JOURNAL OF COMPUTATIONAL PHYSICS 39, 251–271 (1981)

Efficient Integration of the Time Dependent Collisional-Radiative Equations

G. J. PERT

Department of Applied Physics, University of Hull, Hull, HU6 7RX United Kingdom

Received April 5, 1979; revised January 30, 1980

The properties of a set of conservative positive rate equations, such as the collisional-radiative equations, are investigated. It is shown that the equations are positivity maintaining, and furthermore, converge uniformly onto a unique equilibrium state consistent with physical constraints. Two methods of integration are considered. The first involving eigenvector decomposition yields an exact solution, but is expensive in computer time, and cumbersome if the rate matrix is defective. A simple weighted finite difference approach may be integrated stiffly and unconditionally stably, and is therefore suitable for inclusion in multi-celled fluid codes. The improvement in accuracy obtained by weighting is investigated, and attention is drawn to the importance of a positivity-maintaining form.

INTRODUCTION

Interest in the feasibility of generating population inversions in highly ionised species has led to the need to study ionisation processes in developing plasmas. To this end, simple algorithms were described in an earlier paper [1] to allow the inclusion of time-dependent ionisation and recombination within a multi-celled fluid code. In this paper we further develop this work by considering algorithms for the more general collisional-radiative equations.

The collisional-radiative equations [2] form a general group of rate equations with transitions between arbitrarily designated levels. Although their physical properties, for example, conservatism, equilibration, and positivism, are well known [2, 3], these have not been generally proven. We therefore, in this paper, initially establish these properties by considering the behaviour of the equations with a set of arbitrary rates. These results are used to define classes of numerical operators, with properties which mimic the exact equations.

The differential equations associated with the time-dependent collisional-radiative equations are by their nature stiff [4]. If n levels of the system are considered, the rate equations involve the calculation of $n(n-1)$ different rate coefficients dependent in a complicated fashion on the state variables (principally temperature) of the gas. These should clearly be calculated as infrequently as possible, i.e., on the characteristic time scale of the hydrodynamics of the system. However, in general, the characteristic time scale associated with the rates may be much shorter than that of

0021-9991/81/020251-21$02.00/0

Copyright © 1981 by Academic Press, Inc.
All rights of reproduction in any form reserved.

the hydrodynamics. The algorithm should therefore be capable of stable operation in a fully stiff mode. Accurate multi-step methods for treating this problem are available [4], but these involve the storage of data from several time steps, typically $8n$ elements for each fluid-cell. As this rapidly becomes prohibitively large, we have concentrated on two-step schemes, which have the advantage of A-stability [5]. Two methods are considered: an exact method involving the eigenvector decomposition and a weighted finite difference approach. Since the latter involves approximately only one-sixth as many multiplications as the former, it is to be preferred unless accuracy is at a premium. In this context it should be remembered that the rates themselves are only inaccurately known, so that it is unlikely that a method of high accuracy can be justified in practice.

The Collisional-Radiative Equation

In general the fractional population of a state, i, of some stage of ionisation is determined by an equation of the form [3]:

$$dq_i/dt = \sum_{j \neq i} (X_{ij}q_j - X_{ji}q_i) = -\sum R_{ij}q_j, \tag{1}$$

where the sum is taken over all states of that ionic species with a direct transition to the state, i. In principle, we may extend this sum to include all states of all ionic stages of that species by imposing upon the transition matrix, X, an appropriate sparsity pattern. The matrix, X, for a given element must be irreducible, for by the principle of detailed balance, every transition X_{ij} must be accompanied by its inverse, X_{ji}, and furthermore we may assume that there exists at least one transition into every state. The rate matrix, R, is given by

$$R_{ii} = \sum_{j \neq i} X_{ji}: \qquad R_{ij} = -X_{ij}. \tag{2}$$

The sparsity pattern imposed on the transition matrix, X, and therefore the rate matrix, R, will in practice depend on the energy level structure considered for each ionic stage. This may range from a tri-diagonal form (ground states only) to completely dense matrix (one ionisation stage alone). The rate coefficients, X_{ij}, are by definition real, non-negative and bounded, so that R is an M-matrix form [6].

In the absence of chemical interactions, the transition matrix for different species is separable and we may consider the behaviour of each element alone. We shall, therefore, henceforward consider one element alone.

The set of equations (1) has some important properties, which are necessary for their physical interpretation. In the past these have been assumed on physical grounds; however, since the rates, X_{ij}, are not exactly known, it is important to show that these properties follow as a direct consequence of the above form of the equations, and do not depend on the exact values of the rates.

(a) $\sum_i q_i$ is conserved, for clearly

$$\sum_i dq_i/dt = 0. \tag{3}$$

In conformity with the usual definition we put

$$\sum_i q_i = 1. \tag{4}$$

(b) The equations are positivity preserving; i.e., if the vector set $q^0 \geqslant 0$ at an initial time, t_0, then the set $q \geqslant 0$, at a later time, t. The formal solution of Eq. (1) may be written

$$\begin{aligned} q_i(t) = q_i^0 \exp &\left\{-\int_{t_0}^{t} \sum_{j \neq i} X_{ji}\, dt'\right\} \\ &+ \sum_{j \neq i} \int_{t_0}^{t} X_{ij}(t')\, q_j(t') \exp\left\{-\int_{t'}^{t} \sum_{j \neq i} X_{ji}\, dt''\right\} dt'. \end{aligned} \tag{5}$$

The coefficients X_{ij} are all non-negative and bounded. Therefore if $q_j(t') \geqslant 0$ for $j \neq i$ and $t_0 \leqslant t' < t$, then $q_i\ (t) \geqslant 0$. Hence, by induction, since $q_i^0 \geqslant 0$, it follows that $q_i \geqslant 0$ for all i. Since the matrix X is irreducible we may extend this result to conclude that $q(t) > 0$ for all finite time intervals $t > t_0$.

(c) The equations are stable. The exact solution of the set of differential equations can be obtained by reducing R to its Jordan normal form. The resultant equations are stable if and only if the non-zero eigenvalues of R have positive real parts. The nature of the eigenvalues of R is readily established by consideration of its transpose for

$$R_{ii}^{\mathrm{T}} = -\sum_{j \neq i} R_{ij}^{\mathrm{T}}. \tag{6}$$

The matrix R^{T}, and therefore R, is singular and its non-zero eigenvalues have positive real parts [7]. The equations are therefore stable and, if the rates are constant, converge to an equilibrium state associated with the zero eigenvalue.

We may remark on the relation of this condition to those of (a) and (b), for in the set, q, the term q_i is bounded by the sum $\sum_i q_i$ which is constant. This does not ensure convergence to a steady state since purely oscillatory solutions may occur. These are not permitted by the stable condition.

(d) The equilibrium state is unique. The equilibrium state is described by the eigenvector with zero eigenvalue. However, if this is a multiple root of the characteristic equation, the equilibrium state may not be unique, and the final equilibrium condition may depend on the path by which it is approached. Such phenomena are found in phase equilibria, for example. In this case we show that the matrix, R, has only one zero eigenvalue.

Let q^s be an eigenvector of R with zero eigenvalue, and let Q be the departure of any state q from this equilibrium:

$$Q_i = q_i - q_i^s: \qquad \sum_i Q_i = 0 \tag{7}$$

and

$$dQ_i/dt = -\sum_j R_{ij} Q_j. \tag{8}$$

We may reduce the order of the equation from n to $(n-1)$ by eliminating one component (say Q_n) to give

$$dQ_i/dt = -\sum_{j=1}^{n} R_{ij} Q_j = -\sum_{j=1}^{n-1} S_{ij} Q_i \tag{9}$$

with $Q_n = -\sum_{j=1}^{n-1} Q_i$. Clearly S and R have the same eigenvalues and eigenvectors, except that the eigenvector due to q^s is removed from the matrix S. The components of S are

$$\begin{aligned} S_{ij} &= X_{in} - X_{ij}, \qquad i \neq j, \\ S_{ii} &= X_{in} + \sum_{j \neq 1} X_{ji}. \end{aligned} \tag{10}$$

Since at least one pair (X_{in}, X_{ni}) must be non-zero the matrix S is non-singular (Appendix A). The matrix, R, therefore has only one zero eigenvalue and the equilibrium state is unique. Since the matrix X is irreducible and non-negative, it follows that no component of the equilibrium vector q^s is zero.

(e) The equilibrium state is physical, for if one component of q^s (say q_n^s) is positive, then since R is an irreducible M-matrix and therefore monotone [6], all the remaining $(n-1)$ components are solutions of the $(n-1)$ equations:

$$\sum_{j=1}^{n-1} R_{ij} q_j^s = -R_{in} q_n^s$$

and therefore positive.

(f) The solution of Eq. (1) converges uniformly to the steady state when the rates, X, are constant; i.e., the maximum norm of the fractional deviation vector (Q_i/q_i^s) at a time, t, satisfies

$$\|Q_i/q_i^s\| < \|Q_i^0/q_i^s\|, \tag{11}$$

where Q^0 is the value of Q at an earlier time, $t_0 < t$, and is assumed to be nonzero.

We define the vector

$$\mathcal{Q}_i = q_i/q_i^q, \tag{12}$$

and Eq. (1) becomes

$$d\mathcal{Q}_i/dt = \sum_{j=1}^{n} R'_{ij}\mathcal{Q}_j, \tag{13}$$

where $R'_{ij} = q_j^s/q_i^s R_{ij}$. If L is a uniform vector, i.e., one whose components are equal, then clearly

$$\sum R'_{ij} L_j = 0, \tag{14}$$

for the case $\mathcal{Q}_1 = \mathcal{Q}_2 = \cdots = 1$ corresponds to the steady state. Let L be the smallest component of $\mathcal{Q}^0$ at time, t_0; then

$$(\mathcal{Q} - L) > 0, \qquad (\mathcal{Q}^0 - L) \geqslant 0: \qquad t > t^0, \tag{15}$$

since Eq. (13) must be positivity preserving. Hence, since L is constant in time,

$$\operatorname{Min}(\mathcal{Q}) > \operatorname{Min}(\mathcal{Q}^0). \tag{16}$$

Applying this result to the set $(-\mathcal{Q})$ we obtain

$$\operatorname{Max}(\mathcal{Q}) < \operatorname{Max}(\mathcal{Q}^0). \tag{17}$$

Since $\sum_i q_i = \sum_i q_i^s$, it follows that $\operatorname{Max}(\mathcal{Q}) \geqslant 1$ and $\operatorname{Min}(\mathcal{Q}) \leqslant 1$. Hence since $\mathcal{Q}_i = 1 + Q_i/q_i^s$, the result (11) follows.

That this solution converges on to the equilibrium state follows from the exact solution in terms of the eigenvector projection of q, which decays exponentially.

Numerical Solutions

In general, we wish to evaluate the vector set q at some time, t, given the values of the set, q^0, at an earlier time, t_0, by means of a linear operation[1]

$$q = G(q^0). \tag{18}$$

In view of the nature of the collisional-radiative equation, the operation $G(q^0)$ is

[1] If G is linear in q, then $G(aq_1, aq_2,...) = aG(q_1, q_2,...)$, where a is a constant multiplier of the complete set, q. We note that since the average ionisation, $\bar{Z} = \sum_i Z_i q_i / \sum_i q_i$, where Z_i is the ionic charge of state i, we may include variation of the rates due to ionisation induced changes of electron density, the operator, G, remaining linear.

likely to be one of matrix multiplication. We have identified several important properties which we may require preserved by the operator G, namely:

(a) G is conservative if

$$\sum_i q_i = \sum_i q_i^0; \tag{19}$$

(b) G is positivity maintaining if for any non-negative initial stage, q^0,

$$q \geqslant 0 \qquad \text{if} \quad q^0 \geqslant 0. \tag{20}$$

Furthermore, by analogy with the exact solution, we require that if the operation G is repeated a sufficient number of times, then $q > 0$.

(c) G is equilibrating if

$$q = q^0 \qquad \text{if and only if,} \qquad q^0 = q^s, \tag{21}$$

where q^s is the unique set of steady state values of q corresponding to G. Clearly, if G is consistent with Eq. (1), such a set must exist and be unique.

Two important theorems govern the repreated application of the operation G.

Theorem 1. *If G is conservative and positivity preserving, then G is stable.*

Since the sum, $\sum_i q_i$, of a set of positive values q_i is constant, each value q_i is bounded, and the theorem follows.

Theorem 2. *If, and only if, G is equilibrating and positivity maintaining, then G is convergent; i.e., repeated application of G converges uniformly to the steady state, q^s.*

The proof of this result follows a similar approach to that of the analogous theorem for the differential equation. Thus we consider the behavior of the vector $\mathscr{Q}_i$ by means of the operation

$$\mathscr{Q}_i = q_i^s G(q_1^s \mathscr{Q}_1^0, \ldots) = G'(\mathscr{Q}_1^0, \ldots). \tag{22}$$

Since G is equilibrating, G' is differential; i.e., $\mathscr{Q}_i = \mathscr{Q}_i^0$ if, and only if $\mathscr{Q}_i^0$ is uniform. Furthermore, since G is consistent with Eq. (1), G' must be linear in $\mathscr{Q}$. Thus if L is the uniform vector whose values equal the smallest component of $\mathscr{Q}^0$,

$$(\mathscr{Q}^0 - L) \geqslant 0 \qquad \text{and} \qquad \mathscr{Q} - L = G'(\mathscr{Q}^0 - L) \geqslant 0, \tag{23}$$

since G' is positivity maintaining, and by hypothesis $\mathscr{Q}^0 \neq L$. Thus the smallest value of $\mathscr{Q}$ is not less than that of $\mathscr{Q}^0$. Similarly the largest value of $\mathscr{Q}$ is not greater than that of $\mathscr{Q}^0$. Since repeated application of a positivity-maintaining operator implies $(\mathscr{Q} - L) > 0$, we conclude that the vector $\mathscr{Q}$ is convergent. Furthermore, if the operator G is convergent, it is clearly both positivity maintaining and equilibrating.

Finally, since the operator G is linear in q, we conclude that the solutions, q, must converge to the equilibrium state, q^s.

A convergent operator is clearly stable.

Two-Stage Schemes

As an example of the application of these theorems we consider a general two-stage scheme for which a simple linear form was proposed in [1]. In these methods the rate coefficient matrix is assumed to be tri-diagonal in form. The ionisation change across a pair of isolated levels q_l and q_{l+1} is then calculated, Δ_l. The net ionisation is obtained by recursively evaluating

$$q_l = q_l^0 + \Delta_{l-1} - \Delta_l \tag{24}$$

subject to $\Delta_0 = 0$. The scheme is clearly conservative. It is not, however, positivity maintaining. This condition is readily achieved by a simple limit on Δ_l such that

$$\Delta_l < (q_l^0 + \Delta_{l-1}) \qquad \text{and} \qquad \Delta_l > -q_{l+1}^0 \tag{25}$$

if the recursion in (24) is performed with increasing l. This check, which is readily performed in the recursion, has been routinely included in our codes using this method. In our experience the clipping is extremely rarely required. The two-stage scheme in this form is stable.

If the form of Δ_l is chosen so that

$$\Delta_l = 0 \qquad \text{if} \quad q_l/q_l^s = q_{l+1}/q_{l+1}^s$$

as in [1], then the scheme is equilibrating. The positivity-maintaining form is therefore convergent. These general proofs replace the crude stability analysis given in [1], and are applicable to all two-stage methods.

Eigenvector Decomposition

In principle the exact solution of Eq. (1) can be obtained by a similarity transformation of R into its Jordan normal form. In practice this method, although feasible, is complicated and involves excessive computer operations when compared to the finite difference approximations to be described later. However, it should be borne in mind that it does, in principle, provide exact solutions, whereas those of the finite difference are only approximate.

Since the matrix, R, is singular, we must remove the singularity by evaluating the steady state solution q^s. We may reduce the order of the matrix by one using S, and the set $Q = q - q^s$, as in Eq. (9). The projection of the set Q onto the set of eigenvectors of S is straightforward if S is not defective or nearly defective. However, the problems introduced by defectiveness into a general algorithm, which together with the large number of operations required make this approach unattractive for large repetitive calculation, are discussed in Appendix B.

Finite Difference Algorithms

A more satisfactory approach is to use a finite difference scheme. If we require such a scheme to be unconditionally stable, i.e., for all step lengths, Δt, it is necessary that the finite difference form of Eq. (1) be stable over the complete positive complex half plane of $\lambda\Delta t$, the routine must therefore be A-stable [5]. There is a fundamental theorem [5] which states that multi-step schemes cannot be A-stable if their order is greater than two. Therefore if we require an algorithm which is unconditionally stable for all forms of the rate equation (1), we must restrict ourselves to two-level multistep schemes. Such a restriction is, of course, fully compatible with our storage requirement discussed earlier.

We may write Eq. (1) in a finite difference form, using a weighted mean for the terms on the right-hand side,

$$\begin{aligned}(q_i^{N+1} - q_i^N)/\Delta t = \sum_{j\neq i} X_{ij}[W_{ij}q_j^{N+1} + (1 - W_{ij})q_j^N]\\ -X_{ji}[W_{ji}q_i^{N+1} + (1 - W_{ji})q_i^N],\end{aligned} \tag{26}$$

yielding n equations, which we may solve by matrix inversion. We may reduce the dimensions of the system (and hence the work) by making use of the conservation law

$$\sum q_i^{N+1} = \sum q_i^N \tag{27}$$

to eliminate q_n from the set of equations.

Hence we obtain the set

$$\sum_{j=1}^{n-1} C_{ij}q_j^{N+1} = B_i, \tag{28}$$

where

$$C_{ii} = 1 + \left[X_{in}W_{in} + \sum_{\substack{j\neq i\\ j=1}}^{n-1} X_{ji}W_{ji}\right]\Delta t\text{: } C_{ij} = [X_{in}W_{in} - X_{ij}W_{ij}]\Delta t \tag{29}$$

and

$$B_i = q_i^N \left\{ X_{in} W_{in} - \sum_{\substack{j\neq i\\ j=1}}^{n-1} X_{ji}(1 - W_{ji}) \right] \Delta t \Big\}$$

$$+ \sum_{\substack{j\neq i\\ j=1}}^{n-1} q_j^N \{ X_{in} W_{in} + X_{ij}(1 - W_{ij}) \} \Delta t. \tag{30}$$

An alternative form of (28), suitable for use with an energy limit, is obtained by considering the change in q_i,

$$\Delta_i = q_i^{N+1} - q_i^N: \qquad \Delta_n = - \sum_{j=1}^{n-1} \Delta_j \tag{31}$$

to give

$$\sum_{j=1}^{n-1} C_{ij} \Delta_j = B_i', \tag{32}$$

where

$$B_i' = \sum_{\substack{j\neq i\\ j=1}}^{n} \{ X_{ij} q_j^N - X_{ji} q_i^N \} \Delta t. \tag{33}$$

The matrix C is, in general, dense and Eq. (28) or Eq. (32) must be solved by either an elimination or an iterative method. In view of the fact that one usually only works with a restricted set of states, the demands of a compact elimination method are not severe. We note that since C is the form

$$C = I + S' \Delta t,$$

that C is a matrix of the type considered in Appendix A, and is therefore nonsingular.

The set of equations (26) is clearly conservative. The derived set (28) is obtained using the conservation law (27) and must therefore be conservative.

In equilibrium

$$\sum_{j\neq 1} X_{ij} q_j^s = \sum_{j\neq i} X_{ji} q_i^s \,. \tag{34}$$

Hence if $q^N = q^s$, $B' = 0$ and therefore, since C is non-singular, $\Delta = 0$. Equation (32) and hence (28) and (26) are equilibrating, the steady state being that of the true solution.

This general algorithm is not in general positivity maintaining. To incorporate this requirement we could include a post-evaluation check, as in the two-stage scheme. A more satisfactory approach, however, is to assign the weights so as to ensure this condition is upheld.

We write Eq. (26) in the form

$$q = M^{-1} P q^0 \tag{35}$$

which is positivity-maintaining if, and only if, M is a monotone [8] and P a non-negative matrix. In fact,

$$P_{ii} = 1 - \sum_{j \neq i} (1 - W_{ji}) X_{ji} \Delta t: \qquad P_{ij} = (1 - W_{ij}) X_{ij} \Delta t, \tag{36}$$

$$M_{ii} = 1 + \sum_{j \neq i} W_{ji} X_{ji} \Delta t: \qquad M_{ij} = - W_{ij} X_{ij} \Delta t. \tag{37}$$

Since X_{ij} is non-negative, M is clearly an M-matrix, and therefore monotone and P is non-negative if, and only if, $P_{ii} \geqslant 0$. If this condition is upheld, the algorithm, and those derived from it, are both stable and convergent.

The Choice of Weights

The weights, W_{ij}, must be chosen in such a way as to enable the solution of the finite difference equations to approach the exact one for arbitrary values of the step Δt. Clearly this requires that in the limits

$$\begin{aligned} &W_{ij} \to \tfrac{1}{2} \quad \text{as} \quad \Delta t \to 0 \\ \text{and} \quad & \\ &W_{ij} \to 1 \quad \text{as} \quad \Delta t \to \infty. \end{aligned} \tag{38}$$

The weights are used in an attempt to represent the degree of equilibration of a pair of states i and j during the time step. Thus if the rates are large, $(X_{ij} + X_{ji}) \Delta t \gg 1$, the states will be instantaneously equilibrated throughout the time step, and $W_{ij} = 1$ is appropriate. On the other hand, if $(X_{ij} + X_{ji}) \Delta t << 1$, the states will not equilibrate and $W_{ij} = \frac{1}{2}$ is suitable. It is clear that each weight is associated with a transition and therefore

$$W_{ij} = W_{ji} = f(\lambda_{ij}), \tag{39}$$

where $\lambda_{ij} = (X_{ij} + X_{ji}) \Delta t$, and f is a function with the limits (38).

We may solve Eq. (1) for the case of a two-state system to give

$$\Delta_1 = (X_{12} q_2^0 - X_{21} q_1^0) \Delta t \{1 - \exp(-\lambda_{12})\} / \lambda_{12} \tag{40}$$

in contrast to the finite difference result,

$$\Delta_1 = (X_{12} q_2^0 - X_{21} q_1^0) \Delta t / \{1 + W_{12} \lambda_{12}\}. \tag{41}$$

Thus the finite difference form is exact if the weights

$$W_{12} = 1/\{1 - \exp(-\lambda_{12})\} - 1/\lambda_{12} \tag{42}$$

We note that this function has the limits (38) and is therefore a suitable weight function.

In view of the nature of the weights, to allow for rapid equilibration between two strongly coupled states, we may expect that it is an appropriate form.

Alternative functions, simpler but of similar form, are

$$W = \operatorname{Max}\{\tfrac{1}{2}, (1 - 1/\lambda)\} \tag{43}$$

and

$$= \operatorname{Max}\{\tfrac{1}{2}, 1 - [1/\lambda - \exp(-\lambda)]\}. \tag{44}$$

Figure 1 shows a representative calculation for a simple four-level system with transition matrix as given, for different values of the time step, Δt. The accurate solution was calculated using Gear's routine [4]. For comparison the values obtained by a split time step ($W = \frac{1}{2}$) and fully implicit ($W = 1$) calculations are also shown. The marked improvement obtained using the weights can be clearly seen. Of the three weight functions the two-level form, Eq. (42), shown in Fig. 1 is marginally superior, a result consistent with other tests of this set.

These choices of weights do not in general ensure positivity. This requires that

$$\sum_{j \neq i} (1 - W_{ji}) X_{ji} \Delta t \leqslant 1 \tag{45}$$

for all i. Let us suppose that the level i has N_i non-zero transition elements connecting it to other states j; then the above condition is satisfied if

$$W_{ji} \geqslant 1 - 1/N_i X_{ji} \Delta t. \tag{46}$$

This may be put into a form consistent with (38):

$$W_{ij} = W_{ji} = \operatorname{Max} \left\{ \begin{matrix} \frac{1}{2} \\ 1 - 1/\operatorname{Max}(N_i, N_j)\, \lambda_{ij}) \end{matrix} \right\}, \tag{47}$$

or by analogy with Eq. (42),

$$W_{ij} = W_{ji} = \operatorname{Max} \left\{ \begin{matrix} 1/[1 - \exp(-\lambda_{ij})] - 1/\lambda_{ij}, \\ 1 - 1/[\operatorname{Max}(N_i, N_j)\, \lambda_{ij}]. \end{matrix} \right. \tag{48}$$

Of these, the first, which is computationally simpler, is preferred.

Figure 1 also shows the calculation of the identical matrix problem as before by the positivity-preserving weights (47). It can be seen that the results are considerably more accurate than those obtained without the positivity limit. Calculations show

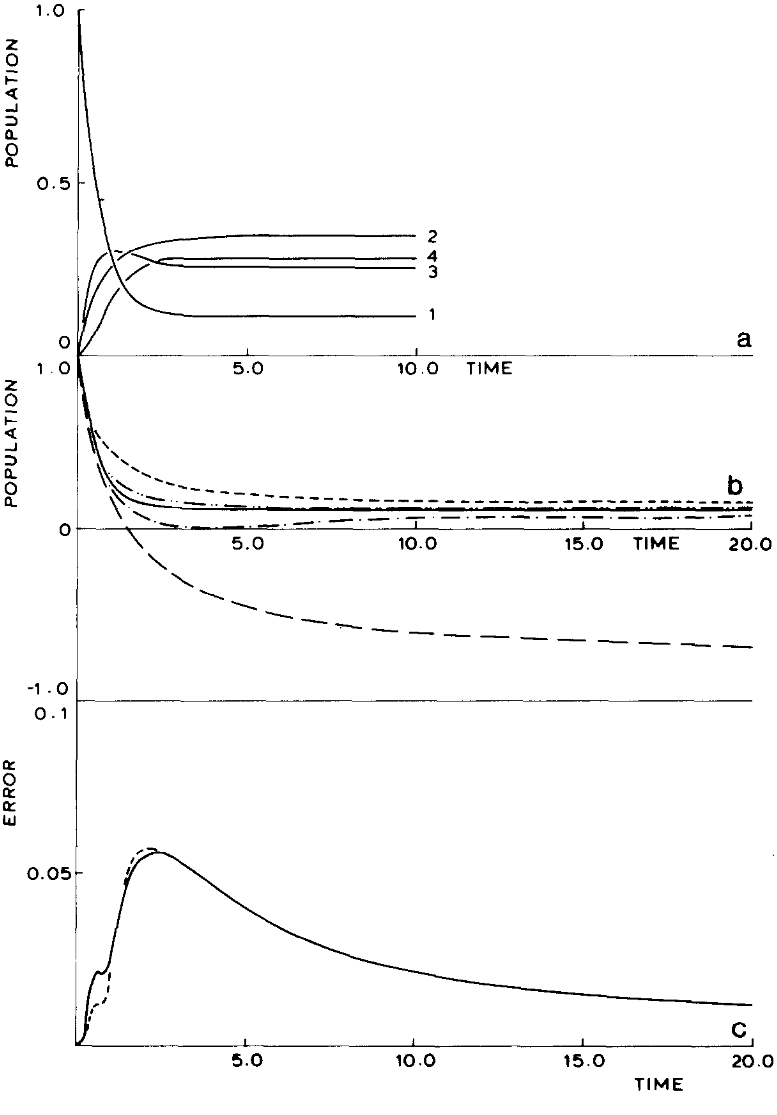


FIG. 1. (a) The exact solution of a four-level system as a function of time. The transition matrix having values (0.1, 0.2, 0.3; 0.5, 0.2, 10^{-3}; 1.0, 0.2, 0.6; 10^{-4}, 10^{-2}, 1.0) and the initial state (1.0, 0.0, 0.0, 0.0). Under these conditons the maximum error incurred by a single finite difference calculation is with level 1. (b) The population of level 1 as a function of time-step for the exact solution ——, and for various weight functions: – – –, $W=\frac{1}{2}$; ···, $W=1$; – · –, $W=\{1/(e^{-\lambda}-1)-1/\lambda\}$; and – · · –, $W=\mathrm{Max}\{(1-1/3\lambda), \frac{1}{2}\}$. Note the markedly improved accuracy of the positivity-maintaining solution. Comparison of the maximum errors of different positivity-maintaining weights: ——, $W=\mathrm{Max}\{(1-1/3\lambda, \frac{1}{2}\}$; and – – –, $W=\mathrm{Max}\{(1-1/3\lambda), (1/(e^{-\lambda}-1)-1/\lambda)\}$.

that the simpler form (47) is marginally more accurate, and since it is computationally simpler it is to be preferred. This result is confirmed by subsequent tests with alternative matrices, for example, Fig. 2.

The Energy Limit

It was shown in [1] that if the system is near equilibrium, the energy exchange between ionisation and the electron temperature resulting from collisional ionisation and three body equilibration can lead to instability. In a similar fashion the inclusion of general collisional excitation and deexcitation energy exchange may also result in an unstable scheme. Thus near equilibrium

$$\Delta_i \simeq (dq_i/dT_e)\,\Delta T_e, \tag{49}$$

and the electron energy change due to a change Δ_i in the population of state i

$$= -V_i'\Delta_i \tag{50}$$

per ion, where $V_i' = V_i + \frac{3}{2}Z_i\,\mathrm{k}\,T_e$, and V_i is the total excitation and ionisation energy (from the ground state of the neutral atom). Thus, as before, considering an iteration for the electron temperature, we obtain on the next iteration, $\Delta T_e'$,

$$\tfrac{3}{2}\bar{Z}\,k\,\Delta T_e' \simeq -\sum V_i' \cdot (dq_i/dT_e)\,\Delta T_e \tag{51}$$

and is unstable if $|\sum V_i' \cdot dq_i/dT_e|/\frac{3}{2}\bar{Z}\mathrm{k} > 1$. This difficulty may be overcome by specifying a limiting energy transfer ΔE by replacing Δ_i by

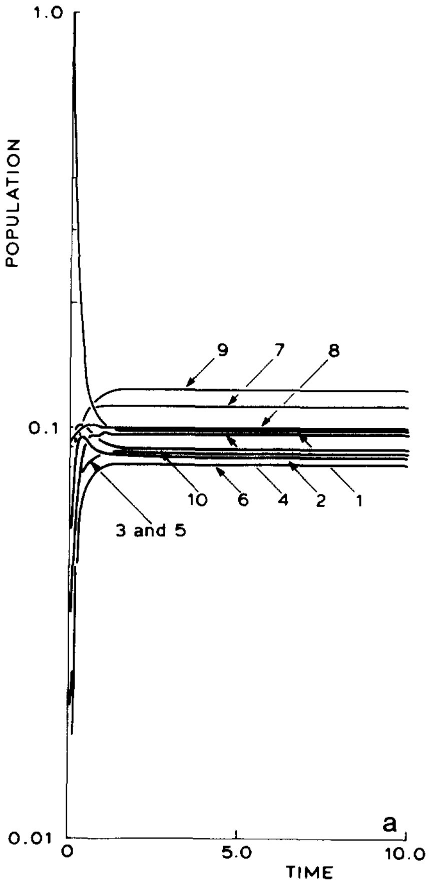
$$\Delta_i = \Delta_i' \Big/ \left(1 + \sum V_i\,\Delta_i'/\Delta E\right), \tag{52}$$

where Δ_i' is the value given by Eq. (31). A suitable value for ΔE is given in [1].

Discussion

The finite difference alogorithm described here has been widely used by the author in several studies. In an extensive series of tests the maximum error of the optimally weighted scheme did not exceed 0.07 and convergence to the equilibrium value was observed to be uniform. Figures 1 and 2 show the results of two such tests. In Fig. 1 a four-level system with rates arbitrarily distributed in the range 10^{-4} to 1 is considered. The maximum error is for state 1 at about two time units, and is about 0.06. The error decreasing to zero at $\Delta t \rightarrow 0$ or $\Delta t \rightarrow \infty$. Figure 2 shows a more difficult test in which 10 states with rates arbitrarily distributed in the range 0.1–1.0 are considered. In this case also the maximum error is less than 0.07 over this range in which all states simultaneously equilibrate. By comparison, the maximum error of the fully implicit scheme is 0.18 and that of centred difference 0.8; the latter method, in common with all non-positivity-maintaining weights ((42)–(44)), yielding negative values.

The most expensive computational element in the algorithm is the solution of the set of linear equations (31) which involves $\frac{1}{3}(n-1)^3$ multiplications. A significant reduction in the computational work required can be achieved by reduction of the algorithm to a two-stage form. In its most extreme form this involves the treatment of each ionisation separately by either a collisional-radiative form or a single level



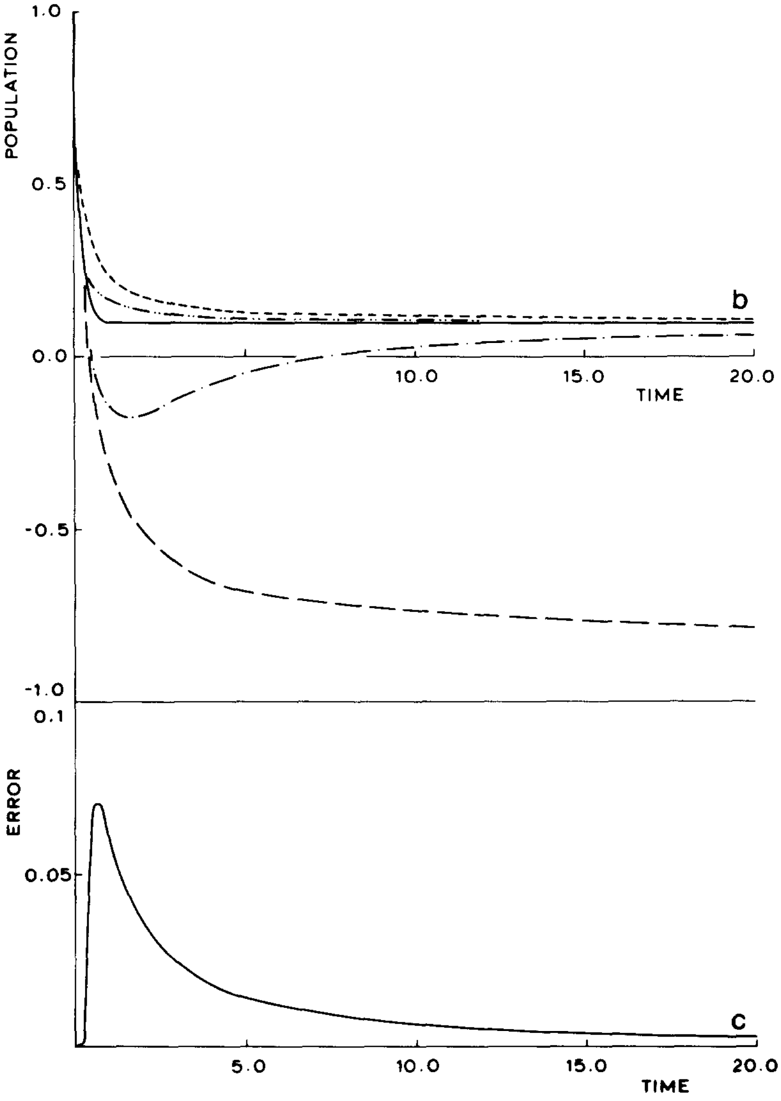
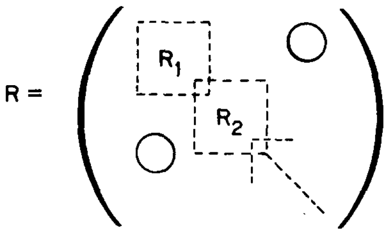


FIG. 2. A test similar to that in Fig. 1 is shown for a 10-level system with transition matrix having values in the range 0.1–1.0 only. The arrangement of the figures is as in Fig. 1. It is noteworthy that this example was the least accurate of all tested in this series of numerical experiments performed to investigate the merits of these weight functions.

(collision-limit) approximation. In this form, the sparse areas of the matrix, R, which has the form of a set of diagonal overlapping blocks



(53)

are separated. The algorithm involves an independent solution for each block, R_i. The number of multiplications involved is then $\frac{1}{3}\prod_{i=1}^{N}(n_i-1)^3$ when there are N blocks each of n_i levels. Clearly

$$n = 1 + \sum_{i=1}^{N}(n_i - 1). \tag{54}$$

This method is particularly convenient if the element contains a large number of ionisation stages, many of which are only briefly involved in the calculation. It is a straightforward generalisation of the two-stage method of [1] and is, of course, stable and convergent.

In principle the number of discrete levels of each ionisation stage is limited only by the depression of the ionisation level. In practice one must introduce some appropriate limit to the number of levels considered. This can only be determined by consideration of the individual characteristics of the system, for example, its L.T.E. limit, and time variation. The levels above this limit are included by some appropriate averaging such as is used in the collision-limit approximation.

If we compare the finite difference solution with the eigenvector decomposition we may note that increased accuracy of the latter is only obtained at the expense of a large increase incomputational work: the finite difference solution requiring approximately $\frac{1}{3}(n-1)^3$ multiplications [9] compared with $2(n-1)^3$ for the eigenvector scheme. Since experience has shown that the ionisation routines typically increase the run-time of a standard hydrocode by a factor of about 4 (depending on the element, number of levels, etc.); this clearly represents a prohibitive further increase for most problems.

The finite difference algorithm described here has been incorporated into the one dimensional Lagrangian fluid code, and the general similarity code described in [1]. With the modification to the energy limit described earlier, the scheme has worked well, and no problems have been encountered. The programmes have been used extensively in several studies in connection with XUV laser action. Figure 3 shows a typical output from the similarity code. The model is used to study the population growth of the hydrogen-like carbon ion C VI during therapid expansion of a laser heated fibre. The code parameters were set to model the experiments reported in [11].

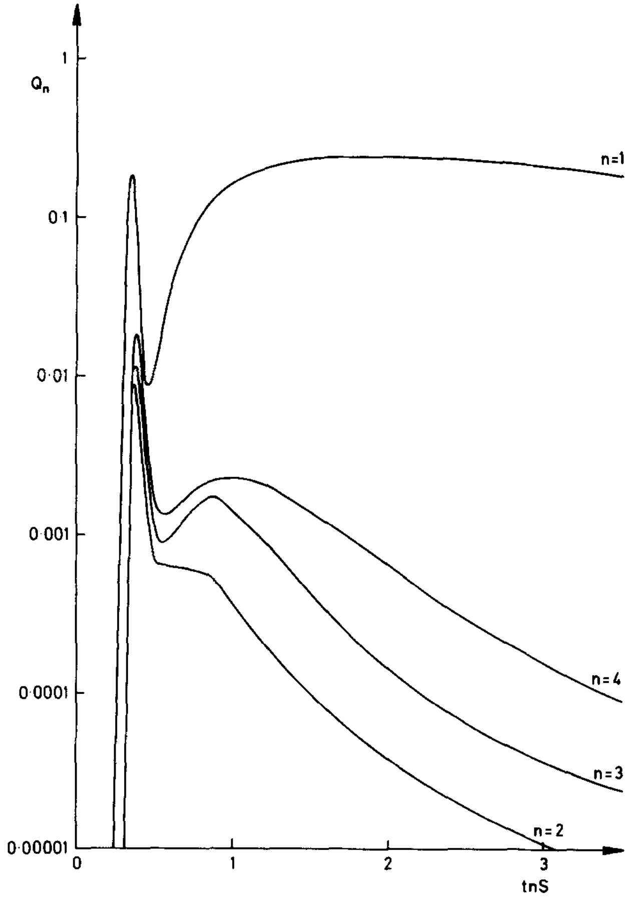


FIG. 3. The development of the populations in the hydrogenic levels $n = 1$ to $n = 4$ of CVI during irradiation of a 5-μm-diameter carbon fibre by a 150-mJ laser pulse of 140 psec duration focussed into a 40-μm-diameter focal spot.

Then run was carried out with 10 energy levels of C VI in a two-stage scheme, although only the populations of the lowest four are shown. It can be seen that the algorithm is well-behaved throughout the run.

APPENDIX A: THEOREM ON DETERMINANTS

LEMMA 1. *If a or its transpose is an irreducibly diagonally dominant real matrix* [6] *of order n, i.e.,*

$$a_{ii} \geqslant \sum_{j \neq i} |a_{ij}| \qquad or \qquad a_{ii} \geqslant \sum_{j \neq i} |a_{ji}| \qquad (A.1)$$

with equality in no more than $(n-1)$ *cases, then the determinant of a*

$$|a_{ij}| > 0. \tag{A.2}$$

Since a is a real matrix, its eigenvalues are real or occur in complex conjugate pairs. Furthermore the eigenvalues of an irreducibly diagonally dominant matrix have positive real parts [6, 7], and their product, the determinant, is therefore positive.

If equality occurs in all cases the matrix is singular and the determinant zero [7].

LEMMA 2. *If b is a real matrix of order n with*

$$b_{11} > 0 : b_{1i} \geqslant 0 : b_{ij} \leqslant 0, \qquad 1 \leqslant j \leqslant n$$

and

$$b_{ii} \geqslant \sum_{\substack{j \neq i \\ j=2}}^{n} |b_{ij}| \qquad \text{or} \qquad b_{ii} \geqslant \sum_{\substack{j \neq i \\ j=2}}^{n} |b_{ji}|, \qquad 2 \leqslant i \leqslant n \tag{A.3}$$

with equality in no more than $(n-2)$ *cases, then the determinant of b*

$$|b_{ij}| > 0. \tag{A.4}$$

Furthermore if equality occurs in all $(n-1)$ *cases, then the determinant of b is nonnegative.*

The cofactor of b_{11} is the determinant of an irreducibly diagonally dominant real matrix, and is therefore positive. The cofactor of b_{1j} is

$$(-1)^{(j+1)} \begin{vmatrix} b_{21} & b_{22} & \cdots & b_{2,j-1} & b_{2,j+1} & \cdots & b_{2n} \\ \vdots & & & & & & \end{vmatrix} = (-1)^{(2j-2)} \begin{vmatrix} -b_{j1} & b_{j2} & \cdots & b_{jn} \\ -b_{21} & b_{22} & \cdots & \\ \vdots & & & \end{vmatrix}, \tag{A.5}$$

and since $b_{ij} \leqslant 0$ $(i \neq j,\ 2 \leqslant i \leqslant n,\ 1 \leqslant j \leqslant n)$, this is a determinant of the same form as b, but order $(n-1)$. Since the lemma is clearly true for $n = 2$, it follows by induction that it holds for all orders n.

THEOREM. *The real matrix r is of order* $(n-1)$ *and is of the form*

$$r_{ij} = s_i + t_{ij} \qquad s_i \geqslant 0 \tag{A.6}$$

and t_{ij} *or its transpose is an irreducibly diagonally dominant M-matix; i.e.,*

$$t_{ii} \geqslant \sum_{j \neq i}^{(n-1)} |t_{ij}| \qquad \text{or} \qquad t_{ii} \geqslant \sum_{j \neq i}^{(n-1)} |t_{ji}| \quad \text{and} \quad t_{ij} \leqslant 0, i \neq j \tag{A.7}$$

with equality in no more than $(n-2)$ *cases.*

The determinant, which may be written,

$$|r_{ij}| = \begin{vmatrix} 1 & 0 & 0 & \cdots \\ -s_1 & r_{11} & r_{12} & \cdots \\ -s_2 & r_{21} & r_{22} & \cdots \end{vmatrix} = \begin{vmatrix} 1 & 1 & 1 & \cdots \\ -s_1 & t_{11} & t_{12} & \\ -s_2 & t_{21} & t_{22} & \end{vmatrix} \tag{A.8}$$

is therefore a determinant of order n of type b. Therefore by Lemma 2

$$|r_{ij}| > 0 \tag{A.9}$$

and the matrix r is non-singular. The matrix r is only singular if there is equality in all $(n-1)$ terms t_{ii}, and the terms s_i are all zero.

APPENDIX B: An Algorithm Using the Exact Solution

In principle the set of differential equations (1) may be exactly integrated if the transformation of R to its Jòrdan normal form is known [9]. If R is not defective this is readily accomplished by the projection of the vector, q, onto the biorthogonal set of eigenvectors of R. In practice it is necessary to first remove the singularity of R by solving the set of $(n-1)$ linear equations

$$\sum_{j=1}^{(n-1)} S_{ij} q_j^s = -R_{in}, \tag{B.1}$$

where S is given by (10) to determine the equilibrium state vector, q^s. Then if S is not defective, the deviation vector, Q, is given by

$$Q = \sum_{k=1}^{(n-1)} (y_k, Q^0)\, x_k \exp\{-\lambda_k(t-t_0)\}, \tag{B.2}$$

where Q^0 is the value of Q at time t_0, and y_k and x_k are left- and right-hand eigenvectors of S with eigenvalue λ_k.

This form can only be used if the matrix is not defective; i.e., no two eigenvectors are parallel [9]. In practice numerical evaluation introduces errors so that exact parallelism is unlikely to occur. However, when the matrix is nearly defective, attention must be paid to the round-off error terms. Consider the example:

$$S = \begin{pmatrix} a & 1 \\ 0 & b \end{pmatrix}, \tag{B.3}$$

which has eigenvectors and eigenvalues,

$$\lambda = a : \begin{pmatrix} 1 \\ 0 \end{pmatrix} : (1, 1/(a-b)) \tag{B.4}$$

$$= b : \begin{pmatrix} 1/(b-a) \\ 1 \end{pmatrix} : (0, 1). \tag{B.5}$$

The solution $Q = \binom{x}{y}$ is therefore

$$x = [x_0 + y_0/(a-b)]e^{at} + y_0/(b-a)e^{bt} \tag{B.6}$$

$$= x_0 - y_0/(b-a)[1 - e^{(b-a)t}]e^{at}, \tag{B.7}$$

$$y = y_0 e^{bt}. \tag{B.8}$$

The first equation (B.6) for x is that generated numerically. Due to evaluation errors, the denominator term $(a-b)$ is not exactly equal to the corresponding one $(b-a)$, nor the difference of the exponential arguments. Thus proceeding to the limit $b \to a$, as the matrix becomes defective,

$$x \to [x_0 + y_0 t]\, e^{at} \tag{B.9}$$

which is an exact solution. The numerical calculation will not achieve this result unless care is taken to ensure the limit by a reduction to the exact solution. This may be accomplished by a series of tests on the eigenvalues (for degeneracy) and subsequently on the eigenvectors (for parallelism). If the order of the non-linear divisors is greater than two, the appropriate higher-order analytic solution must be used [9].

The determination of the eigenvectors of S can be performed by a standard method. Since R is in general an asymmetric matrix with no clearly defined sparsity pattern, a general eigenvalue method must be used. The most efficient such method involves a reduction to Hessenberg form followed by an iterative evaluation of the eigenvalues using the LR or QR schemes [9]. The reduction to Hessenberg form is generally the most computationally expensive involving approximately $5/3\ (n-1)^3$ multiplications for Householder's method [9]. The QR algorithm involves a further $4(n-1)^2$ multiplications per iteration step [9]. Thus, including the necessary preliminary solution of the equilibrium state, this method involves approximately $2(n-1)^3$ multiplications.

In view of the complicated structure of the scheme outlined above and the iterative nature of the calculations, it is natural to consider the suitability of an iterative scheme for calculating eigenvalues, for which the transformation obtained in the previous time step may be used as a first approximation. Jacobi's method [10] provides such a scheme. However, the general method involves about $8(n-1)^3$ multiplications per sweep, and is therefore considerably more expensive than the QR iterative method.

Acknowledgments

This work was carried out as part of the XUV laser programme supported by the Science Research Council.

REFERENCES

1. G. J. PERT, *J. Comput. Phys.* **27** (1978), 241.
2. D. R. BATES, A. E. KINGSTON AND R. W. P. MCWHIRTER, *Proc. Roy. Soc. Ser. A*, **267** (1962), 297; **270** (1962), 185.
3. R. W. P. MCWHIRTER AND A.G. HEARN, *Proc. Phys. Soc.* **82** (1963), 641.
4. C. W. GEAR, "Numerical Initial Value Problems in Ordinary Differential Equations," Prentice–Hall, Englewood Cliffs, N. J., 1971.
5. G. G. DAHLQUIST, *BIT* **3** (1963), 27.
6. R. S. VARGA, "Matrix Iterative Analysis," Prentice–Hall, Englewood Cliffs, N. J., 1962.
7. O. TAUSSKY, *Amer. Math. Monthly* **56** (1949), 672.
8. L. COLLATZ, "The Numerical Treatment of Differential Equations," Springer-Verlag, Berlin, 1960.
9. J. H. WILKINSON, "The Algebraic Eigenvalue Problem," Oxford Univ. Press, Oxford, 1965.
10. P. J. EBERLEIN, *in* "Linear Algebra" (J. H. Wilkinson and C. Reinsch, Eds.), Springer-Verlag, Berlin, 1971.
11. R. J. DEWHURST, D. JACOBY, G. J. PERT, AND S. A. RAMSDEN, *Phys. Rev. Lett.* **37** (1976), 1265.