5.5}$$

The set of least squares normal equations for (5.3) are the same as equation (4.15) with P,Q replaced by M,N. This time C_j also needs to be calculated to obtain the CPL function δU_j from (5.5).

The method is still iterative in that new U_j, X_j, Y_j values are calculated and the step is repeated, as in the original from the AFE method.

From this point it is apparently a short step to a modified MFE method in which V_{jk} is replaced by $\dot{W}_{jk}$, the increments $\delta U, \delta X, \delta Y$ are replaced by $\dot{U}, \dot{X}, \dot{Y}$, and $f(X,Y) - U$ is replaced by $\mathcal{L}U$. But the MFE method is not iterative, the velocities being obtained in one step, so there is a significant difference. An iterated MFE method would be one in which

increments in $\dot{U}, \dot{X}, \dot{Y}$ were calculated rather than $\dot{U}, \dot{X}, \dot{Y}$ themselves. This is consistent with an implicit time stepping approach. Rather than construct such a method, modify it in the manner above and then superimpose time stepping on top of that, it is possible to do the time stepping first and then apply the AFE method. The MBF method is discussed fully in [1].

6. Conclusion

We have discussed a method (AFE) intermediate between the MFE method and MBF methods and used the latter to construct a strategy for avoiding the singularities inherent in both the MFE and AFE methods.

7. References

[1] **Baines, M.J**. *Moving Finite Elements,* OUP (1994)

[2] **Baines, M.J.** *Algorithms for Optimal Discontinuous Piecewise Linear and Constant L_2 Fits to Continuous Functions with Adjustable Nodes in One and Two Dimensions.* Math. Comp. Vol. 62, 645-669 (1994).