COO-2383-0014

DEPARTMENT OF COMPUTER SCIENCE
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN
URBANA, ILLINOIS 61801

MULTIRATE METHODS FOR ORDINARY DIFFERENTIAL EQUATIONS

by

C. W. Gear

*Supported in part by the Atomic Energy Commission under grant US AEC AT(11-1)2383

UIUCDCS-F-74-880
September, 1974

DISCLAIMER

ABSTRACT

Multirate methods have often been considered for integrating systems in which one part changes much more rapidly than another, but, to my knowledge, little theoretical study has been made on the accuracy and stability of such methods. These few comments are intended to suggest some of the problems that might be involved, and some ways in which the methods can be analyzed.

1. INTRODUCTION

Suppose we have the system of differential equations

$$y' = f(y,z,t)$$

$$z' = g(y,z,t)$$

where y and z are two functions of time (they could be vectors, but we will consider scalars for notational convenience). Suppose, also, that y is a rapidly varying function in the sense that if we knew the values of z for any t, we would need a small step size to integrate $y' = f(y,z(t),t)$ by any of the methods under consideration, whereas, if we knew y(t) we could use a much larger step to integrate $z' = g(y(t),z,t)$. In that case, a multirate method may be reasonable.

To simplify the discussion, let us consider a constant step size h for integrating y' and a constant step size kh for integrating z', where k is an integer. (It is not clear whether a non-integral k materially affects either the analysis or the implementation.) For this case, a multirate method can be defined in terms of a standard method as follows. Let the standard method for $y' = f(y,t)$ be given by

$$\underline{y}_{n+1} = M(\underline{y}_n, t_n, h; f) \tag{1}$$

Here, $\underline{y}_n$ represents the set of information saved about y(t) at $t = t_n$ where $t_n = t_0 + nh$. For a one-step method, $\underline{y}_n$ is simply y_n, the computed approximation to $y(t_n)$, but for a multistep method, $\underline{y}_n$ may include $y_{n-1},\ldots,y_{n-k+1}$ and $hy'_n,\ldots,hy'_{n-k+1}$.

Now, let

$$\hat{f}(y,t) = f(y,\hat{z}(t),t) \tag{2}$$

$$\hat{g}(z,t) = g(\hat{y}(t),z,t) \tag{3}$$

where $\hat{z}(t)$ is any function approximating z(t) (the solution of the original

problem) that we can compute from the saved information $\underline{z}_{jk}$ for $t_0 \leq t_{jk} \leq t$. Similarly, $\hat{y}(t)$ is an approximation to $y(t)$. Then one type of multirate method is given by

$$\underline{y}_{nk+i+1} = M(\underline{y}_{nk+i}, t_{nk+i}, h; \hat{f}) \qquad i=0,\ldots,k-1$$

$$\underline{z}_{(n+1)k} = M(\underline{z}_{nk}, t_{nk}, kh; \hat{g}) \tag{4}$$

This means that we integrate for y using a step size h, and evaluate $f(y,z,t)$ wherever needed by interpolating for z. This is reasonable if z is slowly varying as, by definition of slowly varying in this case, the interpolation using the saved values in $\underline{z}$ is reasonably accurate in order that an integration formula based on the same values is reasonably accurate. The integration for z may need to interpolate for y values when g is to be evaluated at off-step points (if, for example, a Runge-Kutta method is used). There are also an adequate number of y values available for this.

Although we have indicated that the same method should be used for both equations, this is obviously not necessary.

Example

If Euler's method is used, we could perform the following sequence of operations

$$y_{nk+i+1} = y_{nk+i} + h\, f(y_{nk+i}, \hat{z}_{nk+i}, t_{nk+i}) \tag{5}$$

$$\hat{z}_{nk+i+1} = z_{nk} + (i+1)h\, g(y_{nk}, z_{nk}, t_{nk})$$

$$i=0,1,\ldots,k-1$$

$$z_{nk+k} = \hat{z}_{nk+k}$$

This would save time if the evaluation of g is very expensive as it is evaluated only 1/k the number of times as would be used in a regular Euler method.

2. ANALYSIS

The traditional error analysis is straightforward, although more messy than in a standard method. The most direct approach is to define the local truncation error $[\underline{d}_y, \underline{d}_z]$ in terms of the actual solution $[y(t), z(t)]$ as follows.

Let the correct values of $\underline{y}_{nk}$ and $\underline{z}_{nk}$ be denoted by $\underline{y}(t_{nk})$ and $\underline{z}(t_{nk})$. Starting from these values, compute

$$\tilde{\underline{y}}_{nk} = \underline{y}(t_{nk}) \tag{6}$$

$$\tilde{\underline{z}}_{nk} = \underline{z}(t_{nk})$$

$$\tilde{\underline{y}}_{nk+i+1} = M(\tilde{\underline{y}}_{nk+i}, t_{nk+i}, h; \tilde{f})$$

$$i=0,\ldots,k-1$$

where

$$\tilde{f}(y,t) = f(y, \tilde{z}(t), t)$$

and $\tilde{z}(t)$ is the approximation to $z(t)$ using the exact rather than computed values of z. Similarly, compute

$$\tilde{z}_{nk+k} = M(\tilde{\underline{z}}_{nk}, t_{nk}, kh; \tilde{g})$$

and define

$$(\underline{d}_y)_{nk} = \tilde{\underline{y}}_{nk+k} - \underline{y}(t_{nk+k}) \tag{7}$$

$$(\underline{d}_z)_{nk} = \tilde{\underline{z}}_{nk+k} - \underline{z}(t_{nk+k})$$

Thus, $[\underline{d}_y, \underline{d}_z]$ is the error introduced in a set of k steps of length h (that is, in one cycle). As long as k is independent of h, it is possible to use standard Taylor series expansions to show that the local truncation error is of the form

$$h^{p+1}[\underline{\phi}_y(t), \underline{\phi}_z(t)] + O(h^{p+2})$$

provided that the solution is sufficiently smooth (in C_{p+2}). $\underline{\phi}_y$ and $\underline{\phi}_z$ will be combinations of derivatives of the solution, and hence bounds of the form

$$||[\underline{d}_y, \underline{d}_z]|| \leq D\, h^{p+1}$$

can be obtained, where D is a combination of bounds on the derivatives.

Example

Consider the previous example in equations (5), and let k=2. For convenience, let n=0 and let y_0, y_0', etc., be the true values of $y(t_0)$, $y'(t_0)$, etc. Then,

$$y_1 = y_0 + hy_0'$$

$$y_2 = y_0 + hy_0' + h\, f(y_0+hy_0',\ z_0+hz_0', t_0+h)$$

$$z_2 = z_0 + 2h\, z_0'$$

Hence,

$$\underline{d}_y = y_2 - y_0 - 2hy_0' - 2h^2 y_0'' + O(h^3)$$

$$\underline{d}_z = z_2 - z_0 - 2hz_0' - 2h^2 z_0'' + O(h^3)$$

or

$$\underline{d}_y = -\, h^2 y_0'' + O(h^3)$$

$$\underline{d}_z = -\, 2h^2 z_0'' + O(h^3)$$

(In fact, for general k, we will get

$$\underline{d}_y = -k\, h^2 y_0''/2 + O(h^3)$$

$$d_z = -\, (kh)^2 z_0''/2 + O(h^3)$$

for Euler's method.)

The order of a multirate method can be determined easily from the

order of the underlying method(s) and the interpolation formula(s) used. If the order of the integrat n method M is p_M in the sense that, for $y' = f(y)$ with y sufficiently smooth

$$y(t_{n+1}) - M(y(t_n),t_n,h;f) = O(h^{p_M+1}),$$

and if the order of the interpolation formula used is p_I in the sense that

$$\tilde{z}(t) - z(t) = O(h^{p_I})$$

when $\tilde{z}$ interpolates z over an interval kh, then order of the multirate method is

$$p = \min(p_M, p_I)$$

Note that the interpolation method can have a local error one power of h less than that of the integration method. This is permissible because interpolation is used to compute $\hat{z}$ (and $\hat{y}$) which are used in terms of the form $hf(y,\hat{z},t)$ and $hg(\hat{y},z,t)$ so that another power of h is included. (At least, this is true in all methods to my knowledge. To prove the above result, we must place some formal restrictions on M, for example, M must be Lipschitz continuous with respect to its arguments and must be "h-Lipschitz continuous" with respect to f, that is, there must exist an L such that

$$||M(y,t,h;f) - M(y,t,h;\tilde{f})|| \leq |h|\ L\ ||f - \tilde{f}||.$$

Such conditions are trivially obtainable for all practical methods.)

In the example above, $p_M = 1$ but $p_I = 2$. We could have used $\hat{z}_1 = z_0$ to evaluate

$$y_2 = y_0 + hy_0' + hf(y_0 + hy_0',\hat{z}_1,t_0 + h)$$

in which case the truncation error in y would have been

$$\underline{d}_y = -h^2 y_0'' - h^2\ f_z\ z_0'$$

If the multirate method is consistent ($p \geq 1$) and stable (a condition

that hasn't been defined yet!), it will converge as $h \to 0$ for fixed k (or even for bounded k) - but that will be left as an exercise for the inquisitive reader, as the concern of multirate methods is to take large rather than small steps. Instead, we want to look at stability for non-zero h, that is, absolute stability and associated concepts.

Stability

Absolute stability is usually discussed by considering the test equation $y' = \lambda y$. To look at multirate methods, we must study a system of equations, for example

$$\begin{bmatrix} y' \\ z' \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} y \\ z \end{bmatrix} + \begin{bmatrix} f(t) \\ g(t) \end{bmatrix} \qquad (7)$$

Since our assumption is that y is rapidly changing while z is not, it seems reasonable to assume that y is not strongly coupled into the equation for z, or, in other words, that $a_{21} = 0$. Hence, let us first restrict ourselves to the problem

$$\begin{bmatrix} y' \\ z' \end{bmatrix} = \begin{bmatrix} \lambda_y & \mu \\ 0 & \lambda_z \end{bmatrix} \begin{bmatrix} y \\ z \end{bmatrix} \qquad (8)$$

λ_y and λ_z are the eigen values of the system, and are characteristic values associated with the two components.

Example

Consider the multirate Euler method again. A little calculation shows that

$$y_{nk+k} = (1+h\lambda_y)^k y_{nk} + h\mu\left[\frac{(1+h\lambda_y)^k - 1}{h\lambda_y} + h\lambda_z \frac{(1+h\lambda_y)^k - kh\lambda_y - 1}{(h\lambda_y)^2}\right] z_{nk} \qquad (9)$$

$$z_{nk+k} = (1+hk\lambda_z) z_{nk} \qquad (10)$$

This is absolutely stable if $|1 + h\lambda_y| < 1$ and $|1 + hk\lambda_z| < 1$. In other words, it is absolutely stable if the Euler method is absolutely stable for the uncoupled equations, using a step of h for y and kh for z. If we had used a backward Euler method for z (as would make sense if the z component were stiff but no longer in its transient region), then equation (10) would be replaced by

$$z_{nk+k} = (1-hk\lambda_z)^{-1} z_{nk} \qquad (11)$$

and the pair (9) and (11) would be absolutely stable for any $Re(\lambda_z) < 0$ and $|1+h\lambda_y| < 1$.

It is quite obvious that this is a general result, namely

Theorem

If a method M_1 is used for y with step size h and M_2 is used for z with step size kh, then the region of absolute stability for the multirate method applied to equation (8) includes the intersections of the strict regions of absolute stability* for M_1 and M_2 provided that bounded values of $\underline{z}$ lead to bounded values of $\hat{z}(t)$.

Proof

The equation for z in (8) is not coupled to that for y, hence, inside the absolute stability region for M_2, the solution for z decays (after a finite number of starting steps). The numerical solution for y is obtained by applying the method M_1 to

$$y' = \lambda_y y + \hat{z}(t)$$

Since $\hat{z}$ is bounded, an absolutely stable method will yield a bounded solution.

---

*By strict absolute stability, we mean that the numerical solution decays for the test equation $y' = \lambda y$.

(Some additional hypotheses on M_1 are needed at this point to make this statement, but are not stated to avoid going into needless detail. The statement is true for all common methods - or one could simply make this the definition of absolute stability since we haven't defined it for other methods!)

In practice, $\frac{\partial g}{\partial y}$ is likely to be small rather than zero, and this could have serious affects on stability of multirate methods. Suppose we use the Euler method of earlier examples on

$$\begin{bmatrix} y' \\ z' \end{bmatrix} = \begin{bmatrix} \lambda_y & \mu \\ \varepsilon & \lambda_t \end{bmatrix} \begin{bmatrix} y \\ z \end{bmatrix}$$

we get

$$\begin{bmatrix} y_{nk+k} \\ z_{nk+k} \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} \\ \varepsilon h & 1 + kh\lambda_z \end{bmatrix} \begin{bmatrix} y_{nk} \\ z_{nk} \end{bmatrix}$$

where

$$c_{11} = (1 + h\lambda_y)^k + \mu\varepsilon h^2 \frac{(1 + h\lambda_y)^k - kh\lambda_y - 1}{(h\lambda_y)^2}$$

and

$$c_{12} = \frac{(1 + h\lambda_y)^k - 1}{h\lambda_y} + h\lambda_z \frac{(1 + h\lambda_y)^k - kh\lambda_y - 1}{(h\lambda_y)^2}]$$

If either $1 + h\lambda_y$ or $1 + kh\lambda_z$ are close to one in magnitude, large values of $\varepsilon\mu$ could place the eigenvalues outside the unit circle and cause instability. It is difficult to estimate how serious this might be in cases of real interest. In this example, problems do not arise when $h\lambda_y$ and $kh\lambda_z$

are near zero, and for large values of these numbers, we should probably consider methods for stiff equations. For example, suppose we used the backward Euler method as

$$y_{nk+i+1} = y_{nk+i} + hf(y_{nk+i+1}, \hat{z}_{nk+i+1}, t_{nk+i+1})$$

$$i = 0,\ldots,k-2$$

$$\hat{z}_{nk+i+1} = z_{nk} + (i+1)hg(y_{nk}, z_{nk}, t_{nk})$$

and

$$y_{nk+k} = y_{nk+k-1} + hf(y_{nk+k}, z_{nk+k}, t_{nk+k})$$

$$z_{nk+k} = z_{nk} + khg(y_{nk+k}, z_{nk+k}, t_{nk+k})$$

we find that

$$\begin{bmatrix} y_{nk+k} \\ z_{nk+k} \end{bmatrix} = B \begin{bmatrix} y_{nk} \\ z_{nk} \end{bmatrix}$$

where

$$B = \begin{bmatrix} 1 - h\lambda_y & -\mu h \\ -\epsilon kh & 1 - hk\lambda_z \end{bmatrix}^{-1} \begin{bmatrix} d_{11} & d_{12} \\ 0 & 1 \end{bmatrix}$$

where

$$d_{11} = (1 - h\lambda_y)^{-k+1} + \epsilon\mu h^2 \frac{(1 - h\lambda_y)^{1-k} + (1-k)h\lambda_y - 1}{(h\lambda_y)^2}$$

and

$$d_{12} = \mu h\left[\frac{1 - (1 - h\lambda_y)^{1-k}}{h\lambda_y(1 - h\lambda_y)} + \lambda_z h \frac{(1 - h\lambda_y)^{1-k} + (1-k)h\lambda_y - 1}{(h\lambda_y)^2}\right.$$

For very negative values of $h\lambda_y$ and $hk\lambda_z$ and modest values of $h\mu$ and $kh\epsilon$ the eigenvalues of B are clearly less than one, so that we can hope that methods for stiff equations will be satisfactory. However, present knowl-

edge leads us to be wary of multirate methods unless the coupling from y to z through g is either extremely small, or is small compared to terms such as λ_y when methods for stiff equations are used. Note that one of the big savings in multirate stiff methods is that the inevitable system of non linear equations only has the dimensionality of y for (k-1) out of k steps.