U.S. Department of Commerce
National Oceanic and Atmospheric Administration
National Weather Service
National Centers for Environmental Prediction
5830 University Research Court
College Park, MD 20740

Office Note 482

doi:10.7289/V59Z92XW

ROBUSTNESS OF IMPLICIT RUNGE-KUTTA SCHEMES WITH RESPECT TO ERRORS IN THE SPECIFICATION OF THE SYSTEM'S COMPLEX FREQUENCIES.

R. James Purser*
IM Systems Group, Rockville, Maryland

May 19, 2016

This is an unreviewed manuscript, primarily intended for informal exchange of information among the NCEP staff members

* email: jim.purser@noaa.gov

Abstract

The note uses the idealized single-mode complex-frequency response analysis to investigate (subject to the limitations inherent in this idealization) the stability and robustness of some time integration schemes that might be considered candidates for incorporation within a model where both long time steps and a high order of formal accuracy are desired. Generally, we would expect that such models would be formulated in the semi-Lagrangian style, although this is not strictly necessary. The family of schemes we consider can broadly be categorized as "implicit Runge-Kutta" (of which the trapezoidal, and decentered generalizations are relatively trivial examples). The numerical robustness of these implicit schemes under simplified conditions depends upon the number of applications of the imperfect corrector iterations, and the amounts of decentering at each stage. The idealized experiments we describe shed light on this dependency and narrow down the choices we might wish to consider in formulating a formally accurate numerical model with a fuller degree of implicitness than is customarily adopted within the semi-Lagrangian paradigm.

1. Introduction

Implicit methods of time dicretization have the well-known benefit of allowing longer critical time steps, and therefore offer the prospect of more efficient numerical integrations, than are possible with purely explicit methods. Implementing a *fully* implicit time integration in the context of atmospheric modeling involves solving several challenging numerical problems. The fully implicit approach in a semi-Lagrangian numerical modeling context was suggested, and some of its implications outlined, by Purser (1983) at about the time when, largely owing to the influence and inspiration of Robert (1981, 1982), there began to be a concerted resurgence of interest in the semi-Lagrangian paradigm. As reported in the review of Staniforth and Côté (1991), Lagrangian ideas were a part of the development of numerical weather prediction models even from the early 1950s, but Robert's key step, which provoked the intense renewal of activity throughout the 1980s and since, was to combine the semi-Lagrangian advection with a semi-implicit treatment of the 'fast' modes which then allowed stable time steps to break the barrier hitherto set by the advective Courant-Friedrichs-Lewy (CFL) criterion (Simmons and Temperton, 1997). Recent years have witnessed a wider interest in the more fully implicit formulations at operational forecasting centers as nonhydrostatic dynamics have gained in importance. Consideration has been given to the more fully implicit approach at the ECMWF (see Cullen et al. 2000) and at Environment Canada (Yeh et al., 2002; Clancy and Pudykiewicz, 2013). Here at the National Centers for Environmental Prediction (NCEP) Kar has pioneered the incorporation of the fully implicit approach into a semi-Lagrangian nonhydrostatic model (Kar et al. 2004, Kar 2005) and has more recently examined an explicit scheme, combining an Adams-Bashforth predictor with a trapezoidal corrector, which also has implications for a fully implicit formulation (Kar 2012).

With a more fully implicit approach, an attempt must be made to incorporate the advective terms and, if feasible, even the physics terms within the set treated implicitly. With regard

to advection, it is important to distinguish the case in which this is dealt with in an Eulerian manner on the one hand, and in semi-Lagrangian manner on the other hand. Although it has been argued by Bartello and Thomas (1996) that the semi-Lagrangian paradigm becomes *less* attractive at the smaller scales, where a shallower energy spectrum is more typical, Côté et al. (1998) have disputed this controversial claim and, in fact, the semi-Lagrangian approach continues to be enthusiastically adopted by ever more operational centers, and seems justified by the evident computational cost savings that the method makes possible. Nevertheless, in cases where it is known a priori that the time step chosen will *not* be violating the advective CFL criterion, and where therefore the fully implicit approach is adopted as much for reasons of numerical accuracy as for reasons of numerical efficiency, then the formulation of a fully implicit scheme in a model possessing an Eulerian advection treatment is perfectly reasonable and has, in fact, been pioneered and demonstrated successfully by Reisner et al. (2005) for hurricane simulations and by Dijkstra et al. (2001) and Weijer et al. (2003) for an ocean model.

The option of using the undoubtedly simpler Eulerian advection framework is essentially eliminated when the dynamical grid adopted is one with widely varying resolution from place to place if one insists that the common time step must not be dictated by the unreasonably restrictive advective CFL limits imposed by the dynamics occurring in the regions where the grid is very fine. It is precisely this situation that pertains in the case of the recently proposed adaptive 'Generalized Fibonacci' grid of Purser (2008). This grid is designed to allow a broad degree of freedom to specify a variable spatial resolution (in space and time) able to adapt to the anticipated movement of highly active dynamical features (e.g., storms). In such situations, an especially high local spatial resolution might be expected to improve the overall simulation. Because the lines of the Fibonacci grid tend to be moderately curved, and because it is also necessary, in this grid, to smoothly blend together alternative ways of formulating the horizontal finite differences in order to preserve dynamical consistency across the domain and consequently it is desirable to adopt high-order spatial differencing formulae to keep the truncation errors of all the resolved scales in check. This is discussed in Swinbank and Purser (2006). When one already has a strong incentive to employ high-order spatial differencing in such a grid, an obvious question to ask is whether it is possible to preserve in the discretization a correspondly accurate temporal numerical treatment in a practical and reliable way, and if so, what are the methods that allow this to be done robustly? It is largely questions of this type that have motivated the present study.

Of course, even when the fully implicit scheme is found to be robust with respect to errors in the solution of the resulting generalized elliptic-type equations that specify the appropriate mutually-consistent adjustments to all the prognostic variables at each time step, in order to guarantee stability over a long period of integration there still remains the need to control or prevent the growth of nonlinear dynamical instability. This is particularly problematic with either the Fibonacci grid (even in Eulerian form) or with most semi-Lagrangian models where it is certainly impractical (if not impossible) to formulate the advective finite-differences in ways that bring about automatic conservation of the quadratic measures of energy (and perhaps, enstrophy) as has been done in simpler situations through the ingenious schemes of Arakawa (1966), Arakawa and Lamb (1977), for example. In the absence of adequate control of the energy cascade between scales in the spatial discretization, nonlinear instability will generally occur regardless of the choice of temporal discretization and, indeed, would occur even *without*

discretizing the time variable. However, provided safeguards in the spatial discretization are in place to ensure that nonlinear instability can be kept in check, the preliminary linear stability analyses that we illustrate for some idealized fully implicit methods suggest that we should be permitted to enjoy not just relatively long common time steps in our adaptive grid model, but we should also be able to benefit from having a high formal order of accuracy apply to the integrations done at such time steps. As we discussed in Purser (2007), a high formal order of accuracy in the time integration scheme seemingly achieved in a purely linear oscillatory stability analysis can be a misleading guide to the formal accuracy actually attained when nonlinearities are taken into account. Properly formulated Runge-Kutta schemes do attain the intended high-order of accuracy, and we shall consequently give such schemes prominence here.

We restrict the scope of the present note to an examination only of the robustness of certain implicit schemes and we simplify the problem by only considering here the robustness with respect to mis-specification of the complex frequency of the single representative mode of the system. This is, of course, a very drastic simplification, given that the actual time truncation errors in a complex simulation can occur from many other causes. Nevertheless, we often find that the robustness (or not) suggested by such simple analyses carries over to the more complete multivariate models.

Section 2 reviews the more familiar two-time-level implicit schemes with three different choices of a 'predictor' step and from one to three iterations of an imperfect 'corrector' step. The main focus, and the chief novelty, of the present study is the preliminary examination of what can be regarded as high-order multi-stage generalizations of the two-time-level schemes, which then take the form of 'Implicit Runge-Kutta' (IRK) schemes. These have been extensively described (but in the context of ordinary differential equation) by Ehle (1968), Butcher (1987), Lambert (1992) and by Hairer and Wanner (1996). Here we exemplify their adoption only in the shape of their two-stage versions, which we analyze in Section 3. The other numerical challenges mentioned above also need to be dealt with, but lie outside the scope of this note, except for a brief discussion of them in the concluding Section 4.

2. Adaptations of the trapezoidal scheme

The simplest family of fully implicit methods are the two-time-level schemes, of which the second-order 'trapezoidal' method is the most celebrated. In this scheme the earlier and later levels' forcing terms are weighted equally:

$$\frac{1}{\delta t}(P^{n+1} - P^n) = \frac{1}{2}(F^{n+1} + F^n), \tag{2.1}$$

where $F^n \equiv \mathcal{F}(P^n)$ denotes the forcing associated with state P^n for any time level n, and δt is the time step. In terms of the ideal linear system equation,

$$\frac{dP}{dt} = \nu P, \tag{2.2}$$

the trapezoidal scheme, (2.1), is known to possess a stability region comprising the left half plane of complex ν, just as the actual dynamical system does. When the stability region of the

numerical scheme includes the whole of the complex left half-plane, the scheme is said to be 'A-stable', a terminology coined by Dahlquist (1963) who also showed that, for multistep schemes that possess this property, the highest possible order is two and that, among these schemes, the trapezoidal scheme is the one with the smallest coefficient of error (see also Gear 1971). But in the context of a multivariate nonlinear dynamical system such as a model of the atmosphere it is never easy to isolate the individual linearized modes and, even after iterating, the practical forcing terms, F^{n+1}, at the new time level are never going to be exactly equal to what we would obtain by evaluating $\mathcal{F}(P^{n+1})$ at the final iteration of P^{n+1}. We therefore need to ensure that, in spite of this inevitable mismatch (which plagues all practical implementations of implicit schemes) we still have a numerical finite difference scheme that does not become unstable when the corresponding differential scheme is stable.

In the practical, and inevitably approximate, implementation of an intended ideal implicit scheme (such as the trapezoidal) we would normally begin the new time step with some crude explicit estimate for the new state, P^{n+1}. For example, if the final estimate of the state at time level n is denoted simply P^n, and the final estimate of the forcing at level n is denoted F^n (but now without any requirement that $F^n = \mathcal{F}(P^n)$ exactly), then we shall denote this preliminary guess for the next level's state, P_0^{n+1}. For example, it might be obtained by the zeroth-order 'null', or 'persistence' predictor:

$$P_0^{n+1} = P^n, \quad \text{('predictor 0')}, \tag{2.3}$$

or by a first-order 'Euler forward' predictor:

$$P_0^{n+1} = P^n + F^n \delta t, \quad \text{('predictor 1')}, \tag{2.4}$$

or, if an earlier time level of P is also stored, by a second-order 'leapfrog' predictor:

$$P_0^{n+1} = P^{n-1} + 2F^n \delta t, \quad \text{('predictor 2')}. \tag{2.5}$$

Given an estimate P_a^{n+1} where a is an iteration index, we need to apply a 'corrector' iteration to establish the improved estimate, P_{a+1}^{n+1}, and the generic approach we take to accomplish this is first to compute the actual exact forcing through a so-called 'function evaluation': $F_a^{n+1} = \mathcal{F}(P_a^{n+1})$, and thus determine the degree of mismatch, or 'residual', which, for the trapezoidal scheme, becomes:

$$R_a = P_a^{n+1} - P^n - \frac{\delta t}{2}(F^n + F_a^{n+1}). \tag{2.6}$$

We assume an approximate model for the Jacobian of this R with respect to P^{n+1}:

$$\delta R \approx J \, \delta P^{n+1} \tag{2.7}$$

where

$$J = I - \frac{\delta t}{2}\nu_0 \approx \frac{\partial R}{\partial P^{n+1}}. \tag{2.8}$$

The complex frequency, ν_0, here represents an approximation to the actual system frequency of the perturbed mode about the present state. The Newton-Raphson 'corrector' step, which involves updating both P^{n+1} and F^{n+1}, is

$$P_{a+1}^{n+1} = P_a^{n+1} - J^{-1} R_a, \tag{2.9}$$

$$F_{a+1}^{n+1} = F_a^{n+1} - \nu_0 J^{-1} R_a. \tag{2.10}$$

Alternatively, and sometimes more conveniently, we can write

$$F_{a+1}^{n+1} = 2(P_{a+1}^{n+1} - P^n)/\delta t - F^n. \tag{2.11}$$

In order to alleviate the well-known fragility of this scheme when the actual frequency, ν, differs for the assumed frequency, ν_0, we have recourse to the modification of the relative weighting of the two F contributions by what is referred to as 'decentering'. The idea is to make a small modification to the pure trapezoidal scheme to give a slightly larger proportion of the weight to the later forcing; that is, to try to satisfy:

$$P^{n+1} - P^n = \frac{\delta t}{2}((1-\gamma)\mathcal{F}(P^n) + (1+\gamma)\mathcal{F}(P^{n+1})). \tag{2.12}$$

Here the decentering parameter γ vanishes in the pure trapezoidal scheme and $\gamma = 1$ implies a fully forward-weighted 'Euler-backward' implicit scheme. In general we shall allow the decentering to change with iteration index, which means that we must also allow the Jacobian to change:

$$J_a = I - \frac{\delta t}{2}(1+\gamma_a) \tag{2.13}$$

for each $a \geq 1$ in the corrector step,

$$P_a^{n+1} = P_{a-1}^{n+1} - J_a^{-1} R_{a-1}. \tag{2.14}$$

We shall also allow a distinct implicit decentering in the definition of the final evaluation of the forcing, which, since it is only ever used to apply to the *following* time interval, we refer to unambiguously as γ_0 and apply in defining (unsubscripted) F^n:

$$F^n = \left[2(P^n - P^{n-1})/\delta t - (1-\gamma_0)F^{n-1}\right]/(1+\gamma_0). \tag{2.15}$$

With this convention, the M-times corrected scheme in general needs the specification of the $M+1$ decentering parameters, $\gamma_0, \ldots, \gamma_M$.

The stability of schemes is examined under the assumption that the true modal (complex) frequency, ν, is an unknown quantity, but is estimated to be ν_0 (which is used in the definitions of the Jacobians, J_a). In the complex ν-plane we plot the loci of points at which that numerical modal amplification has unit magnitude, and is therefore on the verge of instability. The simplest example is the centered scheme ($\gamma_0 = \gamma_1 = 0$) for a null predictor and single corrector, which is displayed in Fig. 1. In this, and other stability plots, panel (a) shows the limiting-stability curves for assumed frequencies non-dimensionalized with a $\delta t = 1$, at $\nu_0 = 0,\ 0.2i,\ 0.4i,\ 0.6i,\ 0.8i,\ 1.0i$. (These values are marked by asterisks along the positive imaginary axis.) It is also revealing to see how the stability curve evolves at frequencies $|\nu_0| \gg 1$ so we show this curve in panel (b) at the relatively large value, $\nu_0 = 5i$. We see that any underestimation of the magnitude of a pure-imaginary frequency leads to numerical instability, as the stable region for small frequencies can be identified as the region contiguous with small real negative ν, and the evolution of this region as ν_0 gradually increases. Matters are somewhat improved with a second corrector iteration (Fig. 2) since, with the second corrector, at least the bounding stability curve makes tangential contact with the imaginary axis where it cuts

through its reference frequency, ν_0. A third corrector iteration in the centered approximation to the trapezoidal scheme (Fig. 3) allows stability when imaginary ν is underestimated by ν_0, but leads to a very weak instability when ν is overestimated.

None of the centered schemes we have shown are robust – an arbitrarily small error in the estimated frequency ν_0 in cases where ν is imaginary can, if the error is in the adverse direction, put the numerical scheme into an unstable configuration. But at least the twice- and thrice-corrected schemes offer hope that a small degree of decentering might recover the desired feature of robustness, making the method lenient with respect to small errors in the frequency estimation.

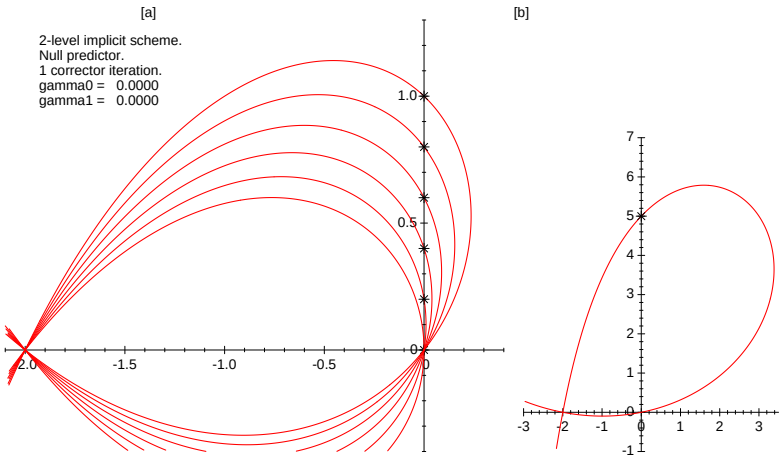

Figure 1. Stability plots for null-predictor, single corrector, two-level trapezoidal implicit scheme (no decentering). The asterisks along the imaginary axis of panel (a) locate the presumed modal complex frequency, ν_0, while the corresponding red curves through each of these locations show the loci of the actual complex frequencies, ν, at which the numerical behavior is neutrally stable, generally oscillating without the amplitude either growing or decaying. Panel (b) in this and the following figures shows the same kind of stability curve, but for a relatively large nominal frequency, $\nu_0 = 5i$, providing a glimpse of the asymptotic behavior as $\nu_0 \to \infty$.

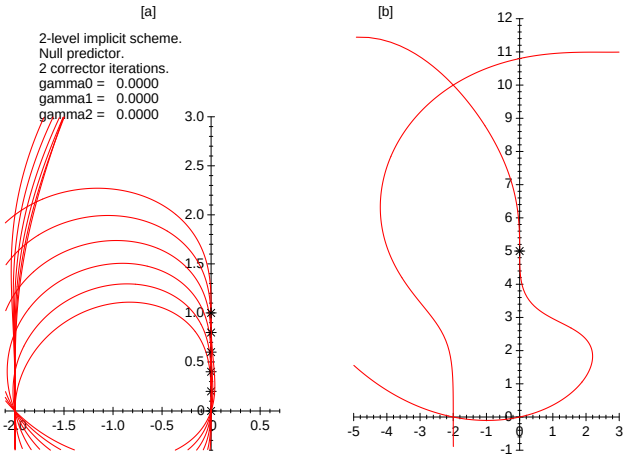

Figure 2. Like Fig. 1 but showing the stability plots for null-predictor, twice corrected, two-level trapezoidal implicit scheme (no decentering).

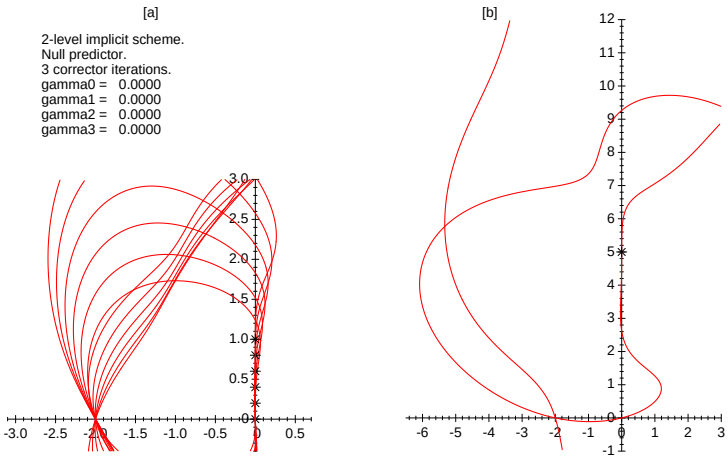

Figure 3. Stability plots for null-predictor, thrice corrected, two-level trapezoidal implicit scheme (no decentering).

Figure 4 shows the robust modification to the scheme of Fig. 2 obtained when all the gamma coefficients are set to 0.1. A generous stretch of the imaginary axis straddling each corresponding ν_0 confirms the robustness of this scheme, although a decentering of this magnitude will degrade the now first-order accuracy of the scheme significantly. A similar robustifying effect is observed when the thrice-corrected scheme is decentered with all gamma parameters set to 0.1, as shown in Fig. 5 .

We might ask whether it is necessary that all of the decentering parameters be made positive in order to achieve robustness. If only $\gamma_0 > 0$, as shown in Fig. 6, the scheme is actually destabilized (this is true for both the twice- and thrice-corrected schemes). It is actually enough to set only the final gamma to a small positive value. The desired robustness is seen for the twice-corrected scheme in Fig. 7. However, it is also important to note that the decentering of this last corrector step will also have the dominant effect upon the overall accuracy, which is therefore only first-order again.

The next question we investigate is whether a higher-order of predictor allows robustness to be achieve with less damage to the accuracy. Therefore, instead of the null 'predictor 0', we apply the Euler forward 'predictor 1', to obtain the once-corrected scheme shown in Fig. 8, the twice-corrected scheme shown in Fig. 9 and the thrice-corrected scheme shown in Fig. 10, where all the decentering parameters in these three schemes are set to zero. The once-corrected scheme remains inadequate without quite drastic measures taken to make it significantly robust, but the twice corrected scheme, even without any decentering, appears to have excellent stability characteristics, and is fully second-order accurate. The thrice corrected scheme with the Euler predictor is *less* robust than the scheme with two corrector stages.

We can further robustify the scheme shown in Fig. 9, if desired, with the gamma coefficients set to very small positive values. For example, Fig. 11 shows the stability plot obtained when $\gamma_0 = \gamma_1 = \gamma_2 = 0.01$.

There does not appear to be any advantage, as far as stability is concerned, in choosing a more accurate predictor. The leapfrog second-order 'predictor 2' is used in the once-, twice-

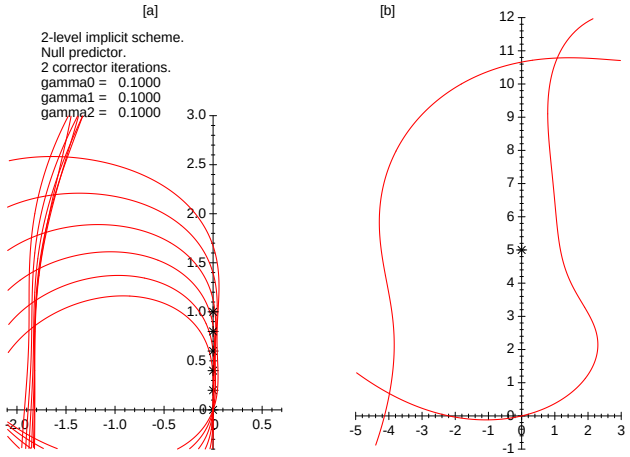

Figure 4. Stability plot for null-predictor, twice corrected, two-level implicit scheme with equal decentering at each stage.

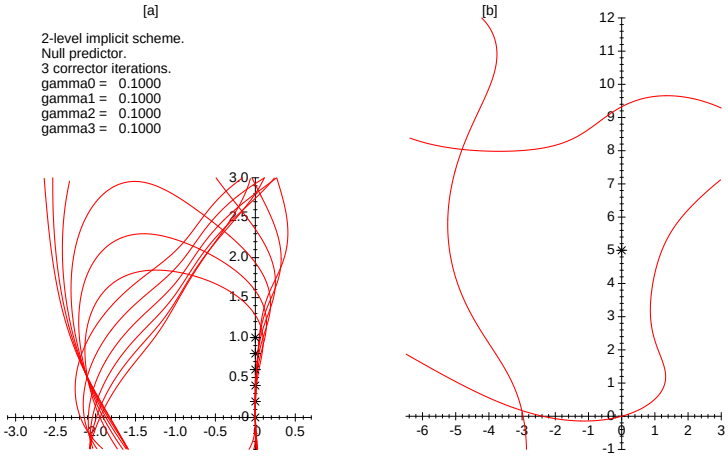

Figure 5. Stability plot for null-predictor, thrice corrected, two-level implicit scheme with equal decentering at each stage.

and thrice-corrected schemes shown respectively in Figs. 12, 13 and 14, in all of which the gamma coefficients are zero. As before, the schemes with one or three corrector stages are not robust as imaginary ν_0 becomes large, while the scheme of Fig. 13 with two corrector stages continues to show robustness with respect to small errors in the magnitude of ν_0 compared with ν. However, this scheme is clearly not as robust as the corresponding scheme of Fig. 9 which used the Euler forward predictor.

The two-level schemes we have examined are second-order accurate provided the last gamma coefficient vanishes. However, as we have shown, it is this same last gamma coefficient that predominantly controls each scheme's robustness and, by invoking this control, the positive coefficient causes the formal accuracy to decrease to only first-order. Of course, the effect is still small provided that the size of this controlling gamma coefficient remains small enough. We have found that the most robust scheme out of all those tested is the one whose predictor is the

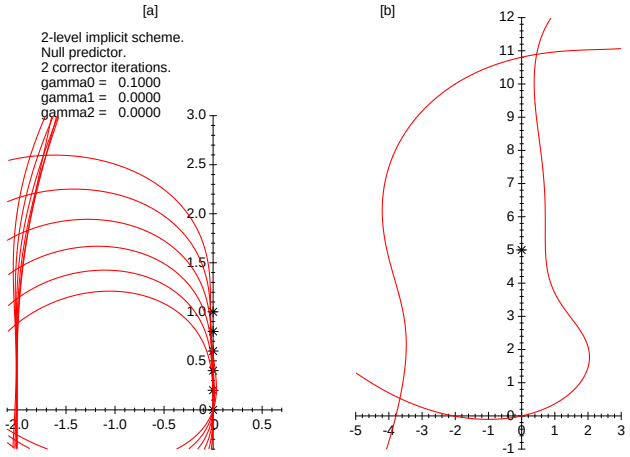

Figure 6. Stability plot for null-predictor, twice corrected, two-level implicit scheme with decentering at only stage zero, showing a destabilizing effect in the case of the small frequencies of panel (a).

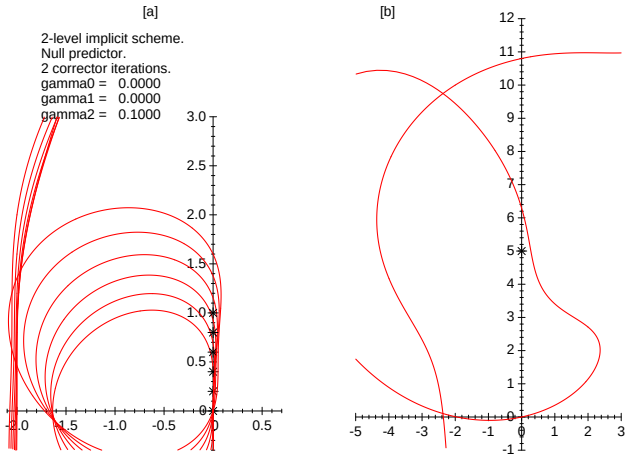

Figure 7. Stability plot for null-predictor, twice corrected, two-level implicit scheme with decentering at only the last stage, which achieves robustness, but at the expense of accuracy.

Eulerian forward step, and that has two corrector steps. This scheme (and the corresponding one using the leapfrog predictor) hardly needs the invocation of the gamma coefficients to achieve robustness, and is therefore the one to be recommended among these two-level schemes.

In cases, such as the adaptive-resolution Fibonacci grid, where it is desirable that the fully implicit methods possess a high formal order of accuracy in both space and time (to reduce spurious reflection or refraction effects at zones of changing resolution) then we need to look beyond the simple implicit two-level schemes we have considered. This brings us to their high-order generalizations, the implicit Runge-Kutta schemes, which we examine in the next section.

3. Implicit Runge-Kutta schemes based on Gaussian quadrature

The implicit Runge-Kutta (IRK) schemes were pioneered mainly by Butcher (1964, 1987) and are based on the various Gaussian quadrature formulae for the uniformly-weighted time-

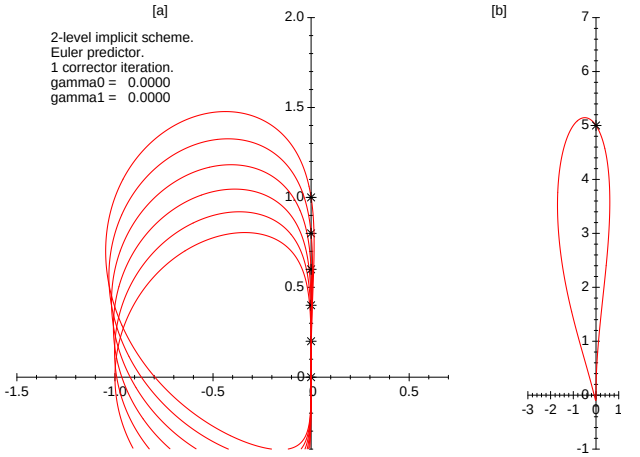

Figure 8. Stability plot for Euler 'predictor 1', once corrected, two-level implicit scheme without decentering.

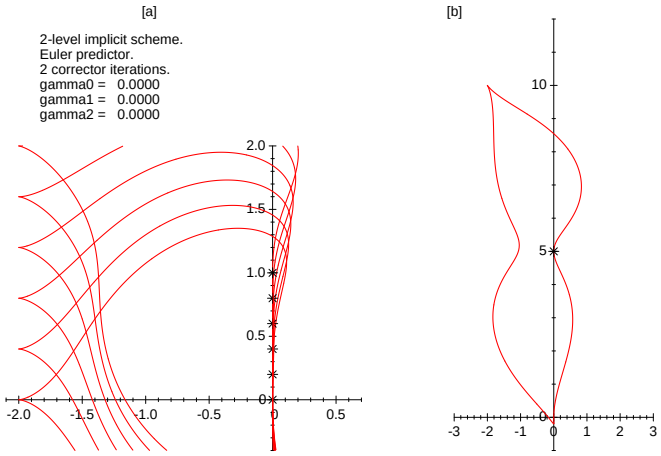

Figure 9. Like Fig. 8 except twice corrected.

step interval. The trapezoidal scheme can be thought of as one of these – the Gauss-Lobatto quadrature with just two nodes. In the theory of Gaussian quadrature over an interval, the Lobatto scheme differs from the better known Legendre quadrature in having two of its nodes at the end points themselves – but sacrifices formal accuracy since the Lobatto scheme of m points is guaranteed true up to polynomials only of degree two less than those for which the corresponding Legendre scheme of m points is guaranteed true. Another quadrature we encounter, the Gauss-Radau scheme, has just one of its nodes an end point, and attains an intermediate accuracy between the Legendre and Lobatto variants. The Gauss-Lobatto scheme of m nodes leads to a scheme of a formal accuracy of order $2m-2$, but unfortunately the higher order time integration schemes based upon the Gauss-Lobatto quadrature in their most straight-forward implementations are not A-stable. (Butcher's book describes the A-stable 'Lobatto IIIC' schemes, but they necessitate an additional copy of the state variable and its associated function evaluation at the start of the time interval.) The schemes based on Gauss-Legendre quadrature with m nodes are A-stable and have a formal order of accuracy of $2m$.

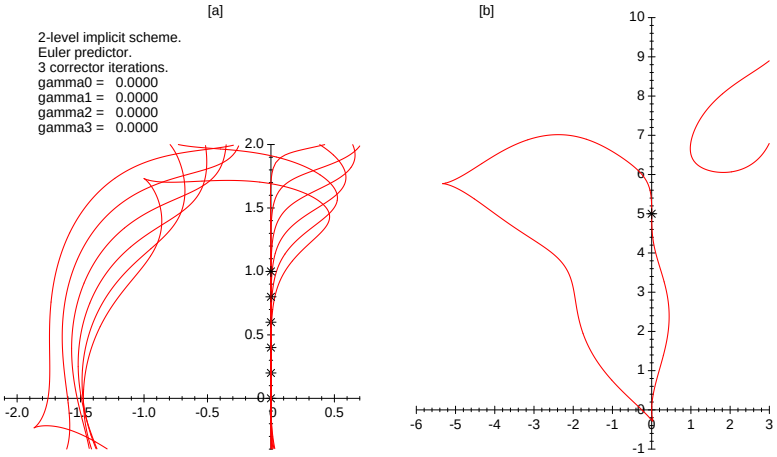

Figure 10. Like Fig. 8 except twice corrected.

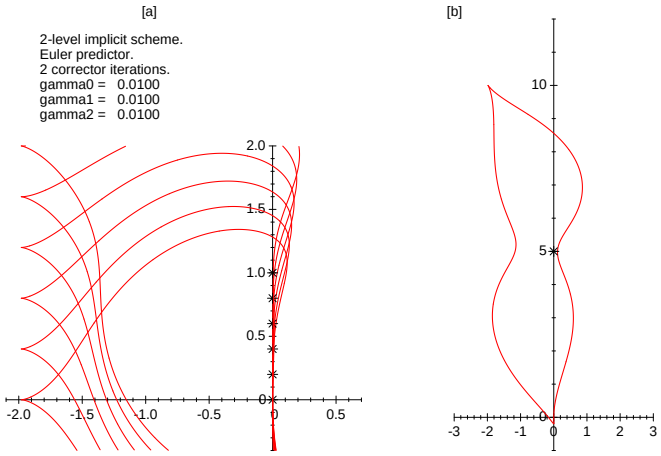

Figure 11. Like Fig. 9 except very small positive decentering coefficients.

Their nodes, where the function evaluations must occur, are all properly interior to the time step interval. We shall only consider the 2-node versions of the implicit Runge-Kutta schemes so, in the case of the Gauss-Legendre scheme, its formal accuracy is of order four.

The two-stage implicit Runge-Kutta scheme attempts to satisfy:

$$P_1 = P^n + (a_{1,1}F_1 + a_{1,2}F_2)\delta t, \tag{3.1a}$$

$$P_2 = P^n + (a_{2,1}F_1 + a_{2,2}F_2)\delta t, \tag{3.1b}$$

$$P^{n+1} = P^n + (b_1F_1 + b_2F_2)\delta t, \tag{3.1c}$$

where $F_1 = \mathcal{F}(P_1)$ and $F_2 = \mathcal{F}(P_2)$ and where P_1 and P_2 are assumed to occupy interior time levels, $t_1 = (n + c_1)\delta t$ and $t_2 = (n + c_2)\delta t$. As before, the practical implementation requires iterative stages to approach a solution of these equations. A second subscript on the two intermediate P values, and on the corresponding forcing terms, denotes this iteration index. As before, there are different strategies to initializing these iterations with a 'predictor'. The null

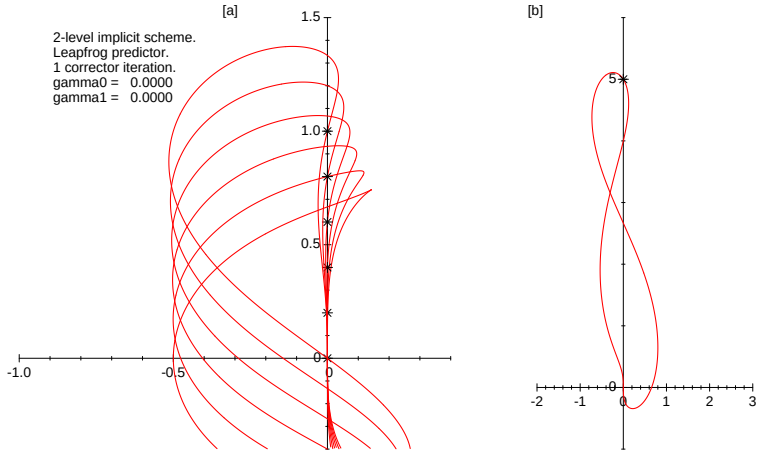

Figure 12. Stability plot for leapfrog 'predictor 2'), once corrected, two-level implicit scheme without decentering.

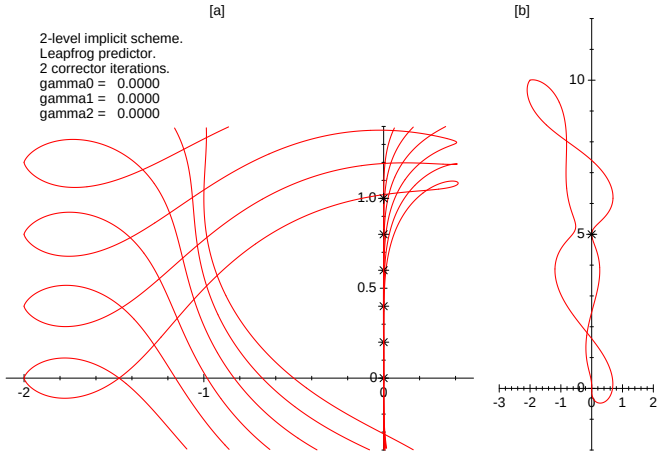

Figure 13. Like Fig. 12 but with two corrector stages.

predictor just sets $P_{1,0} = P^n$ and $P_{2,0} = P^n$. Since all the information determining the next time step is contained in the single state variable, P^n, this means that, for a scalar representation of this scheme, the 'transfer matrix' is trivially of order one. If we use just the additional information in an estimate of the forcing, F^n, at the start of the new time step interval, we will have a transfer matrix of order two. In this latter case it is natural to again use the Euler forward predictor to the two intermediate times $(n + c_1)\delta t$ and $(n + c_2)\delta t$. (Consistency requirements for these schemes are discussed in the appendix.) The Euler predictor steps are:

$$P_{1,0} = P^n + c_1 F^n \delta t, \tag{3.2a}$$

$$P_{2,0} = P^n + c_2 F_n \delta t. \tag{3.2b}$$

The residual at iteration a for the two-stage process is now a 2-vector, $\boldsymbol{R}_a$.

$$\begin{bmatrix} R_{1,a} \\ R_{2,a} \end{bmatrix} = \begin{bmatrix} P_{1,a} - P^n \\ P_{2,a} - P^n \end{bmatrix} - \begin{bmatrix} a_{1,1}, & a_{1,2} \\ a_{2,1}, & a_{2,2} \end{bmatrix} \begin{bmatrix} F_{1,a} \\ F_{2,a} \end{bmatrix} \delta t, \tag{3.3}$$

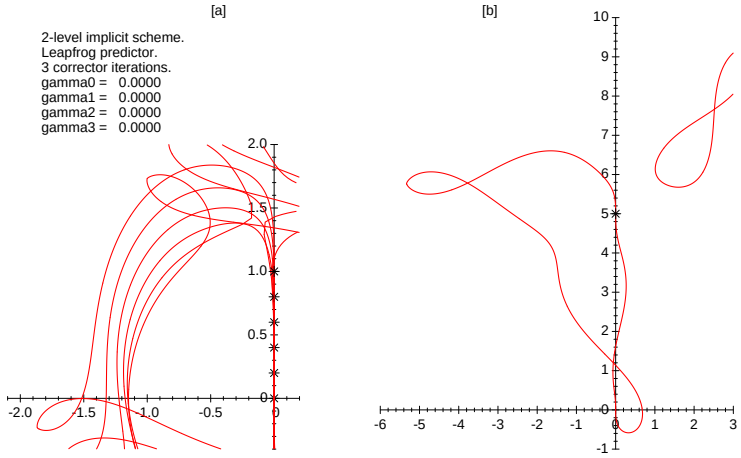

Figure 14. Like Fig. 12 but with three corrector stages.

The Jacobian $\boldsymbol{J}$ for the corrector step is now a 2×2 matrix:

$$\boldsymbol{J} = \boldsymbol{I} - \boldsymbol{a}\nu_0 \delta t, \tag{3.4}$$

where $\boldsymbol{a}$ is the matrix formed by the four a-weight coefficients. The two-vector $\boldsymbol{P} = [P_1, P_2]^T$ of intermediate states is updated,

$$\boldsymbol{P}_{a+1} = \boldsymbol{P}_a - \boldsymbol{J}^{-1}\boldsymbol{R}_a \tag{3.5}$$

The final updated state, P^{n+1}, is obtained by applying (3.1c) to the last available iteration of the intermediate forcings. In the case where the Euler predictor is used, the forcing, F^{n+1}, that we use for the next Euler predictor step is obtained by linearly extrapolating the last available intermediate forcings:

$$F^{n+1} = d_1 F_1 + d_2 F_2, \tag{3.6}$$

with

$$d_1 = \frac{1 - c_2}{c_1 - c_2}, \tag{3.7a}$$

$$d_2 = \frac{1 - c_1}{c_2 - c_1}. \tag{3.7b}$$

The conventional way of exhibiting the Runge-Kutta coefficients is by presenting them in the "Butcher tableau",

$$\begin{array}{c|cc} c_1 & a_{1,1} & a_{1,2} \\ c_2 & a_{2,1} & a_{2,2} \\ \hline & b_1 & b_2 \end{array} \tag{3.8}$$

In the case of the fourth-order two-stage Gauss-Legendre scheme, with the numerical values substituted, this tableau becomes:

$$\begin{array}{c|cc} \frac{3-\sqrt{3}}{6} & \frac{1}{4} & \frac{3-2\sqrt{3}}{12} \\ \frac{3+\sqrt{3}}{6} & \frac{3+2\sqrt{3}}{12} & \frac{1}{4} \\ \hline & \frac{1}{2} & \frac{1}{2} \end{array} \tag{3.9}$$

With null predictor and one corrector iteration, the stability graphs are shown, in the same format used for the simple two-level schemes, in Fig. 15 and, clearly the scheme is not robust for either small or large imaginary ν_0. Matters are made even worse when the Euler predictor is used, as we see in Fig. 16. A considerable improvement occurs with a second corrector iteration; in the case of the null predictor, the result is shown in Fig. 17, and in the case of the Euler predictor, in Fig. 18.

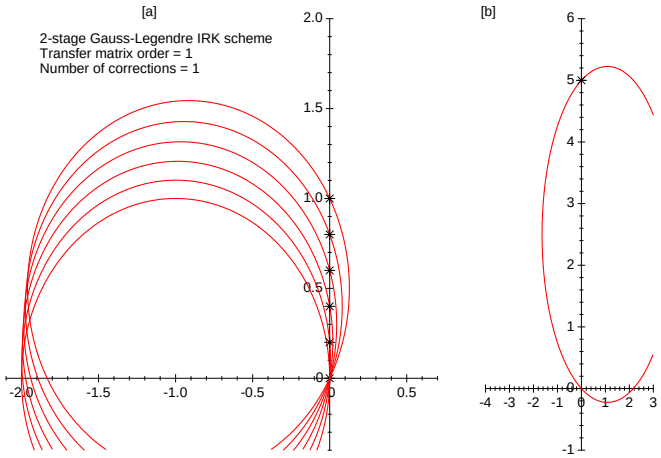

Figure 15. Stability curves for the Gauss-Legendre two-stage IRK scheme with null predictor and once corrected.

If we require greater robustness, we can opt for the corresponding modifications of these Gauss-Legendre quadratures that are associated with alternative, formally less accurate, Gaussian quadratures. The most drastic alternative is to use the Radau scheme that shifts the nodes forward so that the second occurs at the end of the time step, that is, with $c_2 = 1$, while preserving the formal accuracy of order three. The resulting scheme possesses the tableau,

$$\begin{array}{c|cc} \frac{1}{3} & \frac{5}{12} & -\frac{1}{12} \\ 1 & \frac{3}{4} & \frac{1}{4} \\ \hline & \frac{3}{4} & \frac{1}{4} \end{array} \tag{3.10}$$

The scheme with null predictor, single corrector (Fig. 19), is hardly improved since the results at small ν_0 still show a serious lack of robustness (at large ν_0 the results are much better).

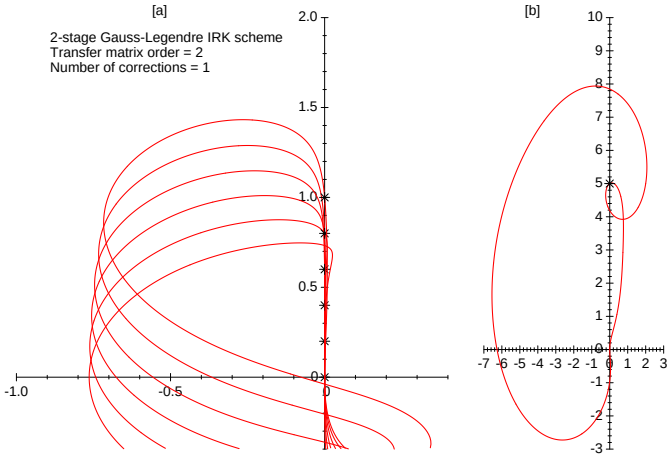

Figure 16. Stability curves for the Gauss-Legendre two-stage IRK scheme with Euler predictor and once corrected.

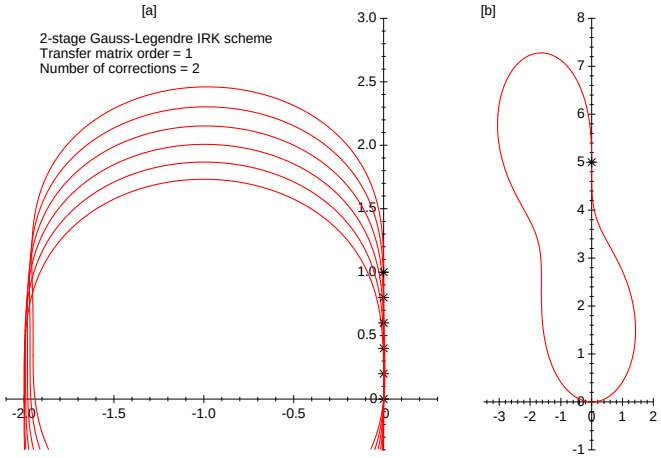

Figure 17. Stability curves for the Gauss-Legendre two-stage IRK scheme with null predictor and twice corrected.

Use of the Euler predictor (Fig. 20) substantially improves the single-corrector Radau scheme's robustness at large and small frequencies. If it is important to *not* repeat the costly corrector iteration, this scheme would probably be the recommended one that maintains the high order of accuracy in time. If we are prepared to perform the corrector iteration twice, we obtain the stability results of Fig. 21 with the null predictor, and those of Fig. 22 with the Euler predictor – both of these schemes are very robust.

The Radau tableau is just the first of an infinite sequence of schemes of order three with rational coefficients, whose principal error coefficients become progressively smaller and with the Gauss-Legendre scheme as their limit, as we outline in the appendix. Given that the twice-corrected Radau schemes are probably more robust than we would normally require, it is informative to see whether we can use the next member of the sequence to achieve a formally slightly more accurate scheme while maintaining an adequate degree of robustness. The next

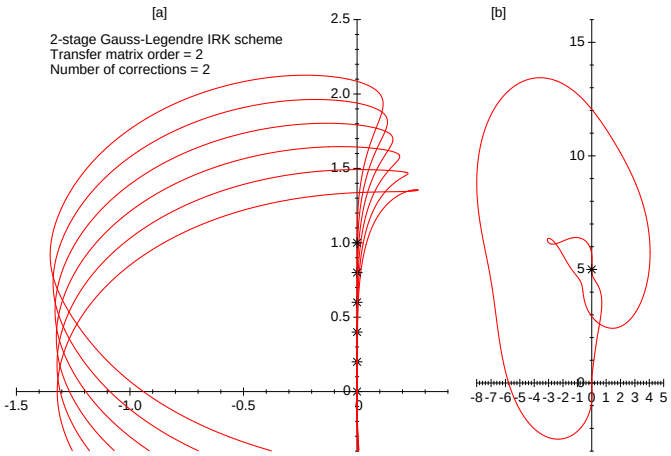

Figure 18. Stability curves for the Gauss-Legendre two-stage IRK scheme with Euler predictor and twice corrected.

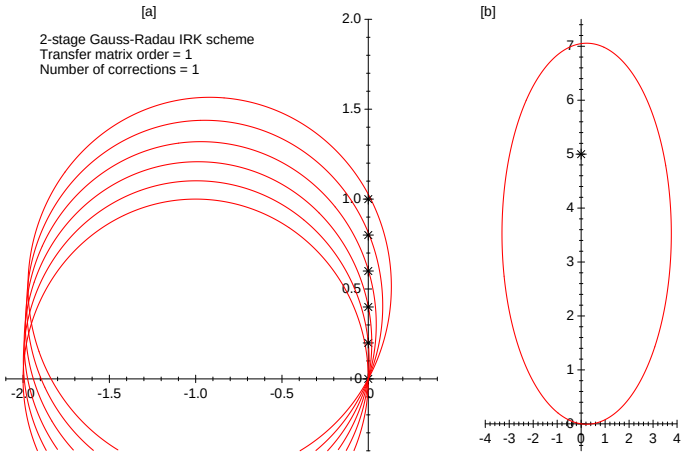

Figure 19. Stability curves for the Gauss-Radau two-stage IRK scheme with null predictor and once corrected.

member of the sequence of rational-coefficient schemes is what we shall refer to as the 'hybrid Legendre-Radau', or briefly, 'Radau' scheme, and has the tableau:

$$\begin{array}{c|cc} \frac{1}{4} & \frac{17}{56} & -\frac{3}{56} \\ \frac{5}{6} & \frac{25}{42} & \frac{5}{21} \\ \hline & \frac{4}{7} & \frac{3}{7} \end{array} \tag{3.11}$$

The hybrid scheme with only a single corrector step does not look sufficiently robust, even when the Euler predictor is used (Fig. 23). But with two corrector iterations, both with the null predictor (Fig. 24) or with the presumably more accurate but less robust Euler predictor (Fig. 25) we still appear to have schemes that are sufficiently robust for practical purposes. Of

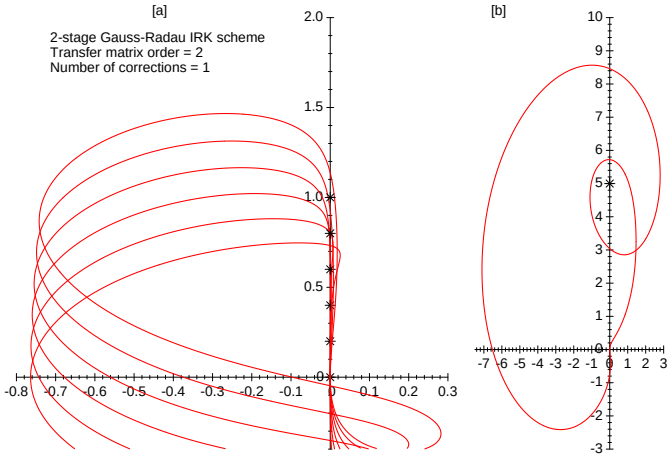

Figure 20. Stability curves for the Gauss-Radau two-stage IRK scheme with Euler predictor and once corrected.

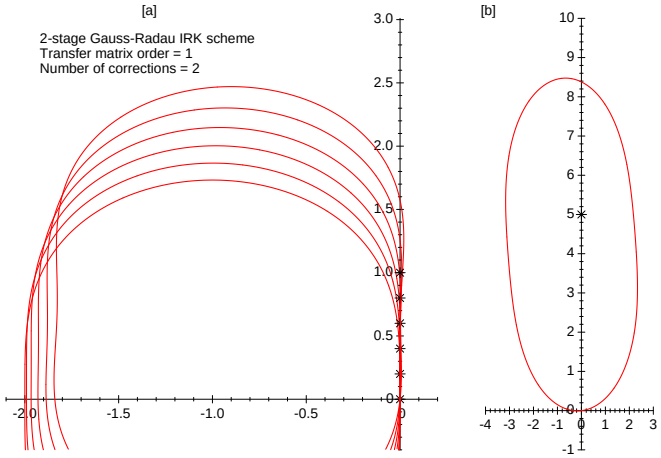

Figure 21. Stability curves for the Gauss-Radau two-stage IRK scheme with null predictor and twice corrected.

course, there is no compelling reason why we must restrict attention to schemes with rational weighting coefficients, so the best practical scheme at third order might be another of this continuum of implicit Runge-Kutta schemes that generalize the Gaussian quadrature.

To conclude, our idealized experiments suggest that:

• The Euler-forward predictor, followed by the two corrector iterations leads to the most stable practical implementation for the trapezoidal-based methods;

• If in future it is desired to achieve a higher order (third or fourth) of accuracy and A-stable schemes, those schemes based on implicit Runge-Kutta of the Gauss quadrature type are recommended;

• For those Gaussian-quadrature-based IRK schemes employing Euler proedictors, satisfactory levels of stability and robustness are best achieved using two corrector iterations;

• Greater robustness in the high order schemes is achieved by abandoning strict fourth-

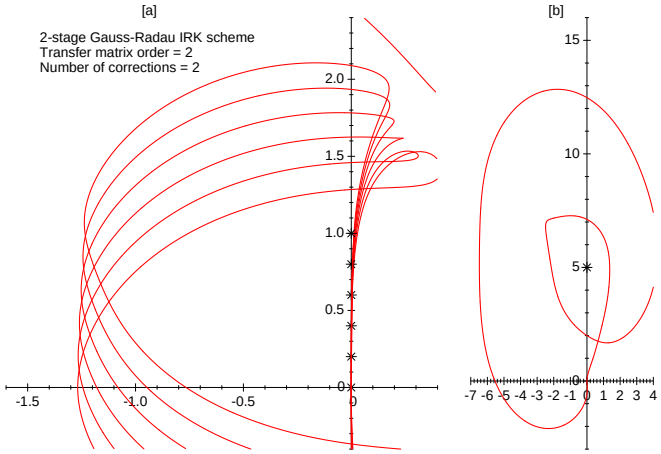

Figure 22. Stability curves for the Gauss-Radau two-stage IRK scheme with Euler predictor and twice corrected.

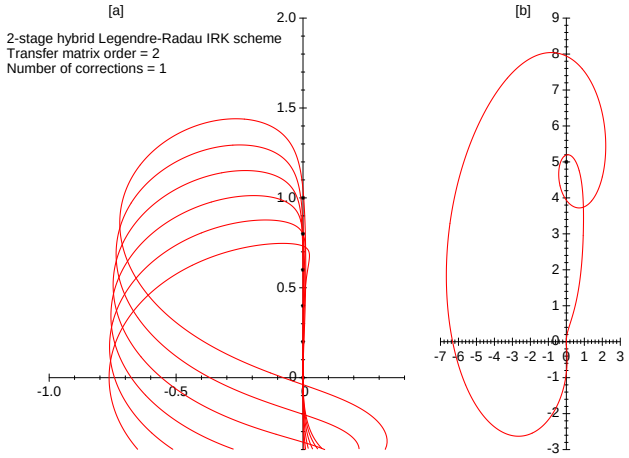

Figure 23. Stability curves for the hybrid two-stage IRK scheme with Euler predictor and once corrected.

order accuracy of the centered schemes and adopting the decentered schemes either of Radau, or hybrid Lengendre-Radau type.

4. DISCUSSION

With a reasonably smooth spatial grid it is quite straightforward to formulate a finite difference atmospheric simulation model with a formally high order or accuracy in space and, while it is legitimate to question the desirability of doing this, it is surely more consistent to take this approach if the accuracy of the time integration is also chosen to be of high order. Several well-known and easily-coded explicit time integration schemes of high order accuracy are available but, in order to take advantage of potentially long time steps for greater numerical efficiency, we face considerable difficulties in identifying practical and robust *implicit* time integration schemes whose formal accuracy is even as much as second order. Dahlquist

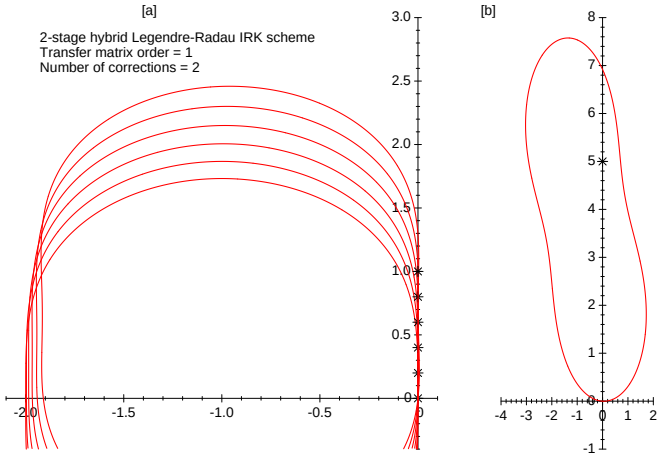

Figure 24. Stability curves for the hybrid two-stage IRK scheme with null predictor and twice corrected.

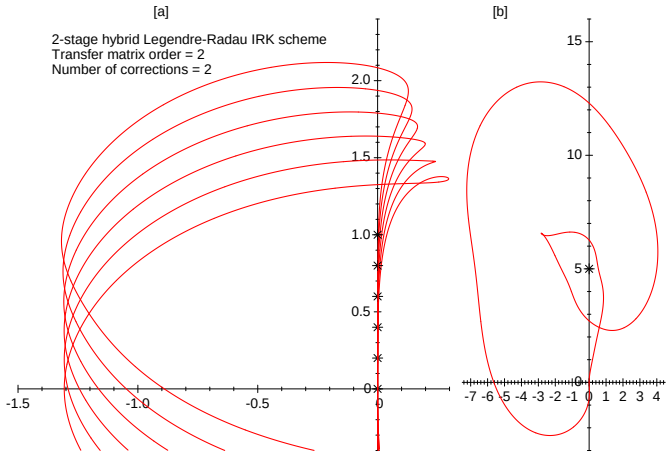

Figure 25. Stability curves for the hybrid two-stage IRK scheme with Euler predictor and twice corrected.

(1963) showed that the trapezoidal scheme is the best of the A-stable implicit multistep schemes but, as we have seen, it is not a trivial matter to choose a practical implementation strategy for this scheme that is sufficiently robust when there is some uncertainty about the actual modal frequencies. A small forward-weighted decentering is a great help in recovering robustness, but at the cost of a formal loss of second-order accuracy. Two iterations of the implicit corrector iteration are confirmed to be dramatically effective at recovering robustness in many cases, even when three iterations lead back to a situation of non-robustness (especially for the very high frequency components). The present experiments suggest that the Euler-forward predictor, followed by two corrector iterations, yields the best (most robust) practical implementation for the trapezoidal method out of those we have tested.

If it is desirable to secure an even higher order of accuracy – third or fourth-order, for example, then, if we want to have A-stable schemes, we need to consider the fully implicit Runge-Kutta schemes based upon the various classical Gauss quadrature methods. We have restricted this study to the two-stage methods of this kind, which already provide us with

schemes of third- or fourth-order accuracy. The analogue of the centered scheme is the formally fourth-order Gauss-Legendre-based IRK scheme. Once again, it is only by performing two corrector iterations that we obtain anything resembling robustness for this scheme and, in practice, errors of adverse sign in the estimated system frequency for these schemes would lead to a weak instability. A safer approach is to abandon strict fourth-order accuracy and decenter the Gauss-Legendre quadrature either to the fully decentered Gauss-Radau scheme, for which the twice-corrected implementations are shown to be more than adequately robust, or an intermediate 'hybrid Legendre-Radau' scheme which is only very slightly decentered (and therefore becomes, in practice, *almost* fourth-order accurate again). This can retain an adequate degree of robustness (though much less than is possessed by the Radau scheme).

One attractive feature of the Gaussian-quadrature-based IRK schemes is that they can be generalized to A-stable schemes of arbitrary formal orders of accuracy. A particularly convenient family of such schemes, possessing rational coefficients derived from a so-called "Pell" series, is provided and discussed in the appendix. The techniques we need to consider for solving the implicit equations (that is, inverting the system whose operator contains our Jacobian, $\boldsymbol{J}$) for the two-stage IRK schemes would in principle apply to implementations at even higher formal order. However, having said that, we must admit that this inversion is far from trivial for the case of any fully implicit atmospheric simulation model, even in the relatively straightforward case of the two-time-level schemes since, if we are expecting to take full advantage of the large time steps that the implicit approach allows, we must also incorporate the highly variable advective terms within that Jacobian operator, and assume that the framework used for these terms is a semi-Lagrangian one. The real challenge is therefore the inversion of the equations for the 'residual', $\boldsymbol{R}$, when, for a full atmospheric model, the components of this quantity are as numerous as (and in direct correspondence with) the prognostic variable components of the model, and when the equations defining increments to $\boldsymbol{R}$ are linearized not about some simple basic rest state, but about the best available approximation to the presently relevant dynamically evolving actual state. Since this linearization is almost precisely what is required in a four-dimensional variational (4D-Var) assimilation schemes, it is evidently at least not beyond the bounds of feasibility. Symbolically, we might express the vector $\boldsymbol{R}$ of the linearized components of the residuals in terms of the vector of linearized 'errors' of the new time-interval's state variables, $\boldsymbol{P}$, through a generic linear spatially-localized operator, $\mathcal{D}$:

$$\boldsymbol{R} \equiv \mathcal{D}\boldsymbol{P}. \tag{4.1}$$

Then, in order to estimate the unkown increment, $-\boldsymbol{P}$, needed to nullify the residual, $\boldsymbol{R}$, we might consider adapting another tool of 4DVar – namely the operator that is formally the *adjoint* (with respect to a suitable inner-product norm – typically an approximate energy norm) of $\mathcal{D}$, which we denote $\mathcal{D}^\dagger$, and constructing the generalized 'elliptic' equation:

$$\mathcal{D}^\dagger\mathcal{D}\boldsymbol{P} = \mathcal{D}^\dagger\boldsymbol{R}. \tag{4.2}$$

The self-adjoint and normally positive-definite operator, $\mathcal{D}^\dagger\mathcal{D}$ leads to a much easier inversion problem than that using just $\mathcal{D}$ by itself, even though, in this very general form, the full set of dynamical variables are involved in a coupled way. Nevertheless, multigrid solution methods are well-adapted to solving problems of this kind where the operators involved are essentially

spatially localized. This should remain true despite the fact that the operator $\mathcal{D}$ will now contain not just in-situ derivatives, but derivatives of the interpolation operators comprising the semi-Lagrangian advection machinery that we are linearizing when, for example, we estimate the residuals along our initial Euler-forward-projected trajectory.

Another essential challenge that must be met concerns the need to massively parallelize the computations involved in solving the implicit scheme's elliptic problem. The difficulty stems from the long-range influence implied by solving any forced elliptic problem. Multigrid procedures for such problems are moderately well adapted to parallel computing environments. It is also possible that compact-support kernels that adequately approximate the Green's function in the inversion of the elliptic operator would enable solutions to this inversion to be obtained, to a sufficient approximation, without the great amount of costly data motion that the unapproximated elliptic problem would seem to require. This is an option deserving further study.

Acknowledgments

I am grateful for several stimulating discussions with Dr. Sajal Kar and for the helpful reviews provided by Drs. Eugene Mirvis and Miodrag Rančić.

Appendix A

Some identities relating to 2-stage Implicit Runge-Kutta (IRK) schemes

It is well-known that, in order for an n-point numerical quadrature over the unit interval, $t \in [0,1]$ to possess at least $(2p-1)$th-order accuracy, the p nodes must reside at the zeroes of a polynomial of the form,

$$f(t) = L^p(t) - \lambda L^{p-1}(t) \tag{A.1}$$

where $L^p(t)$ and $L^{p-1}(t)$ are the Legendre polynomials of degree p and $p-1$ respectively that are scaled to be orthogonal in this interval and that are normalized to $L^p(1) = L^{p-1}(1) = 1$. When $p = 2$ we find,

$$L^1(t) = 1 - 2t, \tag{A.2a}$$
$$L^2(t) = 1 - 6t + 6t^2, \tag{A.2b}$$

so that the zeroes occur at:

$$c_1 = \frac{1}{6}\left[3 + \lambda - (\lambda^2 + 3)^{1/2}\right], \tag{A.3a}$$
$$c_2 = \frac{1}{6}\left[3 + \lambda + (\lambda^2 + 3)^{1/2}\right] \tag{A.3b}$$

In order that the numerical scheme be exact for linear and quadratic evolution of solution P throughout the time step it is required that:

$$\begin{bmatrix} a_{1,1}, & a_{1,2} \\ a_{2,1}, & a_{2,2} \\ b_1, & b_2 \end{bmatrix} \begin{bmatrix} 1, & 2c_1 \\ 1, & 2c_2 \end{bmatrix} = \begin{bmatrix} c_1, & c_1^2 \\ c_2, & c_2^2 \\ 1, & 1 \end{bmatrix}, \tag{A.4}$$

or,

$$\begin{bmatrix} a_{1,1}, & a_{1,2} \\ a_{2,1}, & a_{2,2} \\ b_1, & b_2 \end{bmatrix} = \begin{bmatrix} c_1 + \frac{c_1^2}{2(c_2-c_1)}, & -\frac{c_1^2}{2(c_2-c_1)} \\ \frac{c_2^2}{2(c_2-c_1)}, & c_2 - \frac{c_2^2}{2(c_2-c_1)} \\ \frac{1}{2} - \frac{1-(c_1+c_2)}{2(c_2-c_1)}, & \frac{1}{2} + \frac{1-(c_1+c_2)}{2(c_2-c_1)} \end{bmatrix}. \tag{A.5}$$

TABLE 1. Coefficients of the 2-stage Implicit Runge-Kutta schemes associated with the solutions to a Pell equation.

n	C_n	S_n	c_1	c_2	$a_{1,1}$	$a_{1,2}$	$a_{2,1}$	$a_{2,2}$	b_1	b_2
1	2	1	$\frac{1}{3}$	1	$\frac{5}{12}$	$\frac{-1}{12}$	$\frac{3}{4}$	$\frac{1}{4}$	$\frac{3}{4}$	$\frac{1}{4}$
2	7	4	$\frac{1}{4}$	$\frac{5}{6}$	$\frac{17}{56}$	$\frac{-3}{56}$	$\frac{25}{42}$	$\frac{5}{21}$	$\frac{4}{7}$	$\frac{3}{7}$
3	26	15	$\frac{2}{9}$	$\frac{4}{5}$	$\frac{31}{117}$	$\frac{-5}{117}$	$\frac{36}{65}$	$\frac{16}{65}$	$\frac{27}{52}$	$\frac{25}{52}$
4	97	56	$\frac{3}{14}$	$\frac{19}{24}$	$\frac{345}{1358}$	$\frac{-27}{679}$	$\frac{2527}{4656}$	$\frac{1159}{4656}$	$\frac{49}{97}$	$\frac{48}{97}$

There is no compelling reason to insist on solutions with rational weights other than the numerical convenience of tabulating them. Nevertheless, we observe that such solutions exist with,

$$\lambda = 1/S, \tag{A.6}$$

for positive integers, S, that satisfy the Diophantine (integer solution) 'Pell' equation,

$$C^2 - 3S^2 = 1. \tag{A.7}$$

All integer pairs, (C, S), satisfying this equation can be located in an infinite sequence generated (in either direction) from any two consecutive solutions, such as,

$$(C_0, S_0) = (1, 0), \tag{A.8a}$$
$$(C_1, S_1) = (2, 1), \tag{A.8b}$$

by the simple three-term recurrence:

$$(C_{n-1}, S_{n-1}) - 4(C_n, S_n) + (C_{n+1}, S_{n+1}) = (0, 0). \tag{A.9}$$

We recognize the Radau scheme in the choice $\lambda = 1/S_1 = 1$ and our 'hybrid' scheme of (3.11) as the second member of the sequence, $\lambda = 1/S_2 = 1/4$. For the generic mth member of this family, with $\lambda = 1/S_m$, we find that resulting rational c_1 and c_2 can also be evaluated as:

$$c_1 = \sqrt{\frac{1}{6}\left(\frac{C_{m-1}+1}{C_m+1}\right)}, \tag{A.10}$$

$$c_2 = \sqrt{\frac{1}{6}\left(\frac{C_{m+1}-1}{C_m-1}\right)}, \tag{A.11}$$

while the weights, b_1 and b_2, become:

$$b_1 = \frac{1}{2} - \frac{1}{2C_m}, \tag{A.12a}$$

$$b_2 = \frac{1}{2} + \frac{1}{2C_m}. \tag{A.12b}$$

The coefficients for the first few of these schemes are provided in Table 1. As $m \to \infty$ the tableau for the IRK scheme of this family converges to that of the Gauss-Legendre scheme.

References

Arakawa, A.	1966	Computational design for long-term numerical integrations of the equations of atmospheric motion. *J. Comp. Phys.,* **1,** 119–143.
Arakawa, A., and V. Lamb	1977	Computational design of the basic dynamical processes of the UCLA general circulation model. *Methods in Computational Physics* Vol. 17, J. Chang, Ed., Academic Press, 173–265.
Bartello, P., and S. J. Thomas	1996	The cost-effectiveness of semi-Lagrangian advection. *Mon. Wea. Rev.,* **124,** 2883–2897.
Butcher, J. C.	1964	Implicit Runge-Kutta Processes. *Math. Comp.,* **18,** 50–64.
Butcher, J. C.	1987	*The Numerical Analysis of Ordinary Differential Equations.* John Wiley, New York. 512 pp.
Clancy, C., and J. A. Pudykiewicz	2013	On the use of exponential time integration methods in atmospheric models. *Tellus A,* **65,** 1–16.
Côté, J., S. Gravel, M. Roch, and A. Staniforth	1998	The operational CMC-MRB Global Environmental Multiscale (GEM) model. Part I: Design considerations and formulation. *Mon. Wea. Rev.,* **126,** 1373–1395.
Cullen, M. J. P., D. Salmond, and P. Smolarkiewicz	2000	Key numerical issues for the future development of the ECMWF numerical model. *Proc. ECMWF Workshop on Development in Numerical Methods for Very High Resolution Global Models,* Reading, U.K., ECMWF, 183–206.
Dahlquist, G.	1963	A special stability problem for linear multistep methods. *BIT,* **3,** 27–43.
Dijkstra, H. A., H. Oksuzoglu, F. W. Wubs, and E. F. F. Botta	2001	A fully implicit model of the three-dimensional thermohaline ocean circulation. *J. Comp. Phys.,* **173,** 685–715.
Ehle, B. L.	1968	High order A-stable methods for the numerical solution of systems of differential equation. *BIT,* **8,** 276–278.
Gear, C. W.	1971	*Numerical Initial Value Problems in Ordinary Differential Equations* Prentice-Hall, Englewood Cliffs, New Jersey, 253 pp.
Hairer, E., and G. Wanner	1996	*Solving Ordinary Differential Equations II; Stiff and Differential-Algebraic Problems.* Springer, Berlin Heidelberg.
Kar, S. K.,	2005	A fully-implicit semi-Lagrangian hydrostatic model of atmospheric dynamics. *AMS 21st Conference on Weather Analysis and Forecasting and 17th Conference on Numerical Weather Prediction, 1–5 August 2005. Extended abstract P1.84, available at https://ams.confex.com*

Kar, S. K.	2012	An explicit time-difference scheme with an Adams-Bashforth predictor and a trapezoidal corrector. *Mon. Wea. Rev.,* **140,** 307–322.
Kar, S. K., R. J. Purser, G. Gopalakrishnan, and G. J. DiMego	2004	A fully-implicit semi-Lagrangian nonhydrostatic model employing a pressure-hybrid vertical coordinate. *AMS 20th Conference on Weather Analysis and Forecasting and 16th Conference on Numerical Weather Prediction, 12–15 January 2004, Extended abstract 25.2, available at https://ams.confex.com*
Lambert, J. D.	1992	*Numerical Methods for Ordinary Differential Systems.* Wiley, New York.
Purser, R. J.	1983	A fully implicit Lagrangian method as an alternative approach to numerical weather prediction. *Met Office MET O 11 Technical Note* **164,** 28 pp.
Purser, R. J.	2007	Accuracy considerations of time-splitting methods for models using two-time-level schemes *Mon. Wea. Rev.,* **135,** 1158–1164.
Purser, R. J.	2008	Generalized Fibonacci grids; A new class of structured, smoothly adaptive multi-dimensional computational lattices. *NOAA/NCEP Office Note* **455,** 37 pp.
Reisner, J. M., V. A. Mousseau, A. A. Wyszogrodzki, and D. A. Knoll	2005	An implicitly balanced hurricane model with physics-based preconditioning. *Mon. Wea. Rev.,* **133,** 1003–1022.
Robert, A.	1981	A stable numerical integration scheme for the primitive meteorological equations. *Atmos. Ocean,* **19,** 35–46.
Robert, A.	1982	A semi-Lagrangian and semi-implicit numerical integration scheme for the primitive meteorological equations. *J. Meteor. Soc. Japan,* **60,** 319–325.
Simmons, A. J., and C. Temperton	1997	Stability of a two-time-level semi-implicit integration scheme for gravity wave motion. *Mon. Wea. Rev.,* **125,** 600–615.
Staniforth, A., and J. Côté	1991	Semi-Lagrangian integration schemes for atmospheric models — a review. *Mon. Wea. Rev.,* **119,** 2206–2223.
Swinbank, R., and R. J. Purser	2006	Fibonacci grids; A novel approach to atmospheroc modeling. *Quart. J. Royal Meteor. Soc.,* **132,** 1769–1793.
Weijer, W., H. A. Dijkstra, H. Öksüzoğlu, F. W. Wubs, and A. C. de Niet	2003	A fully-implicit model of the global ocean circulation. *J. Comp. Phys.,* **192,** 452–470.
Yeh, K.-S., J. Côté, J., S. Gravel, A. Méthot, A. Patoine, M. Roch, and A. Staniforth	2002	The CMC-MRB Global Environmental Multiscale (GEM) Model. Part III: Nonhydrostatic formulation. *Mon. Wea. Rev.,* **130,** 339–356.