LBL-16769
Preprint c.2

Lawrence Berkeley Laboratory
UNIVERSITY OF CALIFORNIA

APPLIED SCIENCE DIVISION

RECEIVED
NOV 1 6 1983
LBL LIBRARY

To be submitted for publication

A STUDY OF NUMERICAL METHODS FOR COMBUSTION KINETICS

T.D. Bui

September 1983

TWO-WEEK LOAN COPY

This is a Library Circulating Copy which may be borrowed for two weeks. For a personal retention copy, call Tech. Info. Division, Ext. 6782.

APPLIED SCIENCE DIVISION

LBL-16769 c.2

Prepared for the U.S. Department of Energy under Contract DE-AC03-76SF00098

DISCLAIMER

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor the Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or the Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or the Regents of the University of California.

A STUDY OF NUMERICAL METHODS FOR COMBUSTION KINETICS

by

T.D. Bui
Professor of Computer Science
Concordia University
Montréal, Québec, Canada

Principal Investigator: A.K. Oppenheim

Lawrence Berkeley Laboratory
University of California
Berkeley, California 94720

Abstract: Typical combustion kinetic problems can be written in the form $y' = Ap$; $y(0)$ given, where A is an M by N matrix and p is an N-vector with $p_j = k_j \prod_{i=1}^{M} y_i^{r_{ji}}$. Here the $r_{ji} > 0$ describe the reactions and the $k_j > 0$ are rate constants. This system has $f_x = 0$ and f_y is readily obtained from $\partial p_j / \partial y_i = r_{ji}\, p_j / y_i$. In this paper, we study in detail a method which takes advantage of this property of the combustion kinetic problems.

This work was supported by the Director, Office of Energy Research, Office of Basic Energy Sciences, Division of Engineering, Mathematics and Geosciences of the U.S. Department of Energy under Contract Number DE-AC03-76SF00098.

1. Introduction

The code LSTIFF[1] was developed in 1979 for the purpose of solving the mathematical modeling problem of a laser medium excited by blast waves. In the process of searching for the basic method upon which the code was to be built, we have studied many possibilities. The Rosenbrock method[9] was chosen for the code for the following reasons: first, the Rosenbrock formulas are not implicit in the sense that backward differentiation formulas (BDF)[3] such as Gear's methods are. We can therefore avoid some implementation difficulties. Second, it is possible to derive Rosenbrock formulas with strong stability properties (A- and L-stable) which are not available in high order BDF codes. Third, for the class of problems where very accurate partial derivatives can be readily obtained such as in combustion kinetic problems,[10] the method proved to be very efficient. Using the problem set proposed by Enright, et al.[2] preliminary tests of the code LSTIFF (containing L-stable methods up to order 4) show that the code is much faster than the BDF code EPISODE.[7] In this paper, we will give a complete development of the modified Rosenbrock method. Our recent work in error estimates for one-step methods will also be discussed.

2. Basic Formulas:

If a stiff system of ordinary differential equations

$$y' = f(x, y) \quad ; \quad y(x_0) = y_0$$

is integrated by a diagonally implicit Runge-Kutta method, one has to solve the following system of nonlinear equations:

$$k_i - hf(x_0 + c_i h,\ y_0 + \sum_{j=1}^{i} \beta_{ij} k_j) = 0 \ ; \quad i = 1, .., s \qquad (2.1)$$

At the i-th stage, if we use k_j for $j = 1, \ldots, i-1$; and $k_i^{(0)} = -\frac{1}{\beta_{ii}} \sum_{j=1}^{i-1} \gamma_{ij} k_j$ as the starting value for k_i , then we obtain a Newton-like iteration:

$$(I - \beta_{ii} h J) \Delta k_i = h f(x_0 + c_i h, y_0 + \sum_{j=1}^{i-1} \beta_{ij} k_j + \beta_{ii} k_i^{(0)}) - k_i^{(0)}$$

$$= h f(x_0 + c_i h, y_0 + \sum_{j=1}^{i-1} (\beta_{ij} - \gamma_{ij}) k_j) - k_i^{(0)}$$

where $\Delta k_i = k_i^{(1)} - k_i^{(0)}$ and J denotes the Jacobian $\partial f / \partial y$. Methods of this type are called semi-implicit Runge-Kutta methods. For autonomous systems, $J = f'(y_0)$ and the modified Rosenbrock method[12] becomes:

$$(I - \gamma h f'(y_0)) k_i = h f(y_0 + \sum_{j=1}^{i-1} \alpha_{ij} k_j) + h f'(y_0) \sum_{j=1}^{i-1} \gamma_{ij} k_j \tag{2.2}$$

or

$$k_i = h f(u_i) + h f'(y_0) \sum_{j=1}^{i} \gamma_{ij} k_j$$

where $\gamma = \gamma_{ii} = \beta_{ii}$, $u_i = y_0 + \sum_{j=1}^{i-1} \alpha_{ij} k_j$ and we take only one iteration of the modified Newton method. The solution y_1 at the next step is given by

$$y_1 = y_0 + \sum_{i=1}^{s} b_i k_i \tag{2.3}$$

Equations (2.2) and (2.3) form a one-step method for stiff O.D.E.'s. The constants γ, α_{ij} , γ_{ij} , b_i define the methods. In each stage, k_i is obtained by solving a system of linear equations with the same matrix. The linear combination of stages in (2.3) advances the solution y_0 at x_0

to y_1 at $x_1 = x_0 + h$.

In implementation, Eq. (2.2) involves the matrix-vector multiplications in the last term. Also, it involves the n^2 multiplications to form $\gamma h f'(y_0)$. To avoid these, we use Wanner's formulation:[11]

$$\left(\frac{I}{\gamma h} - f'(y_0)\right) k_i^* = f\left(y_0 + \sum_{j=1}^{i-1} a_{ij} k_j^*\right) + \sum_{j=1}^{i-1} c_{ij} k_j^* / h$$

$$y_1 = y_0 + \sum_{i=1}^{s} m_i k_i^* \tag{2.4}$$

where:

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

$$a_{ij} = \sum_{k=1}^{i-1} \alpha_{ik} (\delta_{kj} - \gamma c_{kj}) / \gamma \tag{2.5}$$

$$m_j = \sum_{k=1}^{s} b_k (\delta_{kj} - \gamma c_{kj}) / \gamma$$

For non-autonomous systems, we have

$$\left(I - \gamma h f_y(x_0, y_0)\right) k_i = f\left(x_0 + A_i h,\ y_0 + h \sum_{j=1}^{i-1} a_{ij} k_j\right) + B_i h f_x(x_0, y_0) + \sum_{j=1}^{i-1} c_{ij} k_j \tag{2.6}$$

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

where $$B_1 = \gamma\,, \quad B_i = \gamma + \sum_{j=1}^{i-1} c_{ij} B_j \qquad (i = 2, \dots, s)$$

$$A_i = \sum_{j=1}^{i-1} a_{ij} B_j / \gamma \qquad (i = 1, \dots, s) \tag{2.7}$$

Similar to (2.4), we can scale (2.6) and (2.7) to avoid the n^2 multiplications in $\gamma h f_y(x_0, y_0)$.

In section 3, we discuss the order conditions for the autonomous case (Eq. (2.2)). The order conditions for the non-autonomous case can be derived similarly, however, it will be more complicated.

3. Order Conditions.

For the order conditions, we have to study the power series in h of y_1 , k_i and $u_i = y_0 + \sum_{j=1}^{i-1} \alpha_{ij} k_j$ in (2.2) and (2.3). In order to do this in a systematical way, we need the following concepts.[5,6]

Let T denote the set of rooted trees. The empty tree is denoted by φ and the tree with only one node by τ . If the subtrees $t_1, \dots, t_m \in T$ are left after removing the root and the adjacent branches of a tree t , then the tree t is denoted by $[t_1, \dots, t_m]$. We define a monotonically labelled tree (LT) as follows: let n be a non-negative integer and $t : \{2, \dots, n\} \to \{1, \dots, n\}$ be a mapping which satisfies:

$$t(i) < i \quad \text{for } i = 2, \dots, n$$

and we call t a monotonically labelled tree of order $\rho(t) = n$. Examples:

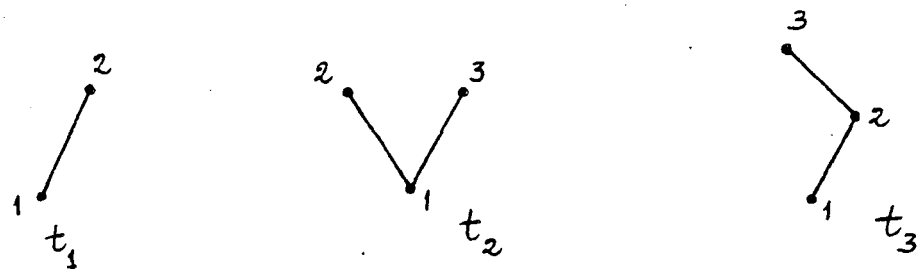


FIG. I

The node 1 is called the root. We denote the number of nodes in t and the number of possible ways of monotonic labellings of the nodes of t by $\rho(t)$ and $\alpha(t)$ respectively (see Hairer and Wanner)[5,6]. There exists a one-to-one correspondence between the set of elementary differentials and the set of monotonically labelled trees.

We define a pair of trees $t \subset w$ to be a tree w together with a subtree t having the same root. We also allow a pair $\varphi \subset w$. We denote the set of all pairs of trees by PT. The number of possible monotonic labellings of the nodes of w, such that the tree t is labelled first is denoted by $\alpha(t \subset w)$. We define the difference $d(t \subset w)$ to be the set of trees that remain after removing the subtree t.

The differential operator D operating on y

$$D \equiv \frac{d}{dy}(\ ) \cdot f(y)$$

can be interpreted by means of monotonically labelled trees as follows:

$$Dy = f$$

$$D^2 y = f' \cdot (f)$$

$$D^3 y = f'' \cdot (f,f) + f' \cdot (f' \cdot f)$$

$$D^4 y = f''' \cdot (f,f,f) + f''(f' \cdot f, f) + f''(f, f' \cdot f) + f''(f \cdot f', f) + f' \cdot (f''(f,f)) + f' \cdot (f' \cdot (f' \cdot f))$$

$f'' \cdot (f' \cdot f, f)$ means the <u>bilinear map</u> f'' operates on the vectors $f' \cdot f$ and f. We represent each f by a node, each f' by a node with one branch, and each f'' by a node with two branches, ...

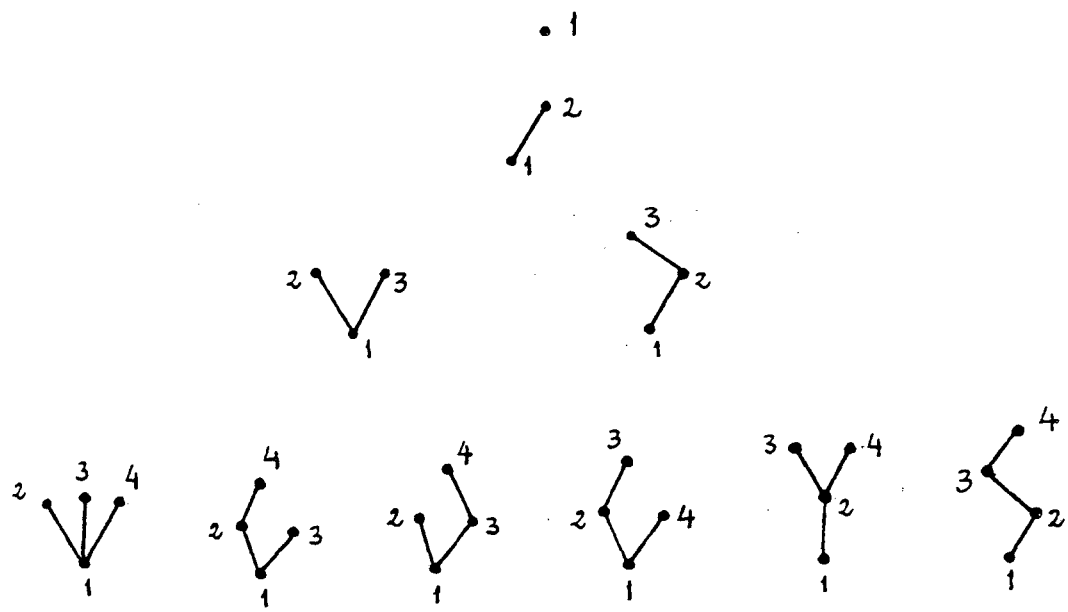


FIG. II

For every tree $t \in T$, we define a function $F(t)\, y : \mathbb{R} \to \mathbb{R}$ recursively by:

$$F(\varphi)(y) = y$$

$$F(t)(y) = f^{(m)}(y) \cdot \left(F(t_1)(y), \ldots, F(t_m)(y) \right)$$

where $t = [t_1, \ldots, t_m]$

The functions $F(t)$ are called elementary differentials. Some examples are:

$$F(\tau)(y) = f(y)$$

$$F(t_1)(y) = f'(y) \cdot f(y)$$

$$F(t_2)(y) = f''(y)\cdot(f(y), f(y))$$
$$F(t_3)(y) = f'(y)\cdot f'(y)\cdot f'(y)$$

where t_1, t_2, t_3 are defined in Figure I.

If $a: T \to \mathbb{R}$ is a mapping, the series:

$$B(a, y_0) = \sum_{t \in T} a(t)\, F(t)(y_0)\, \frac{h^{\rho(t)}}{\rho(t)!} \tag{3.1}$$

is called a Butcher series.[6]

Some examples of Butcher series are: (i) $B(o, y_0) = y_0$ with

$$o(t) = \begin{cases} 1 & \text{if } t = \varphi \\ 0 & \text{if } t \neq \varphi \end{cases} \tag{3.2}$$

(ii) $B(b, y) = hf(y)$ is a Butcher series with

$$b(t) = \begin{cases} 1 & \text{if } t = \tau \\ 0 & \text{if } t \neq \tau \end{cases} \tag{3.3}$$

(iii) $$y(x_0 + h) = \sum_{t \in LT} F(t)(y_0)\, \frac{h^{\rho(t)}}{\rho(t)!} \tag{3.4}$$

is a Butcher series $B(p, y_0)$ with $p = 1$ for all $t \in T$.

<u>Theorem 3.1</u>[5,6]

Let $a: T \to \mathbb{R}$, $b: T \to \mathbb{R}$, then the composition of the

two corresponding Butcher series is again a Butcher series

$$B(b, B(a, y_0)) = B(ab, y_0) \tag{3.5}$$

where the composition $ab: T \to \mathbb{R}$ is defined by

$$(ab)(t) = \sum_{u \subset t \in PT} b(u) \binom{r}{i} \frac{\alpha(u \subset t)}{\alpha(t)} a(u \subset t) \tag{3.6}$$

$$a(u \subset t) = \prod_{z \in d(u \subset t)} a(z) \quad , \quad r = \varrho(t) \quad , \quad i = \varrho(u)$$

$\alpha(t)$ and $\alpha(u \subset t)$ are the numbers of possibilities of monotonic labelling the nodes of t and $u \subset t$ respectively.

<u>Theorem 3.2</u>

The functions u_i, k_i and y_1 of Eqs. (2.2) and (2.3) are Butcher series defined by:

$$u_i(h) = B(\underline{u}_i, y_0)$$

$$k_i(h) = B(\underline{k}_i, y_0)$$

$$y_1(h) = B(\underline{y}_1, y_0)$$

where the coefficients $u_i(t)$, $k_i(t)$ and $y_1(t)$ for $t \in T$ can be expressed recursively by:

$$\underline{u}_i(\varphi) = 1 \quad , \quad \underline{u}_i(t) = \sum_{j=1}^{i-1} \alpha_{ij} k_j \tag{3.7}$$

$$\underline{k}_i(\varphi) = 0 \quad , \quad \underline{k}_i(\tau) = 1 \quad , \tag{3.8}$$

$$\underline{k}_i(t) = \rho(t)\,\underline{u}_i(t_1)\cdots\underline{u}_i(t_m) + \begin{cases} 0 & \text{if } t=[t_1,\dots,t_m] \\ \rho(t)\sum \gamma_{ij}\,\underline{k}_j(t_1) & \text{if } t=[t_1] \end{cases} \tag{3.9}$$

$$\underline{y}_1(\varphi) = 1 \;, \qquad \underline{y}_1(t) = \sum_{i=1}^{s} b_i\,\underline{k}_i(t) \tag{3.10}$$

<u>Theorem 3.3</u>

For every tree $t \in T$ and $t \neq \varphi$, the coefficients of the Butcher series of y_1 are given by

$$\underline{y}_1(t) = \Gamma(t)\sum_{i=1}^{s} b_i\,\psi_i(t) \tag{3.11}$$

where $\Gamma(\tau)=1$, $\Gamma(t) = \rho(t)\Gamma(t_1)\cdots\Gamma(t_m)$ for $t=[t_1,\dots,t_m]$

and $\psi_i(\tau) = 1$,

$$\psi_i(t) = \begin{cases} \sum_{j_1\cdots j_m} \alpha_{ij_1}\cdots\alpha_{ij_m}\,\psi_{j_1}(t_1)\cdots\psi_{j_m}(t_m) & \text{for } t=[t_1,\dots,t_m] \\ \sum_j \beta'_{ij}\,\psi_j(t_1) & \text{for } t=[t_1] \end{cases} \tag{3.12}$$

where we define: $\beta'_{ij} = \alpha_{ij} + \gamma_{ij}$ and $\beta_{ij} = \begin{cases} 0 & \text{for } i \leq j \\ \beta'_{ij} & \text{for } i > j \end{cases}$

Because of the recursive properties of $\psi_i(t)$, they can be obtained by the following procedure:

Attach to each node of t a summation letter starting with i at the

root, then ψ_i is equal to the sum over the product containing

β'_{jk} whenever a single-branched node j is directly connected with an upper node k

α_{jk} whenever a multiple-branched node j is directly connected with an upper node k .

<u>Corollary:</u> Method (2.2) and (2.3) is of order p if

$$\Gamma(t) \sum_i b_i \psi_i(t) = 1 \tag{3.13}$$

for $\rho(t) \leqslant p$ and this equation is not satisfied for at least one tree of order $p+1$.

The principal truncation error $e(t)$ is given by:

$$e(t) = \sum_{\substack{t \in LT \\ \rho(t) = p+1}} \underline{e}(t) F(t)(y_0) \frac{h^{p+1}}{(p+1)!} \tag{3.14}$$

where $\underline{e}(t) = 1 - \underline{y}_1(t)$ is the error coefficients.

Theorem (3.3) and the corollary give us a scheme to write all order conditions for the method defined by Eqs. (2.2) and (2.3). However, these order conditions can be simplified further as follows:

If we move all terms containing $\gamma = \gamma_{ii} = \beta'_{ii}$ to the right hand side. This gives a polynomial $P_k(\gamma)$ where k denoted the tree number. Also, we define $\Phi_i(t)$ exactly as $\psi_i(t)$ in Theorem (3.3) except that β'_{ij} is replaced by β_{ij} . Furthermore, for $t \in T$ and a positive integer j , we define:

$$V(t,j) = \{ s \in T \text{ , } s \text{ is obtained from } t \text{ by removing } j \text{ single branched nodes} \}$$

If t has less than j single branched nodes, then $V(t,j) = 0$.

We also define:

$$N(t,s) = \text{number of possibilities to obtain } s \text{ by removing } j \text{ single branched nodes from } t \text{ .}$$

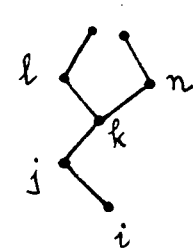
Theorem 3.4

If the order conditions in the Corollary are satisfied for all trees of order $\leqslant \rho(t) - 1$, then the order conditions can be replaced by:

$$\sum_i b_i \Phi_i(t) = \sum_{j \geqslant 0} (-\gamma)^j \sum_{s \in V(t,j)} N(t,s) / \Gamma(s) \tag{3.15}$$

Example:

We will use the result of Theorem (3.4) to derive the order condition for the tree



$$\sum_i b_i \Phi_i(t) = \sum_{ijkln} b_i \beta_{ij} \beta_{jk} \alpha_{kl} \alpha_{kn} \beta_l \beta_n$$

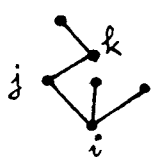
To find the right hand side, we set up the following table:

Single-branched node removed j	1	1	2	2	2	3	3	4
Tree s								
Number of possibilities $N(t,s)$	2	2	1	1	4	2	2	1
$T(s)$	6.5.2.2	6.5.4.2	5.4.3	5.2.2	5.4.2	4.3	4.2	3

Therefore, the order condition is:

$$\sum_{ijk\ell n} b_i \beta_{ij} \beta_{jk} \alpha_{k\ell} \alpha_{\ell n} \beta_\ell \beta_n = \frac{1}{7.6.5.2.2} - \gamma\left(\frac{2}{6.5.2.2} + \frac{2}{6.5.4.2}\right) + \gamma^2\left(\frac{1}{5.4.3} + \frac{1}{5.2.2} + \frac{4}{5.4.2}\right)$$
$$- \gamma^3\left(\frac{2}{4.3} + \frac{2}{4.2}\right) + \gamma^4\left(\frac{1}{3}\right)$$

<u>Examples:</u>

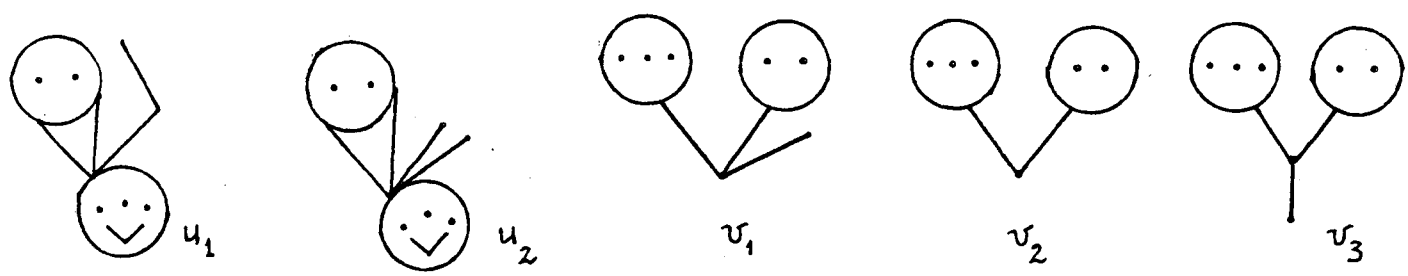
Tree t	$\rho(t)$	Order conditions
	1	$\sum_i b_i = 1$
	2	$\sum_i b_i \beta_i = \frac{1}{2} - \gamma$
	4	$\sum_{ij} b_i \beta_{ij} \alpha_j^2 = \frac{1}{12} - \frac{\gamma}{3}$
	6	$\sum_{ijk} b_i \alpha_{ij} \alpha_i^2 \beta_{jk} \beta_k = \frac{1}{36} - \frac{\gamma}{5} + \frac{\gamma^2}{4}$

4. Simplifications of Order Conditions.[11]

Many order conditions listed in the Table can be eliminated by using the following propositions.

Proposition 4.1:

Let u_1, u_2, v_1, v_2, v_3 be trees as sketched below:



where the encircled parts are identical. Then:

(a) The condition:

$$\sum_k \alpha_{ik} \beta'_k = \frac{\alpha_i^2}{2} \qquad \text{or} \qquad \sum_k \alpha_{ik} \beta_k = \alpha_i \left(\frac{\alpha_i}{2} - \gamma \right) \quad ; \; i = 2, \ldots, s$$

implies that $\underline{y}_1(u_1) = \underline{y}_1(u_2)$.

(b) The condition:

$$\sum_i b_i \beta'_{ik} = b_k (1 - \alpha_k)$$

$$\text{or} \quad \sum_i b_i \beta_{ik} = b_k (1 - \alpha_k - \gamma) \quad ; \quad k = 1, \ldots, s$$

implies that $\underline{y}_1(v_1) = 1$ and $\underline{y}_1(v_2) = 1$ then $\underline{y}_1(v_3) = 1$.

Proof: by direct substitution, note that for the simplifying assumptions, it is easier to work with β'_{ij} instead of β_{ij} .

Example:

$u_1 =$

$u_2 =$

$$\underline{y}_1(u_1) = 6.2 \sum b_i \alpha_{ij} \beta_j' \alpha_i^3$$

$$\underline{y}_1(u_2) = 6.2.2 \sum b_i \alpha_{ij} \beta_j' \alpha_{ik} \beta_k' \alpha_i$$

$$= 6.2 \sum b_i \alpha_{ij} \beta_j' \alpha_i^3 = \underline{y}_1(u_1)$$

In Table 1, trees 5 and 6 are "equivalent" if condition (a) is satisfied. Trees 9, 10 and 11 are all equivalent. When a group of trees (of the same order) satisfies a simplifying condition (a or b), then this group of trees can be eliminated except one together with the simplifying condition. This approach will reduce the number of nonlinear equations to be solved. For example, for methods of order 5, there are 17 order conditions. However, trees 10 and 11 are eliminated (they are equivalent to tree 9) and tree 15 is eliminated due to tree 14. Therefore, 3 equations are eliminated and in their places, we have to add the simplifying condition (a) for s = 5. Note that the simplifying condition (a) for trees 5 and 6 is different from that for trees 9, 10, 11, 14 and 15. Therefore, it is not advantageous to eliminate tree 6.

5. Step Size and Error Controls:

For the implementation, an error estimate is necessary so that the

step size can be automatically controlled. Basically, there are two approaches for error estimates:

(a) use a pair of embedded methods, or

(b) use the Richardson extrapolation approach

Pairs of embedded formulas:

A pair of embedded formulas consists of two formulas which for the sake of efficiency use as many stages as possible in common. The result of one formula - say of order p - is used as the starting value for the next step and the result of the other - say of order q - as error estimate. We denote such a pair by $p(q)$. The method with order q written inside brackets is not used for step continuation.

Richardson extrapolation:

We use one basic method of order p to compute y_1 - the solution of one step with step-size $2h$ and y_2 - the solution of two steps with step-size h. By Richardson extrapolation we obtain a result of order $p+1$:

$$y_{ex} = y_2 + \frac{y_2 - y_1}{2^p - 1} \tag{5.1}$$

6. Stability properties:

Numerical methods for stiff systems should - at least for a certain class of differential equations - yield solutions which show a similar stability behaviour as the exact solutions. One such class is the contractive differential equations:

$$y' = f(y) \tag{6.1}$$

which are characterized by the contractive condition

$$\langle y-z\,,\ f(y)-f(z)\rangle \leqslant 0 \qquad \forall\ y,z \in \mathbb{R}^n \tag{6.2}$$

For any two solutions $y(x)$, $z(x)$ of the differential equation, we have:

$$\|y(x_0+h) - z(x_0+h)\| \leqslant \|y(x_0) - z(x_0)\| \tag{6.3}$$

Eq. (6.2) implies that:

$$\frac{d}{dx}\|y(x) - z(x)\|^2 = 2\langle y(x)-z(x)\,,\ f(y(x)) - f(z(x))\rangle \leqslant 0$$

Therefore, we would like to have $\|y_1 - z_1\| \leqslant \|y_0 - z_0\|$ for the numerical solution. There are implicit Runge-Kutta methods where this can be obtained for all h . These methods however, require the solution of an implicit system of equations. This is usually done in one or two iterations. Then the step-size h is restricted by convergence. Therefore, in the case of semi-implicit methods one cannot expect contractivity for arbitrary large step-size.

We now study the stability of Rosenbrock methods for the scalar test equation:

$$y' = \lambda y \quad , \quad \mathrm{Re}(\lambda) \leqslant 0 \tag{6.4}$$

We define the following:

$$z = \lambda h \quad , \quad \lambda \in \mathbb{C}$$

$$w = \frac{z}{1-\gamma z}$$

$$A = (\alpha_{ij})$$

$$B = (\beta_{ij}) \quad \text{with} \quad \beta_{ij} = \begin{cases} \alpha_{ij} + \gamma_{ij} & 1 \leqslant j < i \leqslant s \\ 0 & \text{otherwise} \end{cases}$$

$$\beta_i = \sum_{j=1}^{s} \beta_{ij} \quad , \quad \alpha_i = \sum_{j=1}^{s} \alpha_{ij}$$

Also, we use the following notations:

$$\underline{u} = (u_1, \dots, u_s)^T \quad , \quad \underline{k} = (k_1, \dots, k_s)^T$$

$$\underline{c} = (c_1, \dots, c_s)^T \quad , \quad \mathbb{1} = (1, \dots, 1)^T$$

Integration of (6.4) by the modified Rosenbrock method (2.2), yields:

$$\underline{u} = y_0 \mathbb{1} + A\underline{k}$$

$$y_1 = y_0 + \underline{b}^T \underline{k}$$

$$\underline{k} = w(y_0 \mathbb{1} + B\underline{k})$$

Because $B = \begin{pmatrix} 0 & 0 & \dots & 0 \\ b_{21} & 0 & \dots & 0 \\ \dots & & & \\ b_{s1} & b_{s2} & \dots & 0 \end{pmatrix}$ hence $B^s = 0$

<u>Proposition 6.1</u>:

The numerical solution of (6.4) obtained by modified Rosenbrock

method (2.2) is given by:

$$y_1 = R(z)\, y_0$$
$$R(z) = 1 + \sum_{j=1}^{s} b^T . B^{j-1} . \mathbb{1}\, w^j \tag{6.5}$$

For the internal stages, we have:

$$u_i = R_i(z) y_0$$
$$R_i(z) = 1 + \sum_{j=1}^{i-1} (\alpha_{i1}, \ldots, \alpha_{i\,i-1}, 0, \ldots, 0) B^{j-1} \mathbb{1}\, w^j \tag{6.6}$$

The $R_i(z)$ are called <u>internal stability functions</u> and $R(z)$ is called the <u>stability function</u>.

The first few internal stability functions are (see 6.6):

$$R_1(z) = 1$$
$$R_2(z) = 1 + \alpha_{21} w = 1 + \alpha_2 w$$
$$R_3(z) = 1 + \alpha_3 w + \alpha_{32} \beta_2 w^2$$
$$R_4(z) = 1 + \alpha_4 w + (\alpha_{42} \beta_2 + \alpha_{43} \beta_3) w^2 + \alpha_{43} \beta_{32} \beta_2 w^3$$

These are obtained easily from equation (6.6). The coefficients of w^j are the left hand sides of the simplifying conditions. If the simplifying conditions are satisfied, we have:

$$R_i(z) = e^{\alpha_i z} + O(z^i)$$

Laguerre polynomials are defined by:

$$L_n^{\alpha}(x) = \sum_{k=0}^{n} (-1)^k \binom{n+\alpha}{n-k} \frac{x^k}{k!} \tag{6.7}$$

The first few Laguerre pólynomials are:

$$L_0^{\alpha}(x) = 1$$

$$L_1^{\alpha}(x) = -x + \alpha + 1$$

$$L_2^{\alpha}(x) = \frac{x^2}{2} - (\alpha+2)x + \frac{1}{2}(\alpha+1)(\alpha+2)$$

From the relation for the generating function for Laguerre polynomials

$$(1-v)^{\alpha+1} \sum_{n=0}^{\infty} L_n^{\alpha}(x)\, v^n = \exp\left(-\frac{xv}{1-v}\right) \; ; \alpha > -1, |v| < 1 \tag{6.8}$$

and the recurrence relation:

$$-\frac{x}{n+1} L_n^1(x) = L_{n+1}^1(x) - 2L_n^1(x) + L_{n-1}^1(x) \tag{6.9}$$

we obtain:

$$1 - x \sum_{n=0}^{\infty} \frac{1}{n+1} L_n^1(x)\, v^{n+1} = \exp\left(-\frac{xv}{1-v}\right) \tag{6.10}$$

By putting $x = \frac{1}{\gamma}$ and $z = \frac{-xv}{1-v}$, i.e. $v = \frac{-\gamma z}{1-\gamma z} = -\gamma w$

$$1 - \frac{1}{\gamma} \sum_{n=0}^{\infty} \frac{1}{n+1} L_n^1\left(\frac{1}{\gamma}\right)(-\gamma w)^{n+1} = \exp(z) \tag{6.11}$$

<u>Proposition 6.2</u>:

(1) The stability function for method (2.2) with s stages and

order s is given by:

$$R(z) = 1 - \frac{1}{\gamma} \sum_{n=0}^{s-1} \frac{1}{n+1} L_n^1\left(\frac{1}{\gamma}\right)(-\gamma w)^{n+1} \tag{6.12}$$

(2) If $\frac{1}{\gamma}$ is a zero of $L_s^1(x)$ (i.e., $L_s^1\left(\frac{1}{\gamma}\right) = 0$), then:

$$R(z) = e^z + O(z^{s+2})$$

(3) $R(z)$ can be rewritten as:

$$R(z) = \frac{P(z)}{Q(z)} = \frac{\sum_{n=0}^{s} L_n^{s-n}\left(\frac{1}{\gamma}\right)(-\gamma z)^n}{(1-\gamma z)^s} \tag{6.13}$$

(4) For the stability at infinity it holds:

$$R(\infty) = L_s\left(\frac{1}{\gamma}\right) \tag{6.14}$$

A comparison of (6.5) and (6.12) gives:

$$\underline{b}^T B^n \mathbb{1} = \frac{1}{\gamma(n+1)} L_n^1\left(\frac{1}{\gamma}\right)(-\gamma)^{n+1}$$

This relation gives the order conditions for the single-branched trees of Table 1.

By putting $x = \frac{a}{\gamma}$ and $z = -\frac{v}{\gamma(1-v)}$, we obtain from (6.10):

$$\exp(az) = 1 - \frac{a}{\gamma} \sum_{n=0}^{\infty} \frac{1}{n+1} L_n^1\left(\frac{a}{\gamma}\right)(-\gamma w)^{n+1} \tag{6.15}$$

Where as before $w = \frac{z}{1-\gamma z}$.

A comparison of (6.6) and (6.15) gives

$$R_i(z) = \exp(\alpha_i z) + O(z^i)$$

if

$$(\alpha_{i1}, \dots, \alpha_{ii-1}, 0, \dots, 0)\, B^n \mathbb{1} = -\frac{a}{\gamma} \frac{1}{n+1} L_n^1\left(\frac{a}{\gamma}\right)(-\gamma)^{n+1}$$

for $0 \leqslant n \leqslant i-2$.

For $n=0$ this condition gives $\alpha_i = a$. For $n \geqslant 1$, these relations are simplifying conditions.

Comparing the first term of (6.12) and (6.15) we have

<u>Proposition 6.3</u>:

Let $R(z,\gamma)$ denote the right hand side of (6.12) in order to indicate the dependence on γ. Let

$$\bar{R}(z,a,\gamma) = 1 - \frac{a}{\gamma} \sum_{n=0}^{s-1} \frac{1}{n+1} L_n^1\left(\frac{a}{\gamma}\right)(-\gamma w)^{n+1}$$

Then:

$$\bar{R}(z,a,\gamma) = \begin{cases} 1 & a = 0 \\ R\left(az, \frac{\gamma}{a}\right) & a \neq 0 \end{cases}$$

The internal stability functions are usually low order approximations to $e^{\alpha_i z}$. Thus, in the following examples, we investigate the stability functions of lower order than (6.12). To this aim, we add on to the right hand side of (6.12) a term $\sum_{j=1}^{s} d_j w^j$.

Note that proposition (6.3) remains true if we replace

$$R(z,\gamma) \text{ by } R(z,\gamma) + \sum_{j=1}^{s} d_j w^j \qquad \text{and}$$

$$\overline{R}(z,a,\gamma) \text{ by } \overline{R}(z,a,\gamma) + \sum_{j=1}^{s} d_j a_j w^j \qquad \text{simultaneously.}$$

Let $R(z) = \dfrac{P(z)}{Q(z)}$ be a stability function of order p, $\gamma > 0$, $Q(z) = (1-\gamma z)^s$, $P(z)$ a polynomial of degree $\leqslant s$. By the maximum principle, $R(z)$ is A-stable iff

$$|R(iy)|^2 \leqslant 1 \qquad \forall\, y \in \mathbb{R}$$

This is equivalent to:

$$E(y) = |Q(iy)|^2 - |P(iy)|^2 \geqslant 0 \qquad \forall\, y \in \mathbb{R}$$

$E(y)$ is called E-polynomial and has the form

$$E(y) = \sum_{j=0}^{2s} e_{2j}\, y^{2j} \qquad \text{with} \quad e_k = 0 \;;\; k \leqslant p .$$

In the following, we study the stability functions of order $s-1$.

<u>Example</u>:

$s = 1$: $R(z) = 1 + w$ (Note: for $s = 1$, $b_1 = 1$ because of order condition 1)

This method is A-stable if $\gamma \geqslant \frac{1}{2}$

$s = 2$: $R(z) = 1 + w + \left(\frac{1}{2} - \gamma + d\right) w^2$

$$e_2 = 8d$$

$$e_4 = (-2d + 4\gamma - 1)(2d + 4\gamma^2 - 4\gamma + 1)$$

$R(z)$ is A-stable if $e_2, e_4 \geqslant 0$, i.e., $0 \leqslant d \leqslant 2\gamma - \frac{1}{2}$

For $d = -\gamma^2 + 2\gamma - \frac{1}{2}$, $P(z)$ is a polynomial of degree 1 . $R(z)$ is L-stable if $|\gamma - 1| \leqslant \frac{\sqrt{2}}{2}$.

$s = 3$: $R(z) = 1 + w + \left(\frac{1}{2} - \gamma\right) w^2 + \left(\frac{1}{6} - \gamma + \gamma^2 + d\right) w^3$

$$e_4 = 3(1 - 12\gamma + 36\gamma^2 - 24\gamma^3 + 24d(1 - 3\gamma))$$

$$e_6 = -(6d + p)^2 + 12\gamma^3(6d + p)$$

$R(z)$ is A-stable for values of (γ, d) satisfying $e_4 \geqslant 0$ and $e_6 \geqslant 0$ ($e_6 \geqslant 0$ is equivalent to $0 \leqslant 6d + p \leqslant 12\gamma^3$). For $d = -\gamma^3 + 3\gamma^2 - \frac{3}{2}\gamma + \frac{1}{6}$, $P(z)$ is a polynomial of degree 2 . $R(z)$ is L-stable if: $0.375603 \leqslant \gamma \leqslant 0.550386$

7. Stability of the extrapolated scheme in Richardson extrapolation.

Let $R(z)$ be a stability function of order p . The stability function of the extrapolated value is:

$$R_{ex}(2z) = \frac{2^p R^2(z) - R(2z)}{2^p - 1} \tag{7.1}$$

Proposition 7.1:

Let $R_{ex}(z)$ be as defined in (7.1). Then $|R_{ex}(\infty)| \leqslant 1$ is equivalent to $-\frac{2^p - 1}{2^p} \leqslant R(\infty) \leqslant 1$.

Proof: Since $R(\infty)$ is real, $-1 \leqslant R_{ex}(\infty)$ holds for any value of $R(\infty)$. The condition $R_{ex}(\infty) \leqslant 1$ gives:

$$-\frac{2^p - 1}{2^p} \leqslant R(\infty) \leqslant 1$$

Example: $R(z)$ is given by (6.12) or (6.13) for $s=4$. From proposition 6.2 (see (1) and (2)), the order of $R(z)$ is $p = 4$ if $L_4^1\left(\frac{1}{\gamma}\right) \neq 0$. $R_{ex}(z)$ is stable at infinity for $-\frac{15}{16} \leqslant R(\infty) \leqslant 1$. From (6.14) it follows:

$$R(\infty) = L_4\left(\frac{1}{\gamma}\right) = \left(\gamma^4 - 4\gamma^3 + 3\gamma^2 - \frac{2}{3}\gamma + \frac{1}{24}\right) / \gamma^4$$

Therefore, R_{ex} is stable at infinity when

$$-\frac{15}{16} \leqslant \frac{\gamma^4 - 4\gamma^3 + 3\gamma^2 - \frac{2}{3}\gamma + \frac{1}{24}}{\gamma^4} \leqslant 1$$

or when $\gamma \in I_1 \cup I_2 \cup I_3$ where:

$$I_1 = [0.105\,663\ ,\ 0.107\,227]$$

$$I_2 = [0.204\,711\ ,\ 0.25]$$

$$I_3 = [0.394\,338\ ,\ \infty]$$

8. Conclusions.

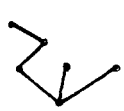
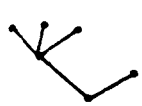
In this paper, we present in details all mathematical tools and their justifications for developing new algorithms for the numerical solutions of stiff differential equations. We briefly summarize the main results here. In section 1, we present the justifications for our belief that the modified Rosenbrock method is suitable for problems in which the partial derivatives

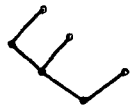
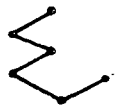
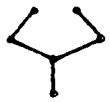
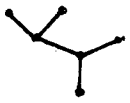
are readily obtainable. In particular, we have in mind the chemical kenetic problems. In sections 2,3 and 4 we present a very powerful method due to Hairer for developing order conditions. Any numerical analyst who wishes to design his own modified Rosenbrock method will find the mathematical tools presented here extremely useful. Methods for estimating the truncation errors are discussed in section 5 and the stability properties are discussed in sections 6 and 7.

We have not developed any particular algorithm in this paper due to two reasons: First, there exist in the literature some very good Rosenbrock algorithms of orders up to four with strong stability properties (see code LSTIFF by Bui[1], GRK4A by Kaps and Rentrop[8], and ROW4A by Gottwald and Wanner[4].) Second, the development of any new algorithm must be presented with a detailed testing. We feel that this paper already contains enough information, a second paper in this sequence is now being prepared.

Table 1. List of order conditions up to order 6.

number	tree	order	
1		1	$\sum b_i = P_1$
2		2	$\sum b_i \beta_i = P_2$
3		3	$\sum b_i \alpha_i^2 = P_3$
4		3	$\sum b_i \beta_{ij} \beta_j = P_4$
5		4	$\sum b_i \alpha_i^3 = P_5$
6		4	$\sum b_i \alpha_i \alpha_{ik} \beta_k = P_6$
7		4	$\sum b_i \beta_{ik} \alpha_k^2 = P_7$
8		4	$\sum b_i \beta_{ik} \beta_{k\ell} \beta_\ell = P_8$
9		5	$\sum b_i \alpha_i^4 = P_9$
10		5	$\sum b_i \alpha_i^2 \alpha_{ik} \beta_k = P_{10}$
11		5	$\sum b_i \alpha_{ik} \beta_k \alpha_{i\ell} \beta_\ell = P_{11}$

12		5	$\sum b_i \alpha_i \alpha_{ik} \alpha_k^2 = P_{12}$
13		5	$\sum b_i \alpha_i \alpha_{ik} \beta_{k\ell} \beta_\ell = P_{13}$
14		5	$\sum b_i \beta_{ik} \alpha_k^3 = P_{14}$
15		5	$\sum b_i \beta_{ik} \alpha_k \alpha_{k\ell} \beta_\ell = P_{15}$
16		5	$\sum b_i \beta_{ik} \beta_{k\ell} \alpha_\ell^2 = P_{16}$
17		5	$\sum b_i \beta_{ik} \beta_{k\ell} \beta_{\ell m} \beta_m = P_{17}$
18		6	$\sum b_i \alpha_i^5 = P_{18}$
19		6	$\sum b_i \alpha_i^3 \alpha_{ik} \beta_k = P_{19}$
20		6	$\sum b_i \alpha_i \alpha_{ik} \beta_k \alpha_{i\ell} \beta_\ell = P_{20}$
21		6	$\sum b_i \alpha_i^2 \alpha_{ik} \alpha_k^2 = P_{21}$
22		6	$\sum b_i \alpha_i^2 \alpha_{ik} \beta_{k\ell} \beta_\ell = P_{22}$
23		6	$\sum b_i \alpha_{ik} \beta_k \alpha_{i\ell} \alpha_\ell^2 = P_{23}$
24		6	$\sum b_i \alpha_{ik} \beta_k \alpha_{i\ell} \beta_{\ell m} \beta_m = P_{24}$
25		6	$\sum b_i \alpha_i \alpha_{ik} \alpha_k^3 = P_{25}$

26		6	$\sum b_i \alpha_i \alpha_{ik} \alpha_k \alpha_{k\ell} \beta_\ell = P_{26}$
27		6	$\sum b_i \alpha_i \alpha_{ik} \beta_{k\ell} \alpha_\ell^2 = P_{27}$
28		6	$\sum b_i \alpha_i \alpha_{ik} \beta_{k\ell} \beta_{\ell m} \beta_m = P_{28}$
29		6	$\sum b_i \beta_{ik} \alpha_k^4 = P_{29}$
30		6	$\sum b_i \beta_{ik} \alpha_k^2 \alpha_{k\ell} \beta_\ell = P_{30}$
31		6	$\sum b_i \beta_{ik} \alpha_{k\ell} \beta_\ell \alpha_{km} \beta_m = P_{31}$
32		6	$\sum b_i \beta_{ik} \alpha_k \alpha_{k\ell} \alpha_\ell^2 = P_{32}$
33		6	$\sum b_i \beta_{ik} \alpha_k \alpha_{k\ell} \beta_{\ell m} \beta_m = P_{33}$
34		6	$\sum b_i \beta_{ik} \beta_{k\ell} \alpha_\ell^3 = P_{34}$
35		6	$\sum b_i \beta_{ik} \beta_{k\ell} \alpha_\ell \alpha_{\ell m} \beta_m = P_{35}$
36		6	$\sum b_i \beta_{ik} \beta_{k\ell} \beta_{\ell m} \alpha_m^2 = P_{36}$
37		6	$\sum b_i \beta_{ik} \beta_{k\ell} \beta_{\ell m} \beta_{mn} \beta_n = P_{37}$

$$P_1 = 1$$

$$P_2 = \frac{1}{2} - \gamma$$

$$P_3 = \frac{1}{3}$$

$$P_4 = \frac{1}{6} - \gamma + \gamma^2$$

$$P_5 = \frac{1}{4}$$

$$P_6 = \frac{1}{8} - \frac{\gamma}{3}$$

$$P_7 = \frac{1}{12} - \frac{\gamma}{3}$$

$$P_8 = \frac{1}{24} - \frac{\gamma}{2} + \frac{3\gamma^2}{2} - \gamma^3$$

$$P_9 = \frac{1}{5}$$

$$P_{10} = \frac{1}{10} - \frac{\gamma}{4}$$

$$P_{11} = \frac{1}{20} - \frac{\gamma}{4} + \frac{\gamma^2}{3}$$

$$P_{12} = \frac{1}{15}$$

$$P_{13} = \frac{1}{30} - \frac{\gamma}{4} + \frac{\gamma^2}{3}$$

$$P_{14} = \frac{1}{20} - \frac{\gamma}{4}$$

$$P_{15} = \frac{1}{40} - \frac{5}{24} + \frac{\gamma^2}{3}$$

$$P_{16} = \frac{1}{60} - \frac{\gamma}{6} + \frac{\gamma^2}{3}$$

$$P_{17} = \frac{1}{120} - \frac{\gamma}{6} + \gamma^2 - 2\gamma^3 + \gamma^4$$

$$P_{18} = \frac{1}{6}$$

$$P_{21} = \frac{1}{18}$$

$$P_{22} = \frac{1}{36} - \frac{\gamma}{5} + \frac{\gamma^2}{4}$$

$$P_{25} = \frac{1}{24}$$

$$P_{27} = \frac{1}{72} - \frac{\gamma}{15}$$

$$P_{28} = \frac{1}{144} - \frac{\gamma}{10} + \frac{3\gamma^2}{8} - \frac{\gamma^3}{3}$$

$$P_{34} = \frac{1}{120} - \frac{\gamma}{10} + \frac{\gamma^2}{4}$$

$$P_{36} = \frac{1}{360} - \frac{\gamma}{20} + \frac{\gamma^2}{4} - \frac{\gamma^3}{3}$$

$$P_{37} = \frac{1}{720} - \frac{\gamma}{24} + \frac{5\gamma^2}{12} - \frac{5\gamma^3}{3} + \frac{5\gamma^4}{2} - \gamma^5$$

References

1. Bui, T.D. and Bui, T.R, Numerical methods for extremely stiff systems of ordinary differential equations, Appl. Math. Modeling 3, 355, (1979).

2. Enright, W.H., Hull, T.E. and Lindberg, B., Comparing numerical methods for stiff systems of ODEs, BIT 15, 10 (1975).

3. Gear, C.W., Numerical initial value problems in ordinary differential equations, Prentice-Hall, Englewood Cliffs, New Jersey (1971).

4. Gottwald, B.A. and Wanner, G., A reliable Rosenbrock integrator for stiff differential equations, Computing 26, 355 (1981).

5. Hairer, E. and Wanner, G., Multistep-multistage-multiderivative methods for ordinary differential equations, Computing 11, 287 (1973).

6. Hairer, E. and Wanner, G., On the Butcher group and general multi-value methods, Computing 13, 1 (1974).

7. Hindmarsh, A.C. and Byrne, G.D., Episode: an effective package for the integration of systems of ordinary differential equations, UCID-30112, Lawrence Livermore Lab. (1977).

8. Kaps, P. and Rentrop, P., Generalized Runge-Kutta methods of order four with stepsize control for stiff ordinary differential equations, Numer. Math. 33, 55 (1979).

9. Rosenbrock, H.H., Some general implicit processes for the numerical solution of differential equations, Computer J. 5, 329 (1963).

10. Shampine, L.F., Implementation of Rosenbrock methods, ACM TOMS 8, 93 (1982).

11. Wanner, G. and Kaps, P.,A study of Rosenbrock-type methods of high order, to be published in Num. Math.

12. Wolfbrandt, A., A study of Rosenbrock processes with respect to order conditions and stiff stability, Thesis, Chalmers University of Technology, Sweden (1977).

This report was done with support from the Department of Energy. Any conclusions or opinions expressed in this report represent solely those of the author(s) and not necessarily those of The Regents of the University of California, the Lawrence Berkeley Laboratory or the Department of Energy.

Reference to a company or product name does not imply approval or recommendation of the product by the University of California or the U.S. Department of Energy to the exclusion of others that may be suitable.

TECHNICAL INFORMATION DEPARTMENT
LAWRENCE BERKELEY LABORATORY
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720