

A CLASS OF ROSENBRCK-TYPE SCHEMES FOR SECOND-ORDER NONLINEAR SYSTEMS OF ORDINARY DIFFERENTIAL EQUATIONS

S. GOYAL

Department of Mathematics, Tuskegee Institute, Tuskegee, AL 36088, U.S.A.

S. M. SERBIN

Department of Mathematics, University of Tennessee, Knoxville, TN 37996-1300, U.S.A.

(Received October 1985)

Communicated by E. Y. Rodin

Abstract—We develop a class of generalized Rosenbrock-type schemes for second-order nonlinear systems of ordinary differential equations. We convert the second-order systems to equivalent first-order form, and then employ the square of the Jacobian. These methods when applied to a linear time-invariant system $U'' + AU = 0$, reproduce a class of schemes given by Baker and Bramble that are derived from a particular class of rational approximations to the exponential with denominators of the form $(1 - \gamma^2 z^2)^s$ for an s -stage method. For our problem, then, an s -stage scheme requires the solution of $2s$ linear algebraic systems at each time step, with the same real matrix. We employ the theory of Butcher [1-4] series to develop order conditions and then present specific examples of fourth-order methods which are unconditionally stable by appropriate choice of parameter γ^2 . Numerical results, confirming the rate of convergence, are presented.

1. INTRODUCTION

In Ref. [5], Baker and Bramble constructed a class of single-step schemes for approximating the solution of certain second-order linear evolution equations with time-independent coefficients. These schemes are based upon a class of rational approximations to the exponential which are analytic in a neighborhood of the imaginary axis and which possess desirable accuracy and stability properties on the imaginary axis, and thus are particularly suited to this class of problems when they are transformed into equivalent first-order systems.

The goal of this investigation is to propose and analyze a class of Rosenbrock-type schemes for the special second-order system

$$\begin{aligned}U'' &= G(U, t) \quad t \in [0, T] \\U(0) &= U_0, \quad U'(0) = V_0,\end{aligned}\tag{1}$$

where $U: [0, \hat{T}] \rightarrow \mathbb{R}^n$ (or, more generally, we could replace \mathbb{R}^n with a Banach space X). These schemes are such that when $G(U, t) = -AU$, for $A \in L(\mathbb{R}^n, \mathbb{R}^n)$ positive definite and symmetric, the Baker-Bramble methods are reproduced, and thus are distinguished from other Rosenbrock methods which have previously appeared.

These include, for example, the original works of Rosenbrock [6], Calahan [7], Wanner [8], Kaps and Rentrop [9] and Kaps and Wanner [10] and many others. A recent analysis of Rosenbrock methods for stiff problems has been given by Verwer [11]. These methods, though, have been proposed for first-order problems. On the other hand, there are many methods which have been proposed recently to deal explicitly with second-order systems [cf. 12-15]. Properties of some of these methods have been analyzed by Thomas [16]. These methods have the property that the stages are implicitly coupled; in contrast, in our method, the stages are determined successively by solving only linear systems.

In order to propose and analyze our schemes, we shall perform the usual conversion of the second-order problem (1) into first-order form by defining $y = [U, V, t]^T$, where $V = U'$, and thus

obtain the autonomous system

$$y' = f(y) = \left[G \begin{pmatrix} \dot{U} \\ 1 \end{pmatrix}, t \right], \quad y(0) = [U_0 \ V_0 \ 0]^T \equiv y_0. \tag{2}$$

We then pose, in Section 2, a Rosenbrock-type method for expression (2), using the square of the Jacobian to generate the linear systems which must be solved at each time step. It is this approach which distinguishes the scheme from other Rosenbrock methods previously developed. In Section 3, we employ the concept of Butcher series to develop order conditions for our general scheme, then fix specifically upon a set of fourth-order schemes, explicitly stating the order conditions and listing several sets of parameters which satisfy these conditions. In Section 4, we justify our contention that our schemes reproduce the Baker–Bramble methods when applied to a linear homogeneous test problem, and having thus established this relation, we use a result of Nørsett and Wanner [17] to obtain conditions on a parameter of the scheme which gives unconditional stability. We conclude in Section 5 with the specific implementation of the two-stage scheme for the original problem (1), showing that for fourth-order, we must solve four $n \times n$ linear systems with the *same* matrix per time step. We provide several examples of the application of our scheme, confirming empirically the expected fourth-order error reduction.

2. FORMULATION OF THE METHOD

Having written the system in first-order form (2), we propose an s -stage method ($s \in \mathbb{Z}^+$) as follows. Define, for $\gamma^2 \geq 0$ a real parameter,

$$E \equiv I - \gamma^2 h^2 f_y^2(y_0) \quad \text{for } i = 1, \dots, s. \tag{3}$$

(Note: we use the notation γ^2 for the parameter to conform with notation established in Ref. [17] which we shall invoke in the stability discussion.) Then, for certain coefficients $\{a_{ij}\}$, $\{b_{ij}\}$, $\{c_{ij}\}$, $\{d_{ij}\}$, $\{e_{ij}\}$, $j < i$, and $\{\phi_i\}$, $\{\theta_i\}$ and $\{m_i\}$ obtain vectors k_i via

$$Ek_i = f(y_0 + h \sum_{j=1}^{i-1} a_{ij} k_j) + \phi_i h f_y(y_0) f(y_0 + h \sum_{j=1}^{i-1} e_{ij} k_j) + \theta_i h f_y(y_0 + h \sum_{j=1}^{i-1} b_{ij} k_j) f(y_0 + h \sum_{j=1}^{i-1} d_{ij} k_j) + \sum_{j=1}^{i-1} c_{ij} k_j. \tag{4}$$

Then, determine y_1 , the approximate solution at the next time station, by

$$y_1 = y_0 + h \sum_{i=1}^s m_i k_i. \tag{5}$$

Due to the specific form of the system (2), we will later see that each step (4) requires that we solve two linear systems, and so in total each time step requires solution of $2s$ linear systems with the same $n \times n$ matrix. The connection with the Baker–Bramble method comes from the use of the square of the Jacobian f_y in the definition of the operator E . This identification, though, as well as the desire to produce in particular a fourth-order scheme with $s = 2$, seems to dictate the inclusion of the terms shown on the r.h.s. of expression (4).

Following Ref. [10], we first pose the method (3)–(5) in an equivalent form for the purpose of derivation of the order conditions.

Lemma 1

The algorithm (3)–(5) is equivalent to the scheme

$$u_i = y_0 + \sum_{j=1}^{i-1} \alpha_{ij} g_j, \tag{6}$$

$$v_i = y_0 + \sum_{j=1}^{i-1} \beta_{ij} g_j, \tag{7}$$

$$w_i = y_0 + \sum_{j=1}^{i-1} \Delta_{ij} g_j, \tag{8}$$

$$z_i = y_0 + \sum_{j=1}^{i-1} e_{ij} g_j, \tag{9}$$

$$g_i = hf(u_i) + \phi_i h^2 f_y(y_0) f(z_i) + \theta_i h^2 f_y(v_i) f(w_i) + h^2 f_y^2(y_0) \sum_{j=1}^i \gamma_{ij} g_j \tag{10}$$

and

$$y_1 = y_0 + \sum_{i=1}^s \mu_i g_i, \tag{11}$$

where the coefficients α_{ij} , β_{ij} , Δ_{ij} , ϵ_{ij} , γ_{ij} , and μ_i are determined implicitly by

$$a_{ij} = \sum_{k=1}^{i-1} \alpha_{ik} (\delta_{kj} - c_{kj}), \tag{12}$$

$$b_{ij} = \sum_{k=1}^{i-1} \beta_{ik} (\delta_{kj} - c_{kj}), \tag{13}$$

$$c_{ij} = \sum_{k=1}^{i-1} \gamma_{ik} (\delta_{kj} - c_{kj}) / \gamma^2, \tag{14}$$

$$d_{ij} = \sum_{k=1}^{i-1} \Delta_{ik} (\delta_{kj} - c_{kj}), \tag{15}$$

$$e_{ij} = \sum_{k=1}^{i-1} \epsilon_{ik} (\delta_{kj} - c_{kj}) \tag{16}$$

and

$$m_j = \sum_{k=1}^s \mu_k (\delta_{kj} - c_{kj}), \tag{17}$$

where $\gamma_{ii} = \gamma^2$ and δ_{kj} is the Kronecker index. A detailed proof of this lemma, patterned after the argument in Ref. [10], may be found in Ref. [18].

3. ORDER CONDITIONS

We shall use the technique of Butcher series, especially as found in Hairer and Wanner [19] and Nørsett and Wolfbrandt [20], to determine order conditions for the method (3)–(5). To begin, we recall briefly certain definitions, notations and properties of rooted trees. We indicate only those ideas which are specifically required in our analysis; for a broader picture, the interested reader may consult Refs [19–21].

Let T denote the set of all rooted trees. The notation $t = [t_1, \dots, t_m]$ expresses the notion that the trees t_1, \dots, t_m remain after the root of t and the adjacent arcs have been removed (Note: in this section, t stands for a tree, not the independent time variable; there should be no difficulty in distinguishing the meaning of t from the context.) We denote by $\rho(t)$ the number of nodes in t ; ϕ and τ denote the trees with zero and one nodes, respectively. LT denotes the set of monotonically labeled trees.

Now, for $t = [t_1, \dots, t_m] \in T$, the elementary differentials $F(t)$ are defined recursively by

$$F(\phi)(y) = y; \quad F(\tau)(y) = f(y)$$

and

$$F(t)(y) = f^{(m)} \cdot [F(t_1)(y), \dots, F(t_m)(y)]. \tag{18}$$

Then, for a map $a: T \rightarrow \mathbb{R}$, the Butcher series $B(a, y_0) \in \mathbb{R}^n$ about $y_0 \in \mathbb{R}^n$ is defined by

$$B(a, y_0) = \sum_{t \in LT} a(t) F(t)(y_0) \frac{h^r}{r!}, \quad r = \rho(t). \tag{19}$$

We also require the notion of the derivative of the map $a(t)$: if $t = \phi$, $a'(t) = 0$; if $t = \tau$, $a'(\tau) = a(\phi) = 1$, and if $t = [t_1, \dots, t_m]$, $a'(t) = \rho(t)a(t_1) \dots a(t_m)$.

The following two theorems, which may be found in Ref. [20] or Ref. [21], are instrumental in the development of the order conditions for our scheme.

Theorem 1 [20]

If $y = B(\psi, y_0)$, with $\psi(\phi) = 1$, then $hf(y) = B(\psi', y_0)$.

Theorem 2 [20]

Let $B(a, y_0)$ and $B(b, y_0)$ be two Butcher series, with $b(\phi) = 0$ and $a(\phi) = 1$. Then $hf'(B(a, y_0))B(b, y_0)$ is a Butcher series $B(a \circ b, y_0)$, where the composition is defined, for $t = [t_1 \dots t_m]$, by

$$(a \circ b)(t) = \begin{cases} \rho(t)b(t_1) & \text{for } m = 1 \\ \rho(t) \sum_{i=1}^m \left(\prod_{j=1, j \neq i}^m a(t_j) \right) b(t_i), & \text{for } m > 1. \end{cases} \tag{20}$$

Further, $(a \circ b)(\phi) = (a \circ b)(\tau) = 0$.

We now use the above machinery to develop the order conditions specific to our method via the following theorem. We shall assume that the numerical solution y_1 given by expression (5) has a Butcher series representation.

Theorem 3

The terms u_i, v_i, w_i, z_i, g_i and y_i defined by expressions (6)–(11), which are functions of the stepsize h , are Butcher series given by

$$u_i(h) = B(u_i, y_0), v_i(h) = B(v_i, y_0), w_i(h) = B(w_i, y_0)$$

and

$$z_i(h) = B(z_i, y_0), g_i(h) = B(g_i, y_0), y_1(h) = B(y_1, y_0), \tag{21}$$

where the coefficients $u_i(t), v_i(t), w_i(t), x_i(t), g_i(t)$ and $y_1(t)$ are determined recursively by:

$$u_i(\phi) = 1, u_i(t) = \sum_{j=1}^{i-1} \alpha_{ij} g_j(t), t \neq \phi; \tag{22}$$

$$v_i(\phi) = 1, v_i(t) = \sum_{j=1}^{i-1} \beta_{ij} g_j(t), t \neq \phi; \tag{23}$$

$$w_i(\phi) = 1, w_i(t) = \sum_{j=1}^{i-1} \Delta_{ij} g_j(t), t \neq \phi; \tag{24}$$

$$x_i(\phi) = 1, x_i(t) = \sum_{j=1}^{i-1} \epsilon_{ij} g_j(t), t \neq \phi; \tag{25}$$

and

$$y_1(\phi) = 1, y_1(t) = \sum_{i=1}^s \mu_i g_i(t); \tag{26}$$

where, for $t = [t_1 \dots t_m]$, $\mathcal{G}_i(\phi) = 0$, $\mathcal{G}_i(\tau) = 1$ and

$$\mathcal{G}_i(t) = \omega'_i(t) + \phi_i \begin{cases} \rho(t)x'_i(t_1), & m = 1 \\ 0, & m > 1 \end{cases} + \theta_i \begin{cases} \rho(t)\omega'_i(t_1), & m = 1 \\ \rho(t) \sum_{l=1}^m \left(\prod_{\substack{j=1 \\ j \neq l}}^m v_i(t_j) \right) \omega'_i(t_l), & m > 1 \end{cases} + \begin{cases} \rho(t)[\rho(t) - 1] \sum_{j=1}^i \gamma_{ij} \mathcal{G}_j(\hat{t}_1), & t = [[\hat{t}_1]] \\ 0, & \text{otherwise} \end{cases} \quad (27)$$

(Note: we interpret any vacuous sum to be zero.)

Proof. By assumption, $y_1(h)$ is a Butcher series, so $g_i = B(\mathcal{G}_i, y_0)$ is a Butcher series. Then, from expression (6),

$$\begin{aligned} u_i(h) &= y_0 + \sum_{j=1}^{i-1} \alpha_{ij} \mathcal{G}_j(h) \\ &= y_0 + \sum_{j=1}^{i-1} \alpha_{ij} B(\mathcal{G}_j, y_0) \\ &= y_0 + \sum_{j=1}^{i-1} \alpha_{ij} \sum_{t \in LT} \frac{h^{\rho(t)}}{\rho(t)!} \mathcal{G}_j(t) F(t)(y_0) \\ &= y_0 + \sum_{t \in LT} \frac{h^{\rho(t)}}{\rho(t)!} \left[\sum_{j=1}^{i-1} \alpha_{ij} \mathcal{G}_j(t) F(t)(y_0) \right] \\ &= \sum_{t \in LT} \frac{h^{\rho(t)}}{\rho(t)!} \omega'_i(t) F(t)(y_0) = B(\omega'_i, y_0). \end{aligned}$$

Similarly, expressions (23)–(26) follow from expressions (7)–(9) and (11), respectively.

Now, to prove expression (27), we consider each term in expression (10). First, by Theorem 1, $hf(u_i) = B(\omega'_i, y_0)$. Next, since y_0 is the Butcher series $B(\sigma, y_0)$, where $\sigma(\phi) = 1$ and $\sigma(t) = 0$, $\rho(t) \geq 1$, by Theorem 1, $hf(z_i) = B(x'_i, y_0)$, and then identifying $b = x'_i$ and $a = \sigma$ in Theorem 2, we have $h^2 f_y(y_0) f(z_i) = hf'(B(\sigma, y_0)) B(x'_i, y_0)$ is a Butcher series $B(\sigma \circ x'_i, y_0)$, with

$$(\sigma \circ x'_i)(t) = \begin{cases} \rho(t)x'_i(t_1), & m = 1 \\ 0, & m > 1. \end{cases}$$

For the third term in expression (27), we apply Theorem 2, with $B(b, y_0) = B(\omega'_i, y_0) = hf(\omega_i)$ and $B(a, y_0) = B(v_i, y_0)$, so that

$$h^2 f_y(v_i) f(\omega_i) = hf'(B(v_i, y_0)) B(\omega'_i, y_0) = B(v_i \circ \omega'_i, y_0),$$

where

$$(v_i \circ \omega'_i)(t) = \begin{cases} \rho(t)\omega'_i(t_1), & m = 1 \\ \rho(t) \sum_{l=1}^m \left[\prod_{\substack{j=1 \\ j \neq l}}^m v_i(t_j) \right] \omega'_i(t_l), & m > 1. \end{cases}$$

Lastly, the final term of expression (10) can be treated by two successive applications of Theorem 2. Alternatively, as motivated by Ref. [10], note that

$$[f_y(y_0)]^2 F(\hat{t}_1)(y_0) = F[[\hat{t}_1]](y_0).$$

Then, expressing

$$g_f(h) = \sum_{i \in LT} \frac{h^{\rho(\hat{t}_1)}}{\rho(\hat{t}_1)!} \mathcal{G}_f(\hat{t}_1) F(\hat{t}_1)(y_0),$$

we obtain

$$h^2 f_y^2(y_0) g_j(h) = \sum_{\hat{t}_1 \in LT} \frac{h^{\rho(\hat{t}_1)+2}}{\rho(\hat{t}_1)!} \mathcal{G}_j(\hat{t}_1) F([\hat{t}_1])(y_0). \tag{28}$$

Letting $t = [[\hat{t}_1]]$ and noting that $\rho(\hat{t}_1) + 2 = \rho(t)$, it follows that

$$h^2 f_y^2(y_0) \sum_{j=1}^i \gamma_{ij} g_j = \sum_{j=1}^i \gamma_{ij} \left\{ \sum_{\substack{t \in LT \\ t = [[\hat{t}_1]]}} \rho(t) [\rho(t) - 1] \frac{h^{\rho(t)}}{\rho(t)!} \mathcal{G}_j(\hat{t}_1) F(t)(y_0) \right\},$$

which establishes the last term of expression (10). ■

(Note: expression (28) is a Butcher series whose nonzero terms belong to trees with a single-branched root whose first node above the root is also singly branched, e.g. $\begin{array}{c} \bullet \\ \nearrow \\ \bullet \end{array} = [[\tau]].$)

Now, the order conditions are easily established in the usual way by noting that exact solution at the next time station, say $t_0 + h$, is by Taylor series a Butcher series $B(\pi, y_0)$, where $\pi(t) = 1 \forall t \in T$. We then have that the method (3)–(5) has order at least ν if $\mathcal{Y}_1(t)$, as given by expression (26), satisfies $\mathcal{Y}_1(t) = 1 \forall t$ such that $\rho(t) \leq \nu$. We thus must establish recursively $\{\mathcal{G}_j(t), \rho(t) \leq \nu\}$.

In Table 1, we show for the case $s = 2, \nu = 4$ these computations. We use the facts that [by expressions (12)–(17)] $\alpha_{21} = a_{21}, \beta_{21} = b_{21}, \gamma_{21} = \gamma^2 c_{21}, \Delta_{21} = d_{21}, \epsilon_{21} = e_{21}$, and setting $\eta_1 = \phi_1 + \theta_1$, we employ expression (22)–(27) to establish $\mathcal{G}_1(t)$ and $\mathcal{G}_2(t)$.

In turn, then, since from expression (17), $\mu_1 = m_1 + c_{21} m_2, \mu_2 = m_2$, the conditions for order 4, from equations (26), are then determined as listed in Table 2.

Of course, it is possible to consider other cases than $s = 2, \nu = 4$; for the sake of brevity here, we shall not present any further order conditions, but rather refer the reader to Ref. [18]. We will remark, though, that it is not possible to achieve order $\nu = 2s$ for $s > 2$, as the number of structurally different trees of order up to ν increases much more rapidly than the number of parameters.

We have approached the question of solving the order equations of Table 2 from two different points of view. First, since there are 10 parameters and only 8 equations, we have the freedom to

Table 1. Values of $\mathcal{G}_1(t)$ and $\mathcal{G}_2(t)$ for trees of order ≤ 4









Number	t	$\mathcal{G}_1(t)$	$\mathcal{G}_2(t)$
1		τ	1
2		$[\tau]$	$2(a_{21} + \phi_2 + \theta_2)$
3		$[\tau, \tau]$	$3(a_{21}^2 + 2\theta_2 b_{21})$
4		$[[\tau]]$	$6[a_{21}\eta_1 + \phi_2 e_{21} + \theta_2 d_{21} + \gamma^2(1 + c_{21})]$
5		$[\tau, \tau, \tau]$	$4(a_{21}^3 + 3\theta_2 b_{21}^2)$
6		$[[\tau, \tau]]$	$12(\phi_2 e_{21}^2 + \theta_2 d_{21}^2)$
7		$[\tau, [\tau]]$	$8[a_{21}^2 \eta_1 + \theta_2 b_{21}(\eta_1 + d_{21})]$
8		$[[[\tau]]]$	$24[\gamma^2 a_{21} \quad 24[\gamma^2(2a_{21} + \eta_1 c_{21} + \phi_2 + \theta_2) + \eta_1(\phi_2 e_{21} + \theta_2 d_{21})]$

Table 2. Order conditions for $s = 2, v = 4$

Number		Condition
1	τ	$m_1 + (1 + c_{21})m_2 = 1$
2	$[\tau]$	$m_1\eta_1 + m_2(c_{21}\eta_1 + a_{21} + \phi_2 + \theta_2) = \frac{1}{2}$
3	$[\tau, \tau]$	$m_2\left(\frac{a_{21}^2}{2} + \theta_2 b_{21}\right) = \frac{1}{6}$
4	$[[\tau]]$	$\gamma^2[m_1 + m_2(1 + 2c_{21})] + m_2[a_{21}\eta_1 + e_{21}\phi_2 + \theta_2 d_{21}] = \frac{1}{6}$
5	$[\tau, \tau, \tau]$	$m_2\left(\frac{a_{21}^3}{6} + \theta_2 \frac{b_{21}^2}{2}\right) = \frac{1}{24}$
6	$[[\tau, \tau]]$	$m_2\left(\phi_2 \frac{e_{21}^2}{2} + \theta_2 \frac{d_{21}^2}{2}\right) = \frac{1}{24}$
7	$[\tau, [\tau]]$	$m_2[a_{21}^2\eta_1 + \theta_2 b_{21}(\eta_1 + d_{21})] = \frac{1}{8}$
8	$[[[\tau]]]$	$\gamma^2[m_1\eta_1 + m_2(2c_{21}\eta_1 + 2a_{21} + \phi_2 + \theta_2)] + m_2\eta_1(\phi_2 e_{21} + \theta_2 d_{21}) = \frac{1}{24}$

choose some of these parameters, as long as the resulting equations are still consistent. Of course, it would be most prudent to try to set $b_{21} = 0$; unfortunately, this produces an inconsistency. We can always set $\phi_1 = 0$, and thus $\theta_1 \equiv \eta_1$, without losing any degree of freedom. Now, the choice of parameters must necessarily be influenced by stability considerations. We shall establish (by appealing to known results given by Nørsett and Wanner [17]) that our scheme is unconditionally stable for

$$\gamma^2 \geq \frac{3 + \sqrt{7}}{12} \approx 0.47048.$$

In our first approach, to streamline the method somewhat, we set $d_{21} = 0, e_{21} = a_{21}$ and $\gamma^2 = \frac{1}{2}$. It then follows that θ_1 must satisfy a certain eleventh-degree polynomial equation:

$$98304\theta_1^{11} - 270336\theta_1^{10} + 236544\theta_1^9 + 29184\theta_1^8 - 207808\theta_1^7 + 139552\theta_1^6 + 18032\theta_1^5 - 64412\theta_1^4 + 10080\theta_1^3 + 23400\theta_1^2 - 113824\theta + 23349 = 0. \quad (29)$$

We then solve for θ_1 by using the IMSL polynomial solver ZPOLR, and then, appealing to the routine ZSCNT, return to the original nonlinear system of Table 2 and reduce the parameter γ^2 to $(3 + \sqrt{7})/12$, yielding the parameter set below in Table 3.

In an attempt to further simplify the scheme, we have tried to set $e_{21} = a_{21}, \phi_1 = \phi_2 = 0$ and $c_{21} = 0$. In this case, we have obtained a solution with $\gamma^2 \approx 1.62$, which again yields an unconditionally stable method, but as the error constant of this scheme will necessarily be larger than that of the previous one, we shall not consider it further.

Finally, we observe that by replacing the fourth and eighth equations of Table 3 by the equations

$$m_1 + m_2(1 + 2c_{21}) = 0, \quad (30)$$

$$m_2(a_{21}\eta_1 + e_{21}\phi_2 + \theta_2 d_{21}) = 1/6, \quad (31)$$

$$m_1\eta_1 + m_2(2c_{21}\eta_1 + 2a_{21} + \phi_2 + \theta_2) = 0 \quad (32)$$

and

$$m_2\eta_1(\phi_2 e_{21} + \theta_2 d_{21}) = \frac{1}{24}, \quad (33)$$

we may then attempt a solution which leaves the parameter γ^2 free. This produces 10 equations in the remaining 10 unknowns. After tedious algebraic reductions, we determine that η_1 must satisfy the cubic equation

$$24\eta_1^3 - 12\eta_1^2 - 4\eta_1 + 1 = 0, \quad (34)$$

which has the real roots 0.6571366762993064, -0.3423473692310087 and 0.1852106929317023.

Table 3. Solution of order equations; forced $d_{21} = 0, e_{21} = a_{21}$

$a_{21} = -0.7777536224724765$	$m_1 = 1.022753184288266$
$b_{21} = 1.117655988539988$	$m_2 = 0.2080352101413627$
$c_{21} = -1.109377052294547$	$\gamma_2 = (3 + \sqrt{7})/12$
$\eta_1 = \theta_1 = 0.5444631141603234$	
$\phi_2 = 0.6622450174040982$	
$\theta_2 = 0.4462326530351922$	

Table 4. Two solutions of order equations; γ^2 free parameter

	$\eta_1 = 0.1852106929317023$	$\eta_1 = 0.6571366762993064$
a_{21}	0.8812757541276511	0.1629806272136976
b_{21}	0.5397522691127052	0.5325697649852304
c_{21}	2.799573347440441	-1.037190241336529
d_{21}	0.3923761529540218	0.04590171220992117
e_{21}	-1.267840885278147	1.649701575706587
ϕ_2	0.006569219401703724	0.03152698463159001
θ_2	-1.583910034725311	0.2996484372403217
m_1	2.357197285405745	1.035856721220791
m_2	-0.357197285405745	0.964143278779209

For each of these values of η_1 , further algebraic reduction shows that a_{21} must satisfy a fifth-degree polynomial equation. After having examined the various solution sets, we show in Table 4 below two such sets of parameters which seem to give the most acceptable results.

4. STABILITY

Our linear stability analysis is predicated upon the fact that the schemes we have proposed are generalizations of the Baker–Bramble method to nonlinear equations, which we now establish. It suffices to consider the scalar equation

$$U'' = -\lambda^2 U, \tag{35}$$

which becomes, in first-order form,

$$y' = Ay, \tag{36}$$

with

$$A = \begin{bmatrix} 0 & 1 \\ -\lambda^2 & 0 \end{bmatrix}. \tag{37}$$

The Baker–Bramble scheme is obtained by approximating the solution for equation (36) over one time step $y(h) = \exp(hA)y(0)$ by

$$y_1 = R(hA)y_0, \tag{38}$$

where

$$R(z) = P(z)/(1 - \gamma^2 z^2)^s \tag{39}$$

and $P(z)$ is a polynomial of degree $2s$ given specifically by [cf. 17]

$$P(z) = \sum_{j=0}^{2s} z^{2s-j} \sum_{i=0}^s \binom{s}{i} \frac{(-\gamma^2)^{s-i}}{(2i-j)!}, \quad i_0 = \left\lfloor \frac{j+1}{2} \right\rfloor. \tag{40}$$

Letting

$$P(z) \equiv \sum_{j=0}^{2s} p_j z^j,$$

it is easy to show that

$$R(h\lambda) = \frac{1}{[1 + \gamma^2(h\lambda)^2]^s} \left[\begin{array}{c} \sum_{j=0}^s (-1)^j p_{2j}(h\lambda)^{2j} h \sum_{j=0}^{s-1} (-1)^j p_{2j+1}(h\lambda)^{2j} \\ \lambda \sum_{j=0}^{s-1} (-1)^{j+1} p_{2j+1}(h\lambda)^{2j+1} \sum_{j=0}^s (-1)^j p_{2j}(h\lambda)^{2j} \end{array} \right] \quad (41)$$

We wish to show that when the scheme (3)–(5) is applied to expression (36), there results a method of the form (38), with $R(h\lambda)$ having the same structure as in expression (41). Since the number of order conditions for the s -stage method when applied to the linear problem will be much less than in the nonlinear case, we will then justifiably assume that the parameters can be chosen to obtain the specific rational function used by Baker and Bramble, i.e. with $P(z)$ given as in expression (40). That is, if the order is assumed to be at least $2s$ for the linear problem, we will have obtained exactly the Baker–Bramble scheme.

Referring back to expression (3), we first see that when $f(y) = Ay$

$$E = \begin{bmatrix} 1 + \gamma^2 z^2 & 0 \\ 0 & 1 + \gamma^2 z^2 \end{bmatrix}, \quad (42)$$

where we let $z = h\lambda$. Then, expression (4) can easily be expressed as

$$k_i = \frac{1}{1 + \gamma^2 z^2} \left\{ \begin{bmatrix} \psi_i z^2 & h \\ -\lambda^2 h & \psi_i z^2 \end{bmatrix} y_0 + \sum_{j=1}^{i-1} \begin{bmatrix} \omega_j z^2 & \alpha_j h \\ -\lambda^2 h \alpha_{ij} & \omega_j z^2 \end{bmatrix} k_j \right\}, \quad (43)$$

where $\psi_i = \phi_i + \theta_i$ and $\omega_{ij} = \psi_i \beta_{ij} + \gamma_{ij}$. Write expression (43) as

$$k_i = \frac{1}{1 + \gamma^2 z^2} \left(P_i y_0 + \sum_{j=1}^{i-1} Q_{ij} k_j \right). \quad (44)$$

The following is then easily established by induction.

Proposition 1

$$k_i = \frac{1}{(1 + \gamma^2 z^2)^i} \left\{ \begin{array}{cc} \Psi_{11}^i(z) & h\Psi_{12}^i(z) \\ \lambda\Psi_{21}^i(z) & \Psi_{22}^i(z) \end{array} \right\} y_0$$

where $\Psi_{11}^i(z)$ and $\Psi_{22}^i(z)$ are of degree $2i$ and $\Psi_{12}^i(z)$ is of degree $(2i - 2)$, $\Psi_{21}^i(z)$ is of degree $(2i - 1)$. ■

Now, considering expression (5), we have

$$\begin{aligned} y_1 &= y_0 + \sum_{i=1}^s \mu_i k_i = \left\{ 1 + \sum_{i=1}^s \frac{\mu_i}{(1 + \gamma^2 z^2)^i} \begin{bmatrix} \Psi_{11}^i(z) & h\Psi_{12}^i(z) \\ \lambda\Psi_{21}^i(z) & \Psi_{22}^i(z) \end{bmatrix} \right\} y_0 \\ &= (1 + \gamma^2 z^2)^{-s} \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \sum_{i=1}^s \mu_i (1 + \gamma^2 z^2)^{s-i} \begin{bmatrix} \Psi_{11}^i(z) & h\Psi_{12}^i(z) \\ \lambda\Psi_{21}^i(z) & \Psi_{22}^i(z) \end{bmatrix} \right\} y_0 \\ &= (1 + \gamma^2 z^2)^{-s} \begin{bmatrix} R_{11}(z) & hR_{12}(z) \\ \lambda R_{21}(z) & R_{22}(z) \end{bmatrix} y_0, \end{aligned} \quad (45)$$

with $\deg R_{11} = 2s = \deg R_{22}$, $\deg R_{12} = 2s - 2$, $\deg R_{21} = 2s - 1$. Then, comparing (45) with (41), we see that both methods have exactly the same form.

While a direct stability analysis for the case $s = 2$ is easily performed, as is presented in Ref. [18], we may appeal to the result of Nørsett and Wanner [17] to establish the next stability result.

Theorem 4

The method (3)–(5) with $s = 2$ and parameters selected so that the order is $\nu = 4$ is I-stable [i.e. unconditionally stable for the test problem (36)–(37)] iff $\gamma^2 \geq (3 + \sqrt{7})/12$. ■

We also show in Ref. [18] by elementary means that for $0 \leq \gamma^2 \leq (3 - \sqrt{7})/12$, the scheme is conditionally stable, with the requirement that

$$(h\lambda)^2 \leq \frac{576\gamma^4 - 288\gamma^2 + 8}{-1152\gamma^6 + 624\gamma^4 - 48\gamma^2 + 1}. \tag{46}$$

5. IMPLEMENTATION AND EXAMPLES

For the original problem (1), the two-stage scheme (3)–(5) can be implemented as follows. Let $k_1 = [p_i, q_i, \tau_i]^T$. From (2), the Jacobian

$$f_y \equiv J = \begin{bmatrix} 0 & I & 0 \\ G_U & 0 & G_t \\ 0 & 0 & 0 \end{bmatrix}, \quad \text{so} \quad J^2 = \begin{bmatrix} G_U & 0 & G_t \\ 0 & G_U & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Let us write the scheme in a form which exhibits progression from time station t_n to $t_{n+1} = t_n + h$. When arguments are unspecified, we are evaluating at $t = t_n$.

Let $L = I - \gamma^2 h^2 G_U$. We determine $\{p_i, q_i\}_{i=1,2}$ via

$$Lp_1 = V_n + \eta_1 hG + \gamma^2 h^2 G_t, \tag{47}$$

$$Lq_1 = G + \eta_1 [G_U V_n + G_t], \tag{48}$$

$$Lp_2 = \gamma^2 h^2 (1 + c_{21})G_t + V_n + a_{21}hq_1 + \phi_2 hG(U_n + e_{21}hp_1, t_n + e_{21}h) + \theta_2 hG(U_n + d_{21}hp_1, t_n + d_{21}h) + c_{21}p_1, \tag{49}$$

$$Lq_2 = \phi_2 hG_U \cdot (V_n + e_{21}hq_1) + G(U_n + a_{21}hp_1, t_n + a_{21}h) + \phi_2 hG_t + \theta_2 hG_U(U_n + b_{21}hp_1, t_n + b_{21}h) \cdot (V_n + hd_{21}q_1) + \theta_2 hG_t(U_n + b_{21}hp_1, t_n + b_{21}h) + c_{21}q_1, \tag{50}$$

$$U_{n+1} = U_n + h\{m_1 p_1 + m_2 p_2\} \tag{51}$$

and

$$V_{n+1} = V_n + h\{m_1 q_1 + m_2 q_2\}. \tag{52}$$

We have applied our scheme to approximate the solution of two model problems. The first is a modification of a nonlinear wave-propagation model suggested by Fermi *et al.* [22].

Problem 1. Seek $u(t) \in \mathbb{R}^n$ satisfying

$$\ddot{u}_j(t) = F(u_{j+1} - u_j) - F(u_j - u_{j-1}) + g_j(t), \tag{53}$$

where

$$F(u) \equiv \lambda u + \alpha u^p, \quad u_0(t) = u_{n+1}(t) = 0, \quad u_j(0) = \sin\left(\frac{2\pi j}{n+1}\right), \quad \dot{u}_j(0) = 0 \quad j = 1, \dots, n,$$

and we have selected $g_j(t)$ so that the exact solution is given by

$$u_j(t) = \sin\left(\frac{2\pi j}{n+1}\right) \cos t. \tag{54}$$

We have introduced the parameter λ to allow control of the spectral radius of the Jacobian, and thus to introduce stiffness attributes. We report results with $n = 20$.

Problem 2. The equation of motion of a soliton in an exponential lattice given by Toda [23] is the highly nonlinear system

$$\begin{aligned} \ddot{u}_j(t) &= 2e^{-u_j} - e^{-u_{j-1}} - e^{-u_{j+1}}, \\ u_j(0) &= -\log[1 + \beta^2 \operatorname{sech}^2(\alpha j)], \\ \dot{u}_j(0) &= \frac{2\beta^3 \operatorname{sech}^2(\alpha j) \tanh(\alpha j)}{[1 + \beta^2 \operatorname{sech}^2(\alpha j)]}, \end{aligned} \tag{55}$$

which has a solution given by

$$e^{-u} - 1 = \beta^2 \operatorname{sech}^2(\alpha j + \beta t), \tag{56}$$

where

$$\beta = \sinh \alpha.$$

We have chosen $\alpha = 2$ and $n = 20$. For each problem, with M time steps, $Mh = 1$, we record the l_2 -norm of the error in u and u_t at $t = 1$,

$$\|u(1) - u_M\|_2 \equiv \sqrt{\sum_{j=1}^{20} [u_j(1) - u_j]^2 / 20}$$

and

$$\|u_t(1) - v_M\|_2 \equiv \sqrt{\sum_{j=1}^{20} [u_j(1) - v_j]^2 / 20}.$$

We remark that both model problems have tridiagonal Jacobian; in this regard, they are reasonable models for the class of sparse Jacobian problems to which our method, requiring an evaluation of the Jacobian at an off-step point on the r.h.s. of expression (50), is most suited.

In the following tables, we report the errors and the observed rates of error reduction. These computations were performed in double precision arithmetic on an IBM 4341 at the University of Tennessee, Knoxville. In all cases, we have used the scheme whose parameters are given in Table 3, for which γ^2 is at its lower limit for unconditional stability.

First, in Tables 5 and 6, we have situations where stability is not an important factor, and we see clear indication of the anticipated fourth-order convergence in both u and u_t .

In order to introduce stability as a factor in the selection of stepsize, we have performed several experiments on Problem 1 with various choices of parameters. A representative case is shown below in Table 7, where we let $\lambda = 10,000$, $\alpha = 2$, $\rho = 3$. In this case, the spectral radius of the Jacobian at $t = 0$ is about 39787. For the conditionally-stable well-known fourth-order method of Numerov, one finds that $M \geq 79$ steps are required to reach $t = 1$ if blow up is to be avoided. On the other hand, the unconditional stability of our method is apparent from the tabulated results. Additionally, the observed convergence rates are heading toward 4 in both u and u_t , again.

We have not meant to draw any far-reaching conclusions from these experiments, but predicted accuracy and stability behavior appears to be evidenced, at least for some range of moderately stiff problems.

Table 5. Problem 1: $\lambda = 1, \alpha = \rho = 2$

M	$\ u(1) - u_M\ _2$	Rate	$\ u(1) - v_M\ _2$	Rate
5	0.362×10^{-5}		0.198×10^{-4}	
10	0.238×10^{-6}	3.92	0.123×10^{-5}	4.01
20	0.153×10^{-7}	3.97	0.766×10^{-7}	4.01
40	0.971×10^{-9}	3.98	0.478×10^{-8}	4.00

Table 6. Problem 2

M	$\ u(1) - u_M\ _2$	Rate	$\ u(1) - v_M\ _2$	Rate
5	0.463×10^{-6}		0.449×10^{-6}	
10	0.301×10^{-7}	3.94	0.301×10^{-7}	3.90
20	0.189×10^{-8}	3.99	0.193×10^{-8}	3.96
40	0.118×10^{-9}	4.00	0.122×10^{-9}	3.98

Table 7. Problem 1: $\lambda = 10,000, \alpha = 2, \rho = 3$

M	$\ u(1) - u_M\ _2$	Rate	$\ u(1) - v_M\ _2$	Rate
30	0.932×10^{-4}		0.119×10^{-2}	
40	0.241×10^{-4}	4.70	0.771×10^{-3}	1.51
50	0.845×10^{-5}	4.69	0.373×10^{-3}	3.25
60	0.379×10^{-5}	4.40	0.193×10^{-3}	3.61
70	0.199×10^{-5}	4.18	0.108×10^{-3}	3.77
80	0.116×10^{-5}	4.04	0.650×10^{-4}	3.80

REFERENCES

1. J. C. Butcher, Implicit Runge–Kutta processes. *Math. Comput.* **18**, 50–64 (1964).
2. J. C. Butcher, On the Runge–Kutta processes of higher order. *J. Aust. math. Soc.* **4**, 179–194 (1964).
3. J. C. Butcher, An algebraic theory of integration methods. *Math. Comput.* **26**, 79–106 (1972).
4. J. C. Butcher, Stability criteria for implicit Runge–Kutta methods. *SIAM JI numer. Analysis* **16**, 46–57 (1979).
5. G. A. Baker and J. H. Bramble, Semidiscrete and single step fully discrete approximations for second-order hyperbolic equations. *RAIRO Analyse Numer.* **13**, 79–100 (1979).
6. H. H. Rosenbrock, Some general implicit processes for the numerical solution of differential equations. *Comput. J.* **5**, 329–330 (1963).
7. D. A. Calahan, A stable, accurate method of numerical integration for nonlinear systems. *Proc. IEEE* **56**, 744 (1968).
8. G. Wanner, On the choice of γ for singly-implicit RK or Rosenbrock methods. *BIT* **20**, 102–106 (1980).
9. P. Kaps and P. Rentrop, Generalized Runge–Kutta methods of order four with stepsize control for stiff ordinary differential equations. *Numer. Math.* **3**, 55–68 (1979).
10. P. Kaps and G. Wanner, A study of Rosenbrock-type methods of higher order. *Numer. Math.* **38**, 279–298 (1981).
11. J. G. Verwer, An analysis of Rosenbrock methods for nonlinear stiff initial value problems. *SIAM JI numer. Analysis* **19**, 155–170 (1981).
12. F. Costabile and C. Costabile, Two-step fourth order P-stable methods for second order differential equations. *BIT* **22**, 384–386 (1982).
13. E. Hairer, Unconditionally stable methods for second order differential equations. *Numer. Math.* **32**, 373–379 (1979).
14. J. R. Cash, Efficient P-stable methods for periodic initial value problems. *BIT* **24**, 248–252 (1984).
15. M. M. Chawla, Two-step fourth order P-stable methods for second-order differential equations. *BIT* **21**, 190–193 (1981).
16. R. M. Thomas, Phase properties of high-order, almost P-stable formulae. *BIT* **24**, 225–238 (1984).
17. S. Nørsett and G. Wanner, The real-pole sandwich for rational approximations and oscillation equations. *BIT* **19**, 79–94 (1979).
18. S. Goyal, A class of Rosenbrock-type schemes for second-order nonlinear systems of ordinary differential equations. Ph.D. Dissertation, Univ. of Tennessee, Knoxville, Tenn. (1983).
19. E. Hairer and G. Wanner, Multistep–multistage–multiderivative methods for ordinary differential equations. *Computing* **11**, 287–303 (1973).
20. S. Nørsett and A. Wolfbrandt, Order conditions for Rosenbrock-type methods. *Numer. Math.* **32**, 1–15 (1979).
21. E. Hairer and G. Wanner, On the Butcher group and the general multi-value methods. *Computing*, **13**, 1–15 (1974).
22. E. Fermi, J. Pasta and S. Ulam, Studies of nonlinear problems. I. *Lect. appl. Math.* **15**, 143–156 (1974).
23. M. Toda, Waves in nonlinear lattice. *Suppl. Prog. theor. Phys.* **45**, 174–200 (1970).