COMPARISON OF DISCRETE SPACE AND DIFFERENTIAL EQUATION METHODS IN NON-LTE LINE TRANSFER PROBLEMS

by

I. P. Grant and G. E. Hunt

Science Research Council, Atlas Computer Laboratory Chilton, Didcot, Berkshire, England

ABSTRACT

Numerical methods are essential to the treatment of line formation in inhomogeneous non-LTE atmospheres. The new methods due to Hummer and Rybicki and to Feautrier now make it possible to make such calculations, although these are often quite time-consuming.

We shall describe an alternative approach using discrete space techniques depending on concpets of invariance. The solution algorithm is closely related to the method of Hummer and Rybicki, whose equations are obtained as a limiting case. The stability and errors of our algorithm are susceptible to mathematical analysis, and make it possible to identify the critical parameters in the calculation with precision. The results for a two-level problem will be compared with those from an implementation of the Rybicki-Hummer equations and a comparison will be made of the performance of the two procedures in respect of speed of computation and storage requirements.

Key words: line formation, inhomogeneous non-LTE atmospheres, numerical methods.

1. INTRODUCTION

In this paper, we shall be concerned with computational problems arising in the calculation of the line spectrum of a non-LTE model stellar atmosphere. This field has recently been treated in book form by

https://doi.org/10.1017/S0252921100151267 Published online by Cambridge University Press

Jefferies,[1] and Hummer and Rybicki[2] have reviewed the state of the art of computation as it was at October 1966. At the time that these articles were written, the best methods available for atmospheres in which the parameters vary with depth were the Riccati method[2,3] and that of Feautrier.[2,4,5] Since then, we have developed a theory of radiative transfer, based on principles of invariance, which is particularly suited to computation.[6,7] The equations of this theory have a simple structure which is easy to interpret in physical terms, and their solution is a relatively rapid and simple matter. Their form is sufficiently universal to make generalization easy, and it is possible to make a sufficiently precise analysis of error propagation and stability to give us confidence in the numerical results obtained.

We shall present an application of this general theory to the problem of determining the line profile due to non-coherent scattering in an inhomogeneous plane-parallel atmosphere. We have confined ourselves for the present to problems already treated by the Riccati method,[3] and we have therefore assumed that we are dealing with a model atom with two discrete levels only. One of our objects has been to see if our methods are in any way competitive with others already in the field; astrophysicists will want solutions to more elaborate model problems in the future, and reliable methods promising any economy while maintaining accuracy are highly desirable. We shall therefore make a detailed comparison of our technique with the Riccati method, and, in so doing, deal with some problems that we have encountered in carrying out their implementation.

2. GENERAL PRINCIPLES

Our computational procedures depend upon a conservation principle which may be expressed in the form

$$\begin{aligned} u^+(y) &= t(y,x)u^+(x) + r(x,y)u^-(y) + \Sigma^+(y,x) \\ u^-(x) &= r(y,x)u^+(x) + t(x,y)u^-(y) + \Sigma^-(x,y) \end{aligned} \qquad (1)$$

where x,y are depth co-ordinates in the plane-parallel atmosphere such that $a \leq x < y \leq b$. The quantities $u^+(x)$, $u^-(x)$, $u^+(y)$, $u^-(y)$ are the

https://doi.org/10.1017/S0252921100151267 Published online by Cambridge University Press

specific intensities of radiation at the corresponding levels x, y, the superscripts referring to the sign of the cosine, μ, of the angle between the direction in which the radiation travels and the normal to the layers in the sense of increasing depth, x. In the present case, $u^+(x)$, for example, will take values $u^+(x)(\mu,\xi) \equiv I(\mu,\xi;x)$, where $I(\mu,\xi;x)$ is the specific intensity for some particular value of μ, $0 < \mu \leq 1$, and ξ is a scaled frequency variable, $\xi = (\nu-\nu_o)/\Delta_s$, at depth x. The terms $\Sigma^+(y,x)$, $\Sigma^-(x,y)$ represent the emission of radiative energy within the layer [x,y] in the positive (negative) direction at y(x). Finally, the quantities t(y,x), t(x,y), r(y,x), r(x,y) represent operators that describe the transmission and reflection of radiation by the layer. In line transfer problems they will involve, in general, redistribution of radiant energy both in direction and in frequency. Thus, for example, we may write

$$[r(x,y)u^-(y)](\mu,\xi) \equiv$$

$$\frac{1}{2\mu}\int_0^1 d\mu' \int_{-\infty}^{\infty} d\xi'\, R(x,y|\mu,\xi;-\mu',\xi')[u^-(y)(-\mu',\xi')] \qquad (2)$$

for the radiation diffusely reflected from negative directions into the positive direction μ at the surface y of the layer [x,y]. This representation is intractable for computation as it stands, but if, as is common, in practice we use a discrete-ordinate approximation both with regard to μ and to ξ, the intensities $u^+(x)$, $u^-(x)$, and sources $\Sigma^+(y,x)$, $\Sigma^-(x,y)$ can be written as vectors in a finite-dimensional linear vector space, and the quantities t(x,y), t(y,x), r(x,y), r(y,x) become matrix operators mapping this space onto itself. We can then apply familiar computational techniques to deal with equations (1).

Suppose, for the moment, that we know the operators and source vectors for each layer of an arbitrary finite partition of [a,b] into N layers by planes at $x = x_n$, $a = x_1 < x_2 < \ldots < x_{N+1} = b$. Equations (1) then provide a set of 2N equations for the 2N intensity vectors u_n^-, $1 \leq n \leq N$ and u_n^+, $2 \leq n \leq N+1$ in terms of the boundary values $u_1^+ = u^+(a)$, $u_{N+1}^- = u^-(b)$ and internal sources. We may write these equations in the form

https://doi.org/10.1017/S0252921100151267 Published online by Cambridge University Press

$$(\underset{\sim}{I} - \underset{\sim}{L})\ \underset{\sim}{U} = \underset{\sim}{\Sigma} + \underset{\sim}{\Sigma}_s \tag{3}$$

where the vectors $\underset{\sim}{U}$, $\underset{\sim}{\Sigma}$ and $\underset{\sim}{\Sigma}_s$ are defined by

$$\underset{\sim}{U} = \begin{bmatrix} u_2^+ \\ u_1^- \\ u_3^+ \\ u_2^- \\ \cdot \\ \cdot \\ \cdot \\ u_{N+1}^+ \\ u_N^- \end{bmatrix}, \quad \underset{\sim}{\Sigma} = \begin{bmatrix} \Sigma^+(2,1) \\ \Sigma^-(1,2) \\ \Sigma^+(3,2) \\ \Sigma^-(2,3) \\ \cdot \\ \cdot \\ \cdot \\ \Sigma^+(N+1,N) \\ \Sigma^-(N,N+1) \end{bmatrix}, \quad \underset{\sim}{\Sigma}_s = \begin{bmatrix} t(2,1)u_1^+ \\ r(2,1)u_1^+ \\ 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ r(N,N+1)u_{N+1}^- \\ t(N,N+1)U_{N+1}^- \end{bmatrix} \tag{4}$$

and where the matrix $\underset{\sim}{L}$ has the block form

$$\underset{\sim}{L} = \left[\begin{array}{cc|cc|cc|cc} o & o & o & r(1,2) & & & & \\ o & o & o & t(1,2) & & & & \\ \hline t(3,2) & o & o & o & o & r(2,3) & & \\ r(3,2) & o & o & o & o & t(2,3) & & \\ \hline & & t(4,3) & o & o & o & o & r(3,4) \\ & & r(4,3) & o & o & o & o & t(3,4) \\ \hline & & & & r(N+1,N) & o & o & o \\ & & & & t(N+1,N) & o & o & o \end{array}\right] \tag{5}$$

https://doi.org/10.1017/S0252921100151267 Published online by Cambridge University Press

We recognize that this is a particular form of block-tridiagonal matrix; the method of dealing with equations involving such matrices is well-known, and a good account is given in the text book by Isaacson and Keller.[8] When this process is written out, we get an algorithm which may be set out as follows:

Define $r(1,n+1)$, $V^{+}_{n+1/2}$, $V^{-}_{n+1/2}$, $n = 1,2, \ldots .$, N by the equations

$$r(1,1) = 0,\ r(1,n+1) = r(n,n+1) +$$
$$+ t(n+1,n)r(1,n)[I-r(n+1,n)r(1,n)]^{-1}t(n,n+1) \quad (6)$$

$$V^{+}_{1/2} = u^{+}_{1},\quad V^{+}_{n+1/2} =$$
$$\hat{t}(n+1,n)V^{+}_{n-1/2} + \Sigma^{+}(n+1,n) + R_{n+1/2}\Sigma^{-}(n,n+1) \quad (7)$$

$$V^{-}_{n+1/2} = \hat{r}(n+1,n)V^{+}_{n-1/2} + T_{n+1/2}\Sigma^{-}(n,n+1)$$

where

$$T_{n+1/2} = [I-r(n+1,n)r(1,n)]^{-1} \equiv$$
$$I+r(n+1,n)[I-r(1,n)r(n+1,n)]^{-1}r(1,n)$$
$$\hat{t}(n+1,n) = t(n+1,n)[I-r(1,n)r(n+1,n)]^{-1} \quad (8)$$
$$\hat{r}(n+1,n) = r(n+1,n)[I-r(1,n)r(n+1,n)]^{-1}$$

and

$$R_{n+1/2} = \hat{t}(n+1,n)r(1,n)$$

These equations can be solved for increasing values

https://doi.org/10.1017/S0252921100151267 Published online by Cambridge University Press

of n, and the final solutions u^+_{n+1}, u^-_n can be obtained from

$$\begin{aligned} u^+_{n+1} &= r(1,n+1)u^-_{n+1} + V^+_{n+1/2} \\ u^-_n &= \hat{t}(n,n+1)u^-_{n+1} + V^-_{n+1/2} , \end{aligned} \tag{9}$$

starting with the specified boundary value of u^-_{N+1}, for n = N, N-1, ... , 1 in succession. The matrix

$$\hat{t}(n,n+1) = T_{n+1/2}\, t(n,n+1)$$

may be computed at the same time as those of equations (8). More information about these algorithms and their physical interpretation is contained in our earlier papers.[6,7]

Our strategy is therefore to put the equations of the problem into this canonical form, and thence to solve them by the standard technique outlined above. One way of doing this is to remark that in the limit as y tends to x, equations (1) must converge to the conventional differential equation of radiative transfer. In this way, we may identify the leading terms in an expansion of the matrix operators in terms of the layer thickness (y-x); they can be expressed in terms of the usual known physical quantities point-by-point, such as absorption coefficients, scattering coefficients and so on. The source terms may be identified in the same way. The approximations so defined have rather poor truncation errors, and it is better to start with a difference scheme which is known to have higher accuracy (6). That is the approach that we shall adopt in this paper. Of course, the operators and sources obtained in this way must have expansion in powers of the layer thickness agreeing with those obtained from the simpler procedure through first order terms.

3. DISCRETE EQUATIONS FOR LINE TRANSFER PROBLEMS

The purpose of this section is to write the integro-differential equations of transfer for the

https://doi.org/10.1017/S0252921100151267 Published online by Cambridge University Press

formation of spectrum lines in the required form, using a notation rather similar to that of Hummer and Rybicki.[2,3] For a two-level atom, assuming complete redistribution, they write

$$\mu \frac{d}{d\tau} I_{\xi\mu} = [\beta+\phi(\xi)](S_{\xi\mu}-I_{\xi}) \tag{10}$$

where $I_{\xi\mu}$ is the specific intensity at a frequency specified by

$$\xi = (\nu-\nu_o)/\Delta_s, \tag{11}$$

in a direction specified by μ where Δ_s is, as usual, some standard frequency interval. Here $\phi(\xi)$ is the line profile, normalized so that

$$\int_{-\infty}^{\infty} \phi(\xi)d\xi = 1. \tag{12}$$

The mean optical depth is defined in terms of the geometrical depth z by

$$d\tau = k_L(z)dz \frac{h\nu_o}{4\pi\Delta_s}(N_1B_{12}-N_2B_{21})dz \tag{13}$$

where $B_{12}(z)$, $B_{21}(z)$ are the Einstein coefficients, and $N_1(z)$,$N_2(z)$ the population densities of the lower and upper states respectively. The quantity β is the ratio k_C/k_L of opacity due to continuous absorption per unit interval Δ_s to that in the line, and in this event

$$S_{\xi}(z) = \frac{\phi(\xi)}{\beta+\phi(\xi)} S_L(z) + \frac{\beta}{\beta+\phi(\xi)} S_c(z) \tag{14}$$

where $S_C(z)$ is the continuum source function, which, for present purposes, we shall write in the form

$$S_c(z) = \rho(z)\ B(\nu_o,T_e(z)) \tag{15}$$

where B is the Planck function for frequency ν_o at

https://doi.org/10.1017/S0252921100151267 Published online by Cambridge University Press

electron temperature T_e, and both ρ and B are assumed to be independent specified functions of z. The line source function for complete redistribution, $S_L(z)$, is given by

$$S_L(z) = A_{21}N_2(z)/[B_{12}N_1(z) - B_{21}N_2(z)], \quad (16)$$

and, when combined with the statistical equilibrium equation

$$N_1[B_{12}\int_{-\infty}^{\infty}\phi(\xi)J_\xi d\xi + C_{12}] = N_2[A_{21} + C_{21} + B_{21}\int_{-\infty}^{\infty}\phi(\xi)J_\xi d\xi],$$

it becomes

$$S_L(z) = (1-\varepsilon)\int_{-\infty}^{\infty}\phi(\xi)J_\xi d\xi + \varepsilon B \quad (17)$$

where J_ξ is the mean intensity, B is the Planck function and

$$\varepsilon = \frac{C_{21}}{C_{21}+A_{21}[1-\exp(-h\nu_o/kT_e)]^{-1}} \quad (18)$$

is the probability per scatter that a photon will be destroyed by collisional de-excitation. We are, as usual, interested in cases with $\varepsilon << 1$, and we shall assume that this parameter is specified in advance.

Our computational problem is therefore to solve equation (10) with a source function defined by equations (14), (15), and (17) regarding $\beta(z)$, $\rho(z)$, $\varepsilon(z)$, $\phi(\xi) \equiv \phi(z,\xi)$, and B(z) as specified functions. In addition we need boundary conditions specifying the fluxes incident on the external surfaces. This problem will be solved by a suitable discretization in frequency, direction, and space. For the frequency discretization, we choose division points

https://doi.org/10.1017/S0252921100151267 Published online by Cambridge University Press

$\{\xi_i\}$ and weights $\{a_i\}$ such that

$$\int_{-\infty}^{\infty} \phi(\xi) f(\xi) d\xi \approx \sum_{i=-I}^{I} a_i f(\xi_i) \qquad \sum_{i=-I}^{I} a_i = 1 \tag{19}$$

and for the directional discretization we choose abscissae $\{\mu_j\}$ and weights $\{b_j\}$ such that

$$\int_0^1 f(\mu) d\mu \approx \sum_{j=1}^{m} b_j f(\mu_j), \quad \sum_{j=1}^{m} b_j = 1 \ . \tag{20}$$

We define the mxm matrices

$$\underset{\sim}{b} = [b_i \delta_{ij}] \ , \ \underset{\sim}{M}_m = [\mu_i \delta_{ij}]$$

and the m-vectors

$$\underset{\sim}{h} = \begin{bmatrix} 1 \\ 1 \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{bmatrix} \ , \ u^+_{i,n} = \begin{bmatrix} I(\mu_1, \xi_i; \tau_n) \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ I(\mu_m, \xi_i; \tau_n) \end{bmatrix}$$

$$u^-_{i,n} = \begin{bmatrix} I(-\mu_1, \xi_i; \tau_n) \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ I(-\mu_m, \xi_i; \tau_n) \end{bmatrix} \tag{21}$$

https://doi.org/10.1017/S0252921100151267 Published online by Cambridge University Press

Then, if we integrate equation (10) from τ_n to τ_{n+1}, we may write the result as

$$\underset{\sim}{M}_m(u^+_{i,n+1}-u^+_{i,n}) = \tau_{n+1/2}(\rho\beta+\varepsilon\phi_i)_{n+1/2}B_{n+1/2}\underset{\sim}{h}$$

$$+ \frac{1}{2}\tau_{n+1/2}\sigma_{n+1/2}\phi_{i,n+1/2} \times$$

$$\times \sum_{i'=-I}^{I} a_{i',n+1/2}(\underset{\sim}{h}\,\underset{\sim}{h}^T)\underset{\sim}{b}(u^+ + u^-)_{i',n+1/2}$$

$$- \tau_{n+1/2}(\beta+\phi_i)_{n+1/2}u^+_{i,n+1/2} \qquad (22)$$

$$-\underset{\sim}{M}_m(u^-_{i,n+1}-u^-_{i,n}) = \tau_{n+1/2}(\rho\beta+\varepsilon\phi_i)_{u+1/2}B_{n+1/2}\underset{\sim}{h}$$

$$+ \frac{1}{2}\tau_{n+1/2}\sigma_{n+1/2}\phi_{i,n+1/2} \times$$

$$\sum_{i'=-I}^{I} a_{i',n+1/2}(\underset{\sim}{h}\,\underset{\sim}{h}^T)\underset{\sim}{b}(u^+ + u^-)_{i',n+1/2}$$

$$- \tau_{n+1/2}(\beta+\phi_i)_{n+1/2}u^-_{i,n+1/2}$$

Here we have used superscript T to denote transposition, $\tau_{n+1/2} = \tau_{n+1} - \tau_n > 0$, and other subscripts $(n+1/2)$ to denote averages for the layer. Also $\sigma_{n+1/2} = 1-\varepsilon_{n+1/2}$.

Corresponding to each pair of subscripts (i,j), we now define an index k so that

$$(i,j) \equiv k = j + (i-1)m, \quad 1 \le k \le K = mI. \qquad (23)$$

https://doi.org/10.1017/S0252921100151267 Published online by Cambridge University Press

For each value of k, we define coefficients $W_k \equiv W_{k,n+1/2}$ by

$$\phi_k W_k = a_i b_j \quad . \tag{24}$$

Let

$$\underset{\sim}{u}_n^{\pm}(L) = \begin{bmatrix} u_{-1,n}^{\pm} \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ u_{-I,n}^{\pm} \end{bmatrix} , \quad \underset{\sim}{u}_n^{+}(R) = \begin{bmatrix} u_{1,n}^{\pm} \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ u_{I,n}^{\pm} \end{bmatrix} \tag{25}$$

be K-vectors corresponding to frequencies to left (L) and right (R) of the line center. Let $\underset{\sim}{\phi}_{n+1/2}$ and $\underset{\sim}{g}_{n+1/2}$ be K-vectors, where the elements of $\underset{\sim}{g}_{n+1/2}$ are defined by

$$g_{k,n+1/2} = (\rho\beta+\varepsilon\phi_k)_{n+1/2} B_{n+1/2} \tag{26}$$

and let $\underset{\sim}{\Lambda}$, $\underset{\sim}{W}$ and $\underset{\sim}{M}$ be K × K matrices defined by

$$\underset{\sim}{\Lambda}_{n+1/2} = [\Lambda_{kk'}] =$$

$$[(\beta+\phi_k)_{n+1/2}\delta_{kk'}] \ , \ \underset{\sim}{W}_{n+1/2} = [W_{k,n+1/2}\delta_{kk'}], \tag{27}$$

and

$$\underset{\sim}{M} = \begin{bmatrix} \underset{\sim}{M}_m & & \bigcirc \\ & \underset{\sim}{M}_m & \\ & & \ddots \\ \bigcirc & & \underset{\sim}{M}_m \end{bmatrix}$$

https://doi.org/10.1017/S0252921100151267 Published online by Cambridge University Press

Then equations (22) reduce to

$$\underset{\sim}{M}[\underset{\sim}{u}^{+}_{n+1}(r)-\underset{\sim}{u}^{+}_{n}(R)] + \tau_{n+1/2}\underset{\sim}{\Lambda}_{n+1/2}\underset{\sim}{u}^{+}_{n+1/2}(R)$$

$$= \frac{\pi}{2}_{n+1/2}\sigma_{n+1/2}(\underset{\sim}{\phi}\ \underset{\sim}{\phi}^{T})_{n+1/2}\underset{\sim}{W}_{n+1/2}[\underset{\sim}{u}^{+}(L) +$$

$$+ \underset{\sim}{u}^{+}(R)+\underset{\sim}{u}^{-}(L)+\underset{\sim}{u}^{-}(R)]_{n+1/2}$$

$$+ \tau_{n+1/2}\underset{\sim}{g}_{n+1/2}(R)$$

and (28)

$$-\underset{\sim}{M}[\underset{\sim}{u}^{-}_{n+1}(R)-\underset{\sim}{u}^{-}_{n}(R)] + \tau_{n+1/2}\underset{\sim}{\Lambda}_{n+1/2}\underset{\sim}{u}^{-}_{n+1/2}(R)$$

$$= \frac{\pi}{2}_{n+1/2}\sigma_{n+1/2}(\underset{\sim}{\phi}\ \underset{\sim}{\phi}^{T})_{n+1/2}\underset{\sim}{W}_{n+1/2}[\underset{\sim}{u}^{+}(L) +$$

$$+ \underset{\sim}{u}^{+}(R)+\underset{\sim}{u}^{-}(L)+\underset{\sim}{u}^{-}(R)]_{n+1/2}$$

$$+ \tau_{n+1/2}\underset{\sim}{g}_{n+1/2}(R)$$

and two similar equations with L and R interchanged everywhere. When the line profile is symmetric, we have $\underset{\sim}{u}^{\pm}(L) = \underset{\sim}{u}^{\pm}(R)$, $\underset{\sim}{g}(L) = \underset{\sim}{g}(R)$, and equations (24) simplify. In either case, the structure of the equations is similar to those of equations (2.10) of our earlier paper,[6] and hence the same procedures may be used to re-write them in the form of equations (1). The details are presented in the Appendix. In the symmetric case, to which we shall confine ourselves in this paper, we find

$$\underset{\sim}{u}^{+}_{n+1} = \underset{\sim}{t}(n+1,n)\underset{\sim}{u}^{+}_{n} + \underset{\sim}{r}(n,n+1)\underset{\sim}{u}^{-}_{n+1} + \underset{\sim}{\Sigma}^{+}(n+1,n)$$

$$\underset{\sim}{u}^{-}_{n} = \underset{\sim}{r}(n+1,n)\underset{\sim}{u}^{+}_{n} + \underset{\sim}{t}(n,n+1)\underset{\sim}{u}^{-}_{n+1} + \underset{\sim}{\Sigma}^{-}(n,n+1) \quad (29)$$

https://doi.org/10.1017/S0252921100151267 Published online by Cambridge University Press

where

$$\underset{\sim}{r}(n,n+1) = \underset{\sim}{r}(n+1,n) \equiv$$

$$\underset{\sim}{r}_{n+1/2} = \gamma_{n+1/2} (\underset{\sim}{C}\ \underset{\sim}{C}^{T})_{n+1/2} \underset{\sim}{M}\ \underset{\sim}{W}_{n+1/2} \quad (30)$$

and

$$\underset{\sim}{t}(n,n+1) = \underset{\sim}{t}(n+1,n) = \underset{\sim}{t}_{n+1/2} + \underset{\sim}{r}_{n+1/2} \quad (31)$$

in which the first term is a diagonal matrix,

$$\underset{\sim}{t}_{n+1/2} = [\underset{\sim}{I} - \tfrac{1}{2}\tau_{n+1/2} \underset{\sim}{M}^{-1} \underset{\sim}{\Lambda}_{n+1/2}] [\underset{\sim}{I} + \tfrac{1}{2}\tau_{n+1/2} \underset{\sim}{M}^{-1} \underset{\sim}{\Lambda}_{n+1/2}]^{-1}.$$

In these definitions $\underset{\sim}{I}$ is the K × K identity matrix,

$$\underset{\sim}{C}_{n+1/2} = [\underset{\sim}{I} + \tfrac{1}{2}\tau_{n+1/2} \underset{\sim}{M}^{-1} \underset{\sim}{\Lambda}_{n+1/2}]^{-1} \cdot \underset{\sim}{M}^{-1} \underset{\sim}{\phi}_{n+1/2}, \quad (32)$$

and

$$\gamma_{n+1/2} = \sigma_{n+1/2} \tau_{n+1/2} / [1 - 2\delta_{n+1/2}] \quad (33)$$

where

$$\delta_{n+1/2} = \sigma_{n+1/2} \tau_{n+1/2} \times$$

$$\times \sum_{k=1}^{K} \frac{W_{k,n+1/2} \phi^2_{k,n+1/2}}{2\mu_k + \tau_{n+1/2} \Lambda_{kk,n+1/2}} \quad (34)$$

The following properties of the matrices are easy but important consequences of the definitions:

(i) $\underset{\sim}{t}_{n+1/2}$ is a matrix approximation to $\exp [-\tau_{n+1/2} \underset{\sim}{M}^{-1} \underset{\sim}{\Lambda}_{n+1/2}]$, and so represents the unscattered attenuation in the layer.

https://doi.org/10.1017/S0252921100151267 Published online by Cambridge University Press

(ii) $\underset{\sim}{r}_{n+1/2}$ is unsymmetric, but the matrix $\underset{\sim}{S}_{n+1/2}$ defined by

$$\underset{\sim}{S}_{n+1/2} = \underset{\sim}{W}_{n+1/2}\underset{\sim}{Mr}_{n+1/2} \tag{35}$$

(iii) $\underset{\sim}{r}_{n+1/2}$ has non-negative elements (written $\underset{\sim}{r}_{n+1/2} \geq o$). Indeed, from (34), taking account of (24) and (27), we have

$$\delta_{n+1/2} = \sigma_{n+1/2}\tau_{n+1/2}$$

$$\sum_{i=1}^{I}\sum_{j=1}^{M} \frac{a_j b_j \phi_{i,n+1/2}}{2\mu_j + \tau_{n+1/2}(\beta_{n+1/2} + \phi_{i,n+1/2})}$$

$$< \sigma_{n+1/2} \sum_{i=1}^{I}\sum_{j=1}^{M} a_i b_j = \frac{1}{2}\sigma_{n+1/2}$$

so that

$$o < \gamma_{n+1/2} < \frac{\sigma_{n+1/2}\tau_{n+1/2}}{1-\sigma_{n+1/2}}$$

(iv) A sufficient condition for $t_{n+1/2} \geq o$ is

$$\tau_{n+1/2} < \min_k \frac{2\mu_k}{\Lambda_k} = \frac{2\mu_1}{\beta + \phi(0)_{n+1/2}} = \tau^1,$$

say. This is more stringent than one needs to ensure that t(n,n+1) and t(n+1,n) are non-negative. A better sufficient condition is

$$\tau < \tau^1 \min_{i,j}\left[1 - \frac{2(1-\varepsilon)a_i\phi_{i,n+1/2}b_j}{(\beta+\phi_i)_{n+1/2}}\right]^{-1} = \tau_{crit} \tag{36}$$

https://doi.org/10.1017/S0252921100151267 Published online by Cambridge University Press

(v) Let

$$||\underset{\sim}{u}|| = 2\pi \sum_{i=-I}^{I} a_i \sum_{j=1}^{M} b_j u_j \, |u_i(\mu_j)| \tag{37}$$

$$= 2\pi \sum_{k=1}^{K} \underset{\phi_k}{W_k} \, |u_k|$$

Then $||\underset{\sim}{u}||$ is a vector norm satisfying the usual norm conditions, and a matrix norm, $||\underset{\sim}{A}||$, consistent and subordinate to this norm is[7]

$$||A|| = \max_{k'} \sum_{k} \underset{\phi_k}{W_k} \, |A_{kk'}| \, \underset{\phi_{k'}}{W_{k'}^{-1}} \quad . \tag{38}$$

Now define the cell matrix $\underset{\sim}{\Gamma}(n,n+1)$ associated with the layer under consideration by

$$\underset{\sim}{\Gamma}(n,n+1) = \begin{bmatrix} \underset{\sim}{t}(n+1,n) & \underset{\sim}{r}(n,n+1) \\ \underset{\sim}{r}(n+1,n) & \underset{\sim}{t}(n,n+1) \end{bmatrix}$$

By analogy with (38), and assuming (36) holds,

$$||\underset{\sim}{\Gamma}(n,n+1)|| = \max \{ ||\underset{\sim}{t}(n+1,n) +$$

$$+\underset{\sim}{r}(n+1,n)||, \; ||\underset{\sim}{t}(n,n+1)+\underset{\sim}{r}(n,n+1)|| \}$$

$$= ||\underset{\sim}{t}_{n+1/2} + 2\underset{\sim}{r}_{n+1/2}||$$

$$= 1-\tau_{n+1/2} \min_{k} \frac{\{(\beta+\varepsilon\phi_k)_{n+1/2}\}}{\mu_k} + \tag{40}$$

$$+ O(\tau^2_{n+1/2})$$

$$< 1.$$

https://doi.org/10.1017/S0252921100151267 Published online by Cambridge University Press

We have shown[7] that the conditions set out above are necessary and sufficient to ensure the existence of a unique bounded non-negative solution to our equations.

To complete the specification, we need the source vectors, which are given by

$$\Sigma^{+}(n+1,n) = \underset{\sim}{\Sigma}^{-}(n,n+1) \equiv$$

$$\underset{\sim}{\Sigma}_{n+1/2} = \frac{1}{2}\tau_{n+1/2}[\underset{\sim}{I} + \underset{\sim}{t}_{n+1/2} + 2\underset{\sim}{r}_{n+1/2}]\ \underset{\sim}{M}^{-1}\underset{\sim}{g}_{n+1/2}$$

4. IMPLEMENTATION OF THE PROCEDURES FOR LINE TRANSFER

The structure of equations (30) and (31), which represents the reflection operators for a thin layer as a Kronecker product, allows a great simplification of the algorithm. We suppose that we require only to determine the source function and the emergent intensity vector, so that we shall set down only those equations which are directly relevant to the computation in this section. The remaining equations not needed for this part of the computation are set out in the Appendix.

We write

$$\underset{\sim}{S}_{n+1/2} = \underset{\sim}{M}\ \underset{\sim}{W}_{n+1/2}\underset{\sim}{r}(1,n)\ , \tag{42}$$

and define the scalars

$$\lambda_{n+1/2} = \underset{\sim}{C}^{T}_{n+1/2}\underset{\sim}{S}_{n+1/2}\underset{\sim}{C}_{n+1/2}$$

and

$$A_{n+1/2} = \frac{\gamma_{n+1/2}}{1-\gamma_{n+1/2}\lambda_{n+1/2}} \tag{43}$$

Then equations (6) become

$$\underset{\sim}{S}_{n+3/2} =$$

$$\underset{\sim}{W}^{-1}_{n+1/2}\underset{\sim}{W}_{n+3/2}[\underset{\sim}{t}_{n+1/2}\underset{\sim}{S}_{n+1/2}\underset{\sim}{t}_{n+1/2} + A_{n+1/2}\underset{\sim}{X}_{n+1/2}\tilde{\underset{\sim}{X}}_{n+1/2}], \tag{44}$$

https://doi.org/10.1017/S0252921100151267 Published online by Cambridge University Press

to be evaluated recursively for n = 1,2, ··· , N, starting with $\underset{\sim}{S}_{3/2} = 0$, where the K-vectors $\underset{\sim}{X}_{n+1/2}$ and $\tilde{\underset{\sim}{X}}_{n+1/2}$ are defined by

$$\underset{\sim}{X}_{n+1/2} = (\underset{\sim}{M}\, W_{n+1/2} + \underset{\sim}{t}_{n+1/2}\underset{\sim}{S}_{n+1/2})\underset{\sim}{C}_{n+1/2}$$

$$\tilde{\underset{\sim}{X}}_{n+1/2} = \underset{\sim}{C}^{T}_{n+1/2}(\underset{\sim}{S}_{n+1/2}\underset{\sim}{t}_{n+1/2} + \underset{\sim}{M}\, \underset{\sim}{W}_{n+1/2}).$$

Notice that $\tilde{\underset{\sim}{X}}_{n+1/2}$ will equal $\underset{\sim}{X}^{T}_{n+1/2}$ if $S_{n+1/2}$ is symmetric. This is only the case if $\underset{\sim}{W}_{n+1/2}$ is independent of depth, that is to say if (a_i/ϕ_k) is independent of n. This depends on the choice of quadrature forumla; $\underset{\sim}{W}_{n+1/2}$ is independent of n for most choices.

We also define

$$\tilde{\underset{\sim}{V}}_{n+1/2} = \underset{\sim}{M}\, \underset{\sim}{W}_{n+1/2}\underset{\sim}{V}^{+}_{n+1/2} \tag{45}$$

and the K-vector

$$\underset{\sim}{Y}_{n+1/2} = \underset{\sim}{W}^{-1}_{n-1/2}\underset{\sim}{W}_{n+1/2}\tilde{\underset{\sim}{V}}_{n-1/2} + \underset{\sim}{S}_{n+1/2}\underset{\sim}{\Sigma}_{n+1/2} \quad . \tag{46}$$

Then the first of equations (7) becomes

$$\tilde{\underset{\sim}{V}}_{n+1/2} = \underset{\sim}{t}_{n+1/2}\underset{\sim}{Y}_{n+1/2} + B_{n+1/2}\underset{\sim}{X}_{n+1/2} + \underset{\sim}{M}\, \underset{\sim}{W}_{n+1/2}\underset{\sim}{\Sigma}_{n+1/2}, \tag{47}$$

to be evaluated recursively for n = 1,2, ··· ,N, starting with

$\tilde{\underset{\sim}{V}}_{1/2} = \underset{\sim}{M}\, \underset{\sim}{W}_{1/2}\underset{\sim}{u}^{+}_{1}$ ($\underset{\sim}{W}_{1/2} = \underset{\sim}{W}_{3/2}$ being assumed), where

$$B_{n+1/2} = A_{n+1/2}\underset{\sim}{C}^{T}_{n+1/2}\underset{\sim}{Y}_{n+1/2} \quad . \tag{48}$$

https://doi.org/10.1017/S0252921100151267 Published online by Cambridge University Press

Then, given $\underset{\sim}{u}^-_{N+1}$, we can evaluate for n=N, N-1, ...,1 the equation

$$\underset{\sim}{u}^-_n = \underset{\sim}{t}_{n+1/2}\,\underset{\sim}{u}^-_{n+1} + \left[B_{n+1/2} + A_{n+1/2}\left(\tilde{\underset{\sim}{X}}_{n+1/2}\,\underset{\sim}{u}^-_{n+1}\right)\right]\underset{\sim}{C}_{n+1/2} + \underset{\sim}{\Sigma}_{n+1/2} \qquad (49)$$

to give the upward directed intensities at each interface. The line source function, defined by (17), requires the calculation of

$$\bar{J}_n = \int_{-\infty}^{\infty} \phi(\xi)\, J_{\xi,n}\, d\xi$$

$$\approx \left[\underset{\sim}{\phi}^T_{n-1/2}\underset{\sim}{W}_{n-1/2}\underset{\sim}{u}^+_n + \underset{\sim}{\phi}^T_{n+1/2}\underset{\sim}{W}_{n+1/2}\underset{\sim}{u}^-_n\right] \qquad (50)$$

$$= \underset{\sim}{Z}^T_n\,\underset{\sim}{u}^-_n + \underset{\sim}{\phi}^T_{n-1/2}\,\underset{\sim}{M}^{-1}\,\tilde{\underset{\sim}{V}}_{n-1/2} \quad ,$$

where

$$\underset{\sim}{Z}^T_n = \underset{\sim}{\phi}^T_{n=1/2}\underset{\sim}{M}^{-1}\left(\underset{\sim}{W}^{-1}_{n+1/2}\underset{\sim}{W}_{n-1/2}\underset{\sim}{S}_{n+1/2}\right) + \underset{\sim}{\phi}^T_{n+1/2}\underset{\sim}{W}_{n+1/2} \qquad (51)$$

may be computed along with $\underset{\sim}{X}_{n+1/2}$, $\tilde{\underset{\sim}{X}}_{n+1/2}$ and $\underset{\sim}{Y}_{n+1/2}$. Notice that if $\underset{\sim}{u}^-_1$ and the line source function are all that we require, it is unnecessary to compute $\underset{\sim}{u}^+_n$ directly.

We may also wish to compute the emission in some other direction μ than that prescribed by the rule for angular quadrature. Now that we have the line-source function this is straightforward. Writing the required frequency distribution as an I-vector, $\underset{\sim}{u}^-_1(\mu)$, we see that

$$\underset{\sim}{u}^-_1(\mu) = \int_0^{\tau_{N+1}} e^{-\tau\underset{\sim}{\Lambda}(\tau)/\mu}\left[(1-\varepsilon(\tau))\,\underset{\sim}{\phi}_\mu(\tau)\bar{J}(\tau) + \underset{\sim}{g}_\mu(\tau)\right]d\tau \qquad (52)$$

https://doi.org/10.1017/S0252921100151267 Published online by Cambridge University Press

where $\underset{\sim}{\Lambda}(\tau)$ is an I × I diagonal matrix, and $\underset{\sim}{\phi}_\mu(\tau)$, $\underset{\sim}{g}_\mu(\tau)$ are I-vectors defined for the single value of μ as in equations (27) and (26) respectively. The discrete representation of (52) may be written

$$\underset{\sim}{\bar{u}_1}(\mu) \approx \sum_{n=1}^{N} \underset{\sim}{\bar{t}}(1,n+\tfrac{1}{2}) [(1-\epsilon_{n+1/2})\bar{J}_{n+1/2}\underset{\sim}{\phi}_{n+1/2} + \underset{\sim}{g}_{n+1/2}] \tau_{n+1/2} \tag{53}$$

The choice of $\bar{J}_{n+1/2}$ and of $\underset{\sim}{t}(1,n+\frac{1}{2})$ is slightly arbitrary. We have used

$$\underset{\sim}{\bar{t}}(1,n+\tfrac{1}{2}) = \tfrac{1}{2}[\underset{\sim}{\bar{t}}(1,n) + \underset{\sim}{\bar{t}}(1,n+1)]$$

where

$$\underset{\sim}{\bar{t}}(1,n) = \prod_{k=1}^{n-1} \underset{\sim}{t}_{n+1/2}(\mu) \ , \tag{54}$$

with

$$\underset{\sim}{t}_{n+1/2}(\mu) = \left[\delta_{ii'} \frac{1 - \frac{1}{2}\tau_{n+1/2}\Lambda_i/\mu}{1 + \frac{1}{2}\tau_{n+1/2}\Lambda_i/\mu}\right] ,$$

and

$$\bar{J}_{n+1/2} = \tfrac{1}{2}(\bar{J}_{n+1} + \bar{J}_n) \ . \tag{55}$$

It would be possible to use more accurate formulae, but we have not found it necessary to do so.

5. THE RICCATI METHOD

If we pass to the limit as $\tau_{n+1/2} \to 0$, our equations reduce to a set equivalent to those of the

https://doi.org/10.1017/S0252921100151267 Published online by Cambridge University Press

Riccati method of Rybicki and Hummer.[2,3] Equation (44) becomes

$$\frac{d\underset{\sim}{S}(t)}{d\tau} = \underset{\sim}{W}^{-1}(\tau)\ \frac{d\underset{\sim}{W}(\tau)}{d\tau}\ \underset{\sim}{S}(\tau) - \underset{\sim}{M}^{-1}\underset{\sim}{\Lambda}(\tau)\underset{\sim}{S}(\tau) +$$

$$- \underset{\sim}{S}(\tau)\underset{\sim}{M}^{-1}\underset{\sim}{\Lambda}(\tau) + \sigma(\tau)\underset{\sim}{X}(\tau)\tilde{\underset{\sim}{X}}(\tau) \qquad (56)$$

with initial condition $\underset{\sim}{S}(o) = 0$,

$$\underset{\sim}{X}(\tau) = [\underset{\sim}{W}(\tau) + \underset{\sim}{S}(\tau)\underset{\sim}{M}^{-1}]\ \underset{\sim}{\phi}(\tau)$$

$$\tilde{\underset{\sim}{X}}(\tau) = \underset{\sim}{\phi}^{T}(\tau)\ [\underset{\sim}{M}^{-1}\underset{\sim}{S}(\tau) + \underset{\sim}{W}(\tau)]\ .$$

also

$$\frac{d\tilde{\underset{\sim}{V}}(\tau)}{d\tau} = \underset{\sim}{W}^{-1}(\tau)\ \frac{d\underset{\sim}{W}(\tau)}{d\tau}\ \tilde{\underset{\sim}{V}}(\tau) - [\underset{\sim}{\Lambda}(\tau) + \qquad (57)$$

$$- \sigma(\tau)\underset{\sim}{X}(\tau)\underset{\sim}{\phi}^{T}(\tau)]\underset{\sim}{M}^{-1}\tilde{\underset{\sim}{V}}(\tau) + (\underset{\sim}{W}(\tau) + \underset{\sim}{S}(\tau)\underset{\sim}{M}^{-1})\underset{\sim}{g}(\tau)$$

corresponding to (47), with $\tilde{V}(o) = MW(o)\ u^{+}(o)$, and

$$\frac{d\underset{\sim}{u}^{-}(\tau)}{d\tau} = \underset{\sim}{M}^{-1}[\underset{\sim}{\Lambda}(\tau) - \sigma(\tau)\underset{\sim}{\phi}(\tau)\underset{\sim}{X}(\tau)]\ \underset{\sim}{u}^{-}(\tau) +$$

$$- \sigma(\tau)\underset{\sim}{M}^{-1}\underset{\sim}{\phi}(\tau)\underset{\sim}{\phi}^{T}(\tau)\underset{\sim}{M}^{-1}\tilde{\underset{\sim}{V}}(\tau) - \underset{\sim}{M}^{-1}\underset{\sim}{g}(\tau) \qquad (58)$$

while (50) and (51) give

$$\bar{J}(\tau) = \tilde{\underset{\sim}{X}}(\tau)\ \underset{\sim}{u}^{-}(\tau) + \underset{\sim}{\phi}^{T}(\tau)\underset{\sim}{M}^{-1}\tilde{\underset{\sim}{V}}(\tau). \qquad (59)$$

Apart from superficial differences arising from a difference of notation, the set (56) - (59) differ from those of Rybicki and Hummer by terms involving $\underset{\sim}{W}^{-1}\ \frac{d\underset{\sim}{W}}{d\tau}$. The matrix $\underset{\sim}{W}$ is defined by (24). In most applications, the weights for the frequency integra-

https://doi.org/10.1017/S0252921100151267 Published online by Cambridge University Press

tion, a_i, have been chosen proportional to the corresponding value of ϕ. Thus the element

$$W_k = \frac{a_i b_j}{\phi_k}$$

in such cases is independent of τ, and the terms involving $\underset{\sim}{W}^{-1} \frac{d\underset{\sim}{W}}{d\tau}$ disappear. This has a number of useful consequences both for the discrete and the Riccati methods; in particular the operator $\underset{\sim}{S}$ becomes symmetric, a fact which can be used to reduce both the work and the storage required. Other aspects of the choice of quadrature formulae will be discussed in the following section.

We can also solve equations similar to (52) to determine the emissions in directions other than those used in the quadrature formula. This may be conveniently incorporated in the procedure by way of differential equations which can be integrated along with equation (58).

6. COMPUTATIONAL ASPECTS

We have generated solutions of the two-level model using both techniques on the ICT Atlas computer at the S.R.C.'s Chilton Laboratory. The discrete method involves extensive use of a matrix multiplication routine which computes scalar products of two vectors double length and rounds the result correctly to single length. In the Riccati method, we use Merson's version of the fourth-order Runge-Kutta method, which provides an estimate of error that can be used to control the step-length. It is also necessary in the Riccati method to interpolate in tables of $\underset{\sim}{X}_n$ and $\Psi_n = \underset{\sim}{\phi}_n^T \underset{\sim}{M}^{-1} \underset{\sim}{\tilde{V}}_n$ generated on the forward sweep to provide the necessary intermediate values for Runge-Kutta integration in the reverse direction. We have done this by constructing the coefficients of a Newton interpolation polynomial for a given set of nodes for each interval in turn. The coefficients may be preserved, and used to evaluate the interpolation polynomial for each of the required intermediate points. The coefficients are changed only when the nodes change, and this enables us to deal very simply with the end intervals.

https://doi.org/10.1017/S0025292100151267 Published online by Cambridge University Press

6.1 Storage Allocation

Both programs require basic data to generate the following information in the general inhomogeneous case:

$$\underset{\sim}{M}, \underset{\sim}{W}, \tau_n, \varepsilon_{n+1/2}, \gamma_{n+1/2}, \underset{\sim}{g}_{n+1/2}, \underset{\sim}{\Lambda}_{n+1/2}, \underset{\sim}{\phi}_{n+1/2}$$

Total: $3(K+1)N + 2K+1$ cells.

The discrete method requires storage of

$$A_{n+1/2}, B_{n+1/2}, \underset{\sim}{t}_{n+1/2}, \underset{\sim}{X}_{n+1/2}, \underset{\sim}{C}_{n+1/2}, \varepsilon_{n+1/2}, \underset{\sim}{\tilde{V}}_{n+1/2}, \underset{\sim}{Z}_n$$

on the forward sweep, and of $\bar{J}_n$ on the reverse sweep.

Total: $(6K+3)N+1$ cells.

The Riccati method requires storage of

$$\Psi_n, \underset{\sim}{X}_n$$

on the forward sweep, and of $\bar{J}_n$ on the reverse sweep.

Total: $(K+2)N$ cells.

These requirements must be added to the storage needed for program and system routines and for working space. For a typical problem with I = 10, m = 2 (giving K = 20) and N = 50, the additional storage required by the discrete method is about 8 blocks of 512 words, less than 20% of the total. It would be possible, with small modifications of the discrete method, to reduce this excess considerably.

6.2 Timing

We have estimated the amount of computing time required by counting multiplications and divisions needed in both methods. This does not allow for storage references and other operations, but it is reasonable to assume that they are roughly proportional to the number of multiplications and divisions in a problem of this type, and we have found no evidence to contradict the assumption. On this basis, the discrete calculation uses $\{5K^2 + 28K + 4\}$

https://doi.org/10.1017/S0252921100151267 Published online by Cambridge University Press

operations per layer, and the Riccati method $\{\frac{55K^2}{2} + \frac{125K}{2} + \frac{1}{2}p(p+11)(K+1)\}$ operations per layer, where p is the order of interpolation required on the reverse sweep. For practical values of K of about 20, this means that the discrete method is between 5 and 6 times faster than the Riccati method per layer. We have found that it is often necessary in the Riccati method to subdivide the layers in order to maintain accuracy, particularly near the boundaries. This is, of course, done automatically by the step controller. There is no such problem in the discrete method for which we only need to satisfy the inequality (36), so that the practical increase in speed is often larger, although book-keeping overheads diminish the effect. For example in a typical problem with constant Doppler width, having $\tau = 50$, $\beta = 0$, $\varepsilon = 10^{-6}$, $K = 30$, the discrete method completed the computation in 30 seconds, whereas the Riccati method took over 6 minutes.

6.3 Angular Quadrature

There are few problems involved here; we have found that the line profiles and source functions are insensitive to our choice. We have usually employed the abscissae and weights for Gauss quadrature of order m = 2 over [0,1], using an additional integration to compute the normal emergent line profile. We have also used Radau quadrature of order 3 with an abscissa specified at $\mu = 1$. This simplifies the program, but, in addition to increasing the storage requirement, it is rather expensive in computing time, firstly because it increases K by 50%, and secondly because the smallest cosine, μ_1, is nearer to zero than with the Gauss rule. This causes a decrease in the critical step size in the discrete method.

6.4 Frequency Quadrature

The choice of frequency quadrature presents a more difficult problem. For pure Doppler broadening, we can employ the finite bandwidth approximation used by Avrett and Hummer.[9] In our implementation, we have truncated the frequency interval to become $-\alpha \leq \xi \leq \alpha$, and used Gauss quadrature over this finite range. The accuracy depends on the number of

https://doi.org/10.1017/S0252921100151267 Published online by Cambridge University Press

points, and on the value of α, and we have selected a value of I such that the normalization condition (19) is satisfied to prescribed accuracy. For example, choosing $\alpha = 4$, we require I = 10, so that the error in normalization may be less than one part in 10^6. We renormalize the integral to ensure that (19) is satisfied to within the round-off accuracy of the computer, about 11 decimal digits. It is important to do this so as not to introduce spurious sources and sinks in the calculation.[7]

Hummer and Rybicki have pointed out[2,3] that this finite bandwidth approximation is unsatisfactory for Voigt profiles, for the bandwidth α has to be increased substantially, particularly when $\beta>o$. They have discussed methods of overcoming this difficulty which involve splitting the range of integration and renormalizing each segment separately. While it certainly eliminates difficulties arising from lack of normalization, there is an arbitrariness about the procedure which is a little unsatisfactory, and it is not clear what effect changes in the method of splitting have on the final results. The difficulty will be more acute in variable property media, and it seems clear that this problem needs further investigation.

7. NUMERICAL RESULTS

Our main effort has been directed towards the comparison of the discrete and Riccati methods in a few idealized situations. We have computed the line source function $S_L(\tau)$ at a selected set of optical depths in slabs of various thicknesses together with the normal emergent intensity for Doppler broadened lines.

In the case of a homogeneous slab, the results from the two programs are closely similar, as Figures 1 and 2 indicate. However, there are small differences which are best demonstrated by an examination of Table 1.

(a) *Symmetry of $S_L(\tau)$ About the Mid-Plane*. The discrete solution is symmetric to the number of decimal places printed out. The Riccati solution is asymmetric by up to 5% in the worst case of linear interpolation, and up to 2% when cubic or quintic polynomial interpolation is used on the backward sweep.

https://doi.org/10.1017/S0252921100151267 Published online by Cambridge University Press

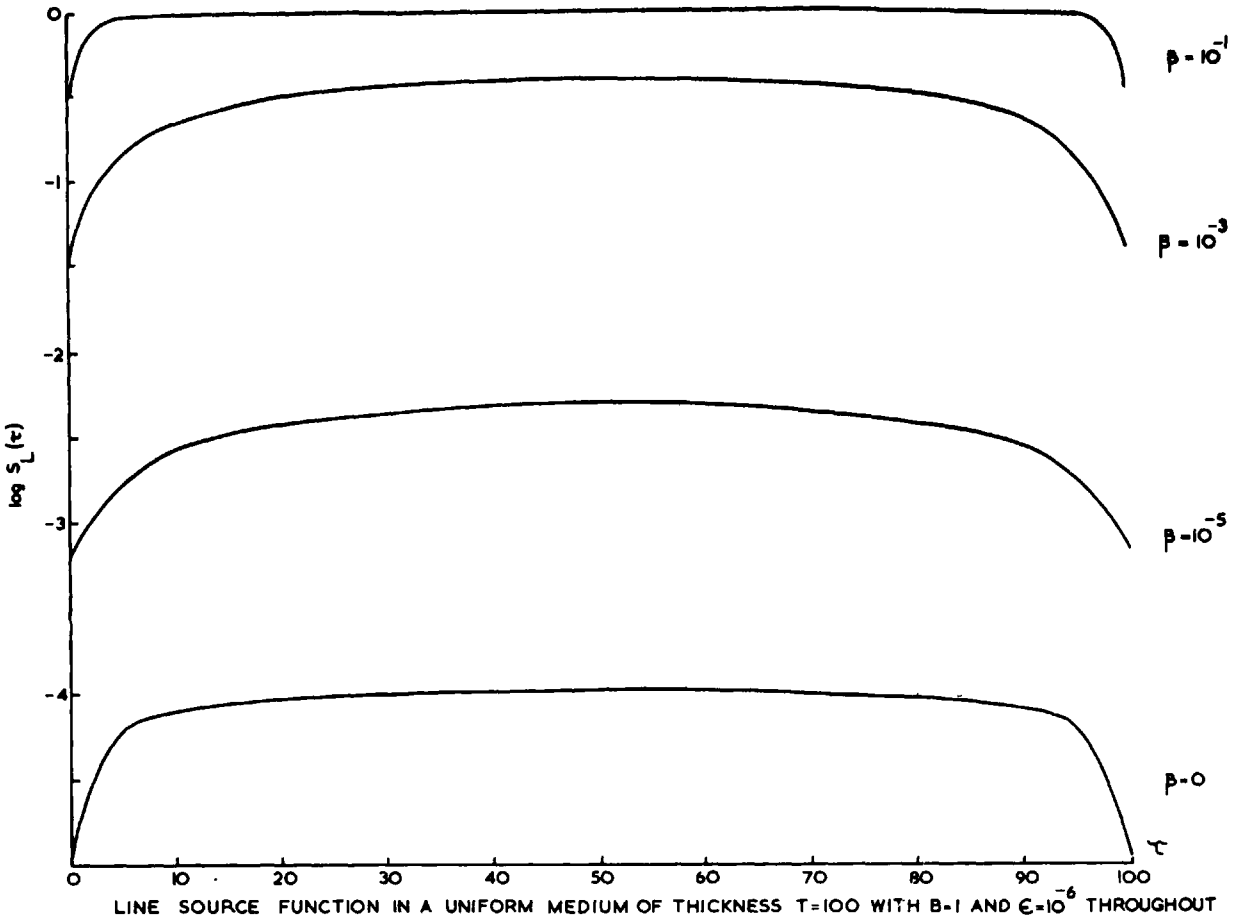
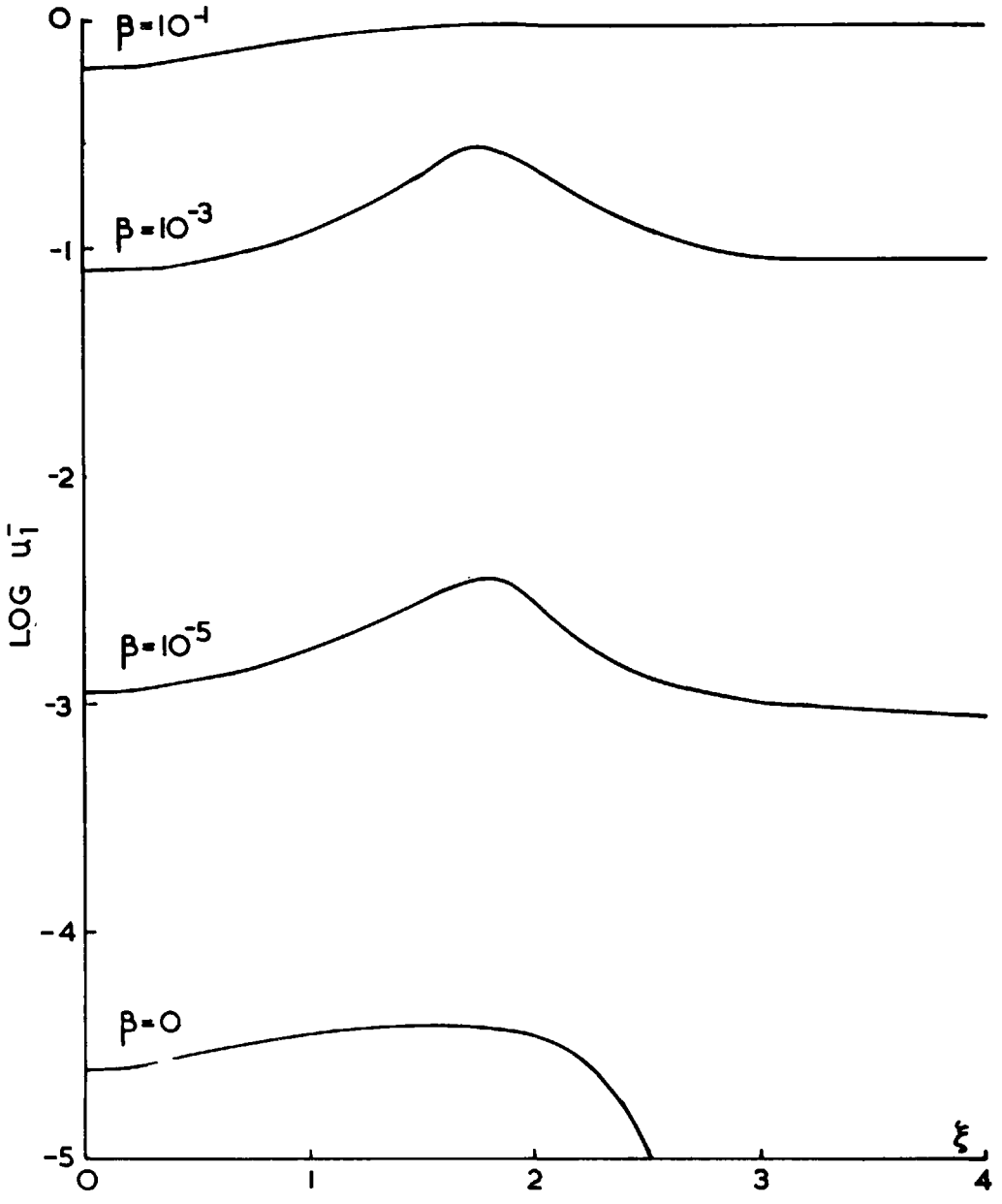


Figure 1. Line source function in a uniform medium, $\tau = 100$, with $B = 1$, $\rho = 1$, and $\varepsilon = 10^{-6}$. Discrete and Riccati solutions are not resolved.

https://doi.org/10.1017/S0252921100151267 Published online by Cambridge University Press



NORMAL EMERGENT INTENSITY FROM A UNIFORM MEDIUM OF OPTICAL THICKNESS T=100, WITH $\epsilon=10^{-6}$ AND B=1, AS A FUNCTION OF DOPPLER WIDTH.

Figure 2. Normal emergent intensities corresponding to the line source functions shown in Figure 1. Discrete and Riccati solutions are not resolved.

https://doi.org/10.1017/S0252921100151267 Published online by Cambridge University Press

(b) *Choice of Angular Quadrature.* The results are insensitive to this choice. Comparing Gauss and Radau solutions, we see that the greatest difference is in the fourth significant figure, both for the line source function and for the normal emergent intensity.

(c) *Variation of Step Length.* The layer thicknesses are presented in the discrete method. We have done some experiments to examine the truncation errors involved, which are summarized in Table 2.

It is clear that we can be confident of reducing the discretization error from this source to a reasonable level if we take $\tau<\tau_{crit}$. Further work to be published elsewhere shows that this error, for the operator approximations we have used, is proportional to the square of the layer thickness. This is verified by the results of Table 2.

In the Riccati method we have asked for an absolute accuracy of 10^{-4} in the S-integration, and a relative accuracy of 10^{-4} in the $\tilde{V}$ and u^- integrations. The program chooses its own step length to meet these criteria. This is always reduced most near the boundaries at the start of a sweep, sometimes as small as 0.03δ (where δ is the Doppler width) though for the most part it is close to the value of 0.7δ quoted as the stability limit by Rybicki and Hummer.[3] The very small step size needed near the boundaries is almost certainly associated with the fact that the source function has a logarithmically infinite slope there,[9] so that the Taylor expansion about $\tau = 0$, on which the Runge-Kutta process is based does not exist in the first interval.

(d) *Effect of Order of Interpolation in the Riccati Method.* Table 2 shows that the biggest improvement comes from replacing linear by cubic interpolation. The use of a fifth-order polynomial produces a less marked effect, and it is clear that there is little point in going further than third-order with step sizes of the order used in this problem.

(e) *The Line Profile.* The most marked discrepancy between the two calculations is at the line center, which is formed in the outermost layers of the slab. There is a deficiency at $\xi = 0$ of about 6% in the Riccati

https://doi.org/10.1017/S0252921100151267 Published online by Cambridge University Press

TABLE 1

This table gives typical solutions for the line source functions S_L and the normal emergent intensity for a plane parallel layer of thickness $\tau = 50$, with $\beta = 0$, $B = 1$. The frequency quadrature used a bandwidth $\alpha = 4$, and a 10-point Gauss rule.

(a) $\varepsilon = 10^{-}$

$\tau_{crit} = 1.0035$ (Gauss), 0.636 (Radau)

Source Function $S_L(\tau)$

	Discrete		Riccati with Radau quadrature		
τ	Gauss	Radau	p=1	p=3	p=5
0	0.3075426	0.3076172	0.2939989	0.3011204	0.3024015
2.5	0.5778	0.5770	0.5732	0.5792	0.5774
5.0	0.6940	0.6941	0.6922	0.6941	0.6934
10.0	0.8066	0.8067	0.8061	0.8067	0.8066
25.0	0.8817	0.8818	0.8817	0.8818	0.8818
40.0	0.8066	0.8067	0.8066	0.8067	0.8067
45.0	0.6940	0.6941	0.6938	0.6939	0.6939
47.5	0.5778	0.5770	0.5766	0.5766	0.5766
50.0	0.3075426	0.3076172	0.3076162	0.3076162	0.3076162

Normal Emergent Intensity*

ξ					
0.3061	5.031 (-1)	5.036 (-1)	4.730 (-1)	4.919 (01)	4.941 (-1)
0.9111	5.941 (-1)	5.943 (-1)	5.741 (-1)	5.881 (-1)	5.888 (-1)
1.4948	7.079 (-1)	7.079 (-1)	7.012 (-1)	7.065 (-1)	7. 064(-1)
2.0435	2.736 (-1)	2.736 (-1)	2.726 (-1)	2.735 (-1)	2.734 (-1)
2.5442	3.321 (-2)	3.321 (-2)	3.310 (-2)	3.319 (-2)	3.319 (-2)
2.9853	2.955 (-3)	2.955 (-3)	2.945 (-3)	2.953 (-3)	2.953 (-3)
3.3565	2.812 (-4)	2.812 (-4)	2.803 (-4)	2.811 (-4)	2.811 (-4)
3.6489	3.625 (-5)	3.625 (-5)	3.614 (-5)	3.623 (-5)	3.623 (-5)
3.8559	7.671 (-6)	7.671 (-6)	7.646 (-5)	7.667 (-6)	7.666 (-6)
3.9725	3.078 (-6)	3.078 (-6)	3.069 (-6)	3.077 (-6)	3.077 (-6)

https://doi.org/10.1017/S0252921100151267 Published online by Cambridge University Press

(b) $\varepsilon = 10^{-6}$

τ_{crit} = 1.0392 (Gauss), 0.7094 (Radau)

Source Function $S_L(\tau) \times 10^5$

	Discrete		Riccati with Radau quadrature		
τ	Gauss	Radau	p=1	p=3	p=5
0	0.8966753	0.8977293	0.8362320	0.8724352	0.8767388
2.5	2.069	2.068	2.048	2.081	2.071
5.0	2.794	2.796	2.784	2.797	2.792
10.0	3.788	3.789	3.784	3.789	3.789
25.0	4.852	4.850	4.848	4.849	4.849
40.0	3.788	3.789	3.789	3.789	3.789
45.0	2.794	2.796	2.795	2.795	2.795
47.5	2.069	2.068	2.066	2.066	2.067
50.0	0.8966753	0.8977293	0.8977239	0.8977239	0.8977239

Normal Emergent Intensity*

ξ					
0.3061	1.763 (-5)	1.766 (-5)	1.649 (-5)	1.726 (-5)	1.731 (-5)
0.9111	2.318 (-5)	2.320 (-5)	2.242 (-5)	2.300 (-5)	2.300 (-5)
1.4908	3.307 (-5)	3.308 (-5)	3.281 (-5)	3.303 (-5)	3.302 (-5)
2.0435	1.324 (-5)	1.324 (-5)	1.320 (-5)	1.324 (-5)	1.323 (-5)
2.5442	1.608 (-6)	1.608 (-6)	1.604 (-6)	1.608 (-6)	1.608 (-6)
2.9853	1.431 (-7)	1.431 (-7)	1.427 (-7)	1.431 (-7)	1.431 (-7)
3.3565	1.362 (-8)	1.362 (-8)	1.358 (-8)	1.362 (-8)	1.362 (-8)
3.6489	1.756 (-9)	1.756 (-9)	1.751 (-9)	1.755 (-9)	1.755 (-9)
3.8559	3.715 (-10)	3.715 (-10)	3.705 (-10)	3.714 (-10)	3.714 (-10)
3.9725	1.491 (-10)	1.491 (-10)	1.487 (-10)	1.491 (-10)	1.490 (-10)

*The representation a(-b) is used for a x 10^{-b}.

https://doi.org/10.1017/S0252921100151267 Published online by Cambridge University Press

TABLE 2

Dependence of Line Source Function and Normal Emergent Intensity on step size in solution by the Discrete method Case (a) of Table 1, $\tau_{crit} = 1.0035$ (Gauss quadrature).

	Percentage changes	
$k = \Delta\tau/\tau_{crit}$	$\lvert S_L(0)_k - S_L(0)_{k=1}\rvert / S_L(0)_{k=1}$	$\lvert \bar{u}(0)_k - \bar{u}(0)_{k=1}\rvert / \bar{u}(0)_{k=1}$
2.0	4	10
0.5	1	2
0.25	0.25	0.3

Note: The maximum change always occurs at the boundary.

https://doi.org/10.1017/S0252921100151267 Published online by Cambridge University Press

solution with linear interpolation falling to 2% if cubic interpolation is used. This difference decreases towards the wings of the line as may be expected.

We have also computed some results with variable line-profile. The results, shown in Figures 3 and 4 are closely similar to those of the corresponding Figures 6 and 7 of Rybicki and Hummer's paper.[3]

8. CONCLUSIONS

The discrete method has a number of advantages compared with the Riccati method which may be summarised as follows:

1. Much greater speed - an order of magnitude.
2. Better control of discretization and round-off errors.
3. Ease and simplicity of programming.

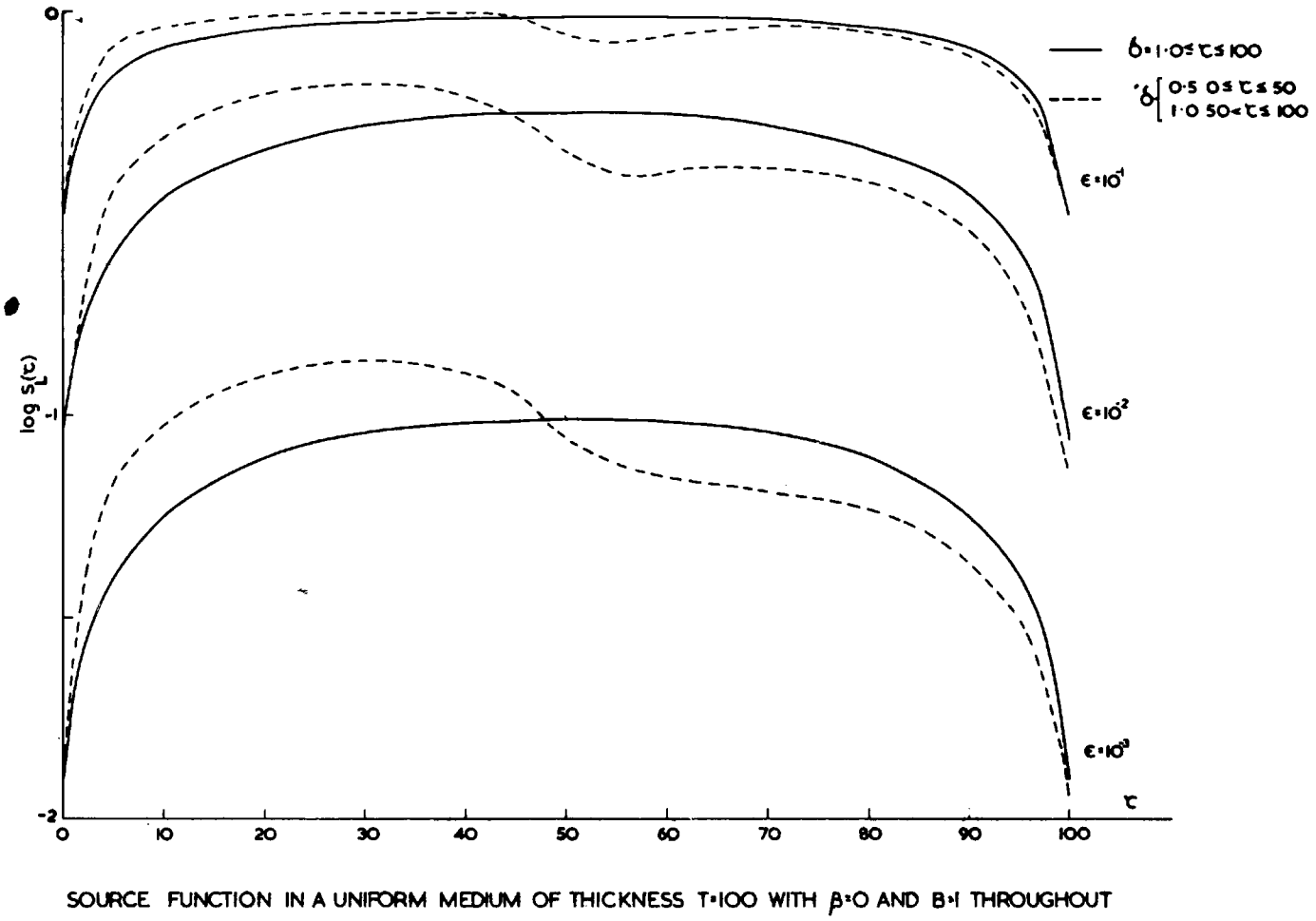
Its greatest disadvantage is that it requires more storage, at least in the current version. It might be possible to circumvent this by reprogramming so that it is unnecessary to retain data for all depths other than certain specified ones. This may be done by computing $\underset{\sim}{Z}_n$ and the scalar product $\phi^T_{n-1/2} \underset{\approx}{M}^{-1} \tilde{\underset{\sim}{V}}_{n-1/2}$ required for (49) for selected points $n = n_1, n_2, \ldots$ on the forward sweep, together with vectors $\bar{\underset{\sim}{\Sigma}}(n_\ell, n_{\ell+1})$ and matrices $\underset{\approx}{t}(n_\ell, n_{\ell+1})$ required to solve the equation

$$\bar{\underset{\sim}{u}}_{n_\ell} = \hat{\underset{\approx}{t}}(n_\ell, n_{\ell+1}) \bar{\underset{\sim}{u}}_{n_{\ell+1}} + \bar{\underset{\sim}{V}}(n_\ell, n_{\ell+1})$$

These vectors and matrices are obtained by simple recurrence forumlae derivable from equation (49).

A similar dodge has been used to limit the storage requirement in our version of the Riccati method, but the larger the intervals chosen, the more strain is thrown on the interpolation polynomial in the backward sweep. There is no such constraint in the discrete method. The accumulation of round-off error in calculating scalar products in the Riccati method also needs some investigation, but it could largely be eliminated at the cost of a little inconvenience.

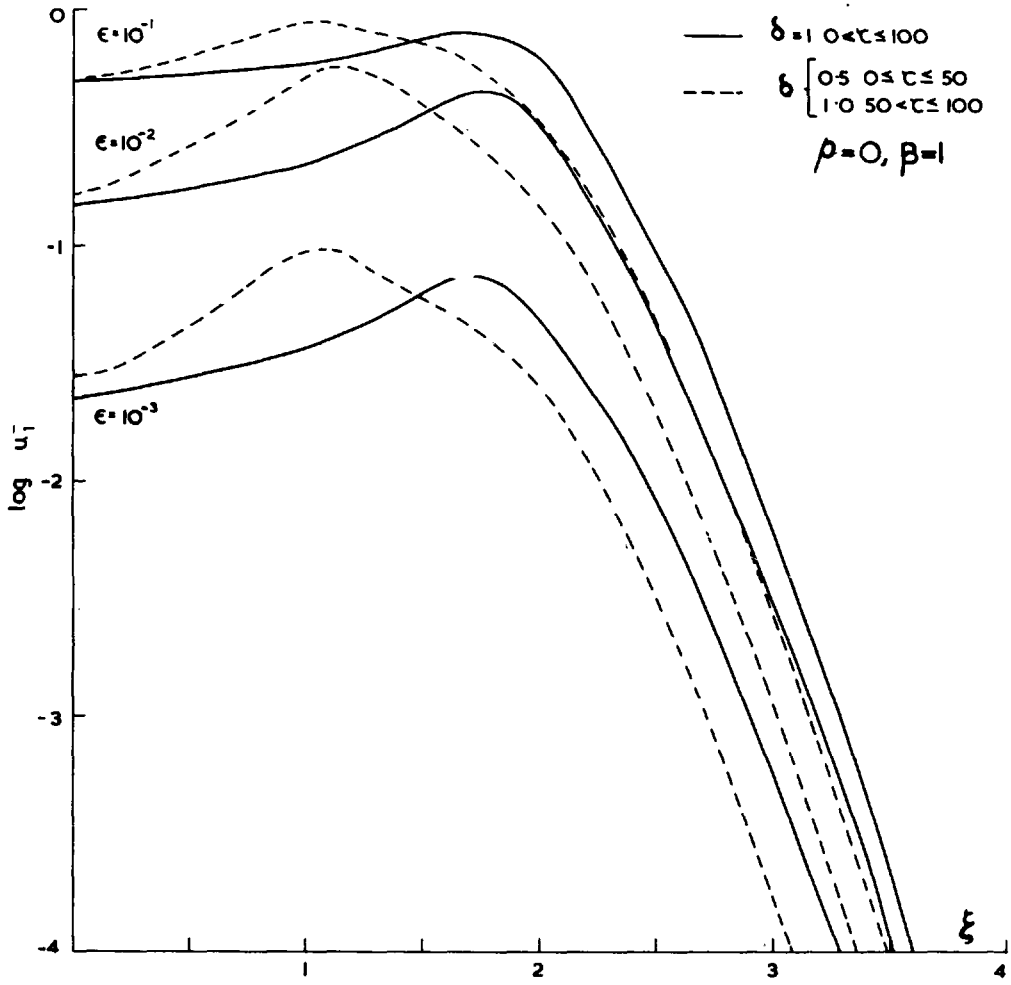
https://doi.org/10.1017/S0252921100151267 Published online by Cambridge University Press



SOURCE FUNCTION IN A UNIFORM MEDIUM OF THICKNESS T=100 WITH β=0 AND B=1 THROUGHOUT

Figure 3. Line source function in a medium of thickness $\tau = 100$ with $\beta = 0$, $B = 1$ throughout, and ε independent of depth. The Doppler width $\delta = 0.5$ for $0 \leq \tau < 50$, and $\delta = 1.0$ for $50 < \tau \leq 100$.

https://doi.org/10.1017/S0252921100151267 Published online by Cambridge University Press



NORMAL EMERGENT INTENSITIES FROM A UNIFORM MEDIUM OF OPTICAL THICKNESS T=100, AS A FUNCTION OF DOPPLER WIDTH

Figure 4. Normal emergent intensities at $\tau = 0$ corresponding to the line source functions shown in Figure 3.

We have not attempted any solutions for semi-infinite media so far, but it would not be difficult to do so using the Asymptotic theory as described by Rybicki and Hummer.[3] Alternatively, since the asymptotic region is of necessity assumed to be homogeneous, the doubling method could be applied.[6,7]

Finally, we may remark that Feautrier's method leads to a set of matrix equations which, like our own equations (3), have a block tridiagonal form and can be solved using the same algorithm.[8] We have tried to see if it is possible to interpret the Feautrier algorithm in terms of our theory. We have not been able to do so because the way in which the positive and negative intensities are combined by Feautrier makes it impossible to disentangle the operators. We have therefore been unable to make any fair comparison of his method and our own.

https://doi.org/10.1017/S0252921100151267 Published online by Cambridge University Press

ACKNOWLEDGMENT

We are indebted to Mrs. Elizabeth Gill for her assistance with the programs.

REFERENCES

1. J. T. Jefferies, *Spectral Line Formation*, Blaisdell (1968).
2. D. G. Hummer and G. Rybicki, *Methods of Computational Physics* 7, 53-127 (1967).
3. G. B. Rybicki and D. G. Hummer, *Astrophys. J.* 150, 607-635 (1967).
4. P. Feautrier, *C. R. Acad. Sci. Paris*, 258, 3189 (1964).
5. D. Mihalas, *Astrophys. J.* 150, 909-925 (1967).
6. I. P. Grant and G. E. Hunt, *Mon. Not. R. Astr. Soc.* 141, 27-42 (1968).
7. I. P. Grant and G. E. Hunt, *Proc. Roy. Soc.* A313, 183-198; ibid. A313, 199-216, (1969).
8. E. Isaacson and H. B. Keller, *Analysis of Numerical Methods*, pp. 58-62. Wiley (1966).
9. E. H. Avrett and D. G. Hummer, *Mon. Not. R. Astr. Soc.* 130, 295-331 (1965).

APPENDIX

A. The results of Sections 3 and 4 depend heavily on the following Lemma. Let $\underset{\sim}{X}$ and $\underset{\sim}{Y}$ be any two K-vectors and let

$$z = \underset{\sim}{Y}^T \underset{\sim}{X}.$$

Then if $|z| < |a|$,

$$[a\underset{\sim}{I} - \underset{\sim}{X}.\underset{\sim}{Y}^T]^{-1} = a^{-1}[\underset{\sim}{I} + (a-z)^{-1}\underset{\sim}{X}.\underset{\sim}{Y}^T].$$

Proof. By the binomial expansion, the result follows directly.

B. Thin Layer Operators and Sources

The approximations given in our paper (6) for thin-layer operators make use of matrices

$$\underset{\sim}{Z} = \underset{\sim}{\Lambda} - \sigma\underset{\sim}{\phi}\underset{\sim}{\phi}^T \underset{\sim}{W}, \quad \underset{\sim}{Y} = \sigma\underset{\sim}{\phi}\underset{\sim}{\phi}^T \underset{\sim}{W} \tag{B1}$$

and

$$\underset{\sim}{\Delta} = [\underset{\sim}{M} + \tfrac{1}{2}\tau\ \underset{\sim}{Z}]^{-1}. \tag{B2}$$

If the "diamond" weights are taken to minimize the local truncation error, equations (A.7) and (A.8) of reference 6 may be written for the present problem in the form

$$\underset{\sim}{r} = \underset{\sim}{\rho}(\underset{\sim}{I} - \underset{\sim}{\rho}^2)^{-1}\underset{\sim}{\theta}, \quad \underset{\sim}{t} = 2(\underset{\sim}{I} - \underset{\sim}{\rho}^2)^{-1}\underset{\sim}{\theta} - \underset{\sim}{I} \tag{B3}$$

where

$$\underset{\sim}{\rho} = \tfrac{1}{2}\tau\ \underset{\sim}{\Delta}\ \underset{\sim}{Y}, \quad \underset{\sim}{\theta} = \underset{\sim}{I} - \tfrac{1}{2}\tau\ \underset{\sim}{\Delta}\ \underset{\sim}{Z} \tag{B4}$$

https://doi.org/10.1017/S0252921100151267 Published online by Cambridge University Press

From (B2) and (B1)

$$\underset{\sim}{\Delta} = [\underset{\sim}{M} + \frac{1}{2}\tau\ \underset{\sim}{\Lambda} - \frac{1}{2}\sigma\tau\underset{\sim}{\phi}\underset{\sim}{\phi}^{T}\underset{\sim}{W}]^{-1}$$

$$= (\underset{\sim}{M} + \frac{1}{2}\tau\underset{\sim}{\Lambda})^{-1}\ [\underset{\sim}{I} - \underset{\sim}{\phi}\underset{\sim}{\phi}^{T}\underset{\sim}{D}]^{-1}$$

where

$$\underset{\sim}{D} = \frac{1}{2}\tau\underset{\sim}{W}\ (\underset{\sim}{M} + \frac{1}{2}\tau\underset{\sim}{\Lambda})^{-1}$$

is a diagonal matrix. Applying the lemma, we get

$$\underset{\sim}{\Delta} = (\underset{\sim}{M} + \frac{1}{2}\tau\underset{\sim}{\Lambda})^{-1}[\underset{\sim}{I} + \frac{1}{1 - \delta}\ \underset{\sim}{\phi}\underset{\sim}{\phi}^{T}\underset{\sim}{D}] \qquad \text{(B5)}$$

where

$$\delta = \underset{\sim}{\phi}^{T}\underset{\sim}{D}\underset{\sim}{\phi}$$

is the quantity written out in equation (34). Thus

$$\underset{\sim}{\rho} = \frac{1}{1 - \delta}\ (\underset{\sim}{M} + \frac{1}{2}\tau\underset{\sim}{\Lambda})^{-1}\underset{\sim}{\phi}\underset{\sim}{\phi}^{T}(\frac{1}{2}\sigma\tau\underset{\sim}{W})$$

and an additional application of the lemma gives

$$(\underset{\sim}{I} - \underset{\sim}{\rho}^{2})^{-1} = \underset{\sim}{I} + \frac{\delta}{1 - 2\delta}\ (\underset{\sim}{M} + \frac{1}{2}\tau\underset{\sim}{\Lambda})^{-1}\underset{\sim}{\phi}\underset{\sim}{\phi}^{T}(\frac{1}{2}\sigma\tau\underset{\sim}{W}) \qquad \text{(B6)}$$

Equations (30) to (33) now follow in a straightforward manner.

C. Discrete Equations which are not used in the present problem

We first list the local operators that we do not use explicitly:

$$\underset{\sim}{T}_{n+1/2} = \underset{\sim}{I} + A_{n+1/2}\underset{\sim}{C}_{n+1/2}\underset{\sim}{C}^{T}_{n+1/2}\underset{\sim}{S}_{n+1/2}$$

https://doi.org/10.1017/S0252921100151267 Published online by Cambridge University Press

$$\hat{\underset{\sim}{t}}(n+1,n) = \underset{\sim}{t}_{n+1/2} + A_{n+1/2}\underset{\sim}{M}^{-1}\underset{\sim}{W}^{-1}_{n+1/2}\underset{\sim}{X}_{n+1/2}\underset{\sim}{C}^{T}_{n+1/2}\underset{\sim}{MW}_{n+1/2}$$

$$\hat{\underset{\sim}{r}}(n+1,n) = A_{n+1/2}\underset{\sim}{C}_{n+1/2}\underset{\sim}{C}^{T}_{n+1/2}\underset{\sim}{M}\,\underset{\sim}{W}_{n+1/2} \qquad \text{(C1)}$$

$$\underset{\sim}{R}_{n+1/2} = \underset{\sim}{M}^{-1}\underset{\sim}{W}^{-1}_{n+1/2}\{\underset{\sim}{t}_{n+1/2} + A_{n+1/2}\underset{\sim}{X}_{n+1/2}\underset{\sim}{C}^{T}_{n+1/2}\}\underset{\sim}{S}_{n+1/2}$$

$$\hat{\underset{\sim}{t}}(n,n+1) = \underset{\sim}{t}_{n+1/2} + A_{n+1/2}\underset{\sim}{C}_{n+1/2}\tilde{\underset{\sim}{X}}_{n+1/2}$$

The equation that is not used explicitly on the forward sweep is

$$\bar{\underset{\sim}{V}}_{n+1/2} = \underset{\sim}{\Sigma}_{n+1/2} + B_{n+1/2}\underset{\sim}{C}_{n+1/2} \quad , \qquad \text{(C2)}$$

and that which is not used explicitly on the backward sweep is

$$\overset{+}{\underset{\sim}{u}}_{n+1} = \underset{\sim}{M}^{-1}\underset{\sim}{W}^{-1}_{n+3/2}\underset{\sim}{S}_{n+3/2}\overset{-}{\underset{\sim}{u}}_{n+1} + \underset{\sim}{M}^{-1}\underset{\sim}{W}^{-1}_{n+1/2}\tilde{\underset{\sim}{V}}_{n+1/2}. \qquad \text{(C3)}$$

https://doi.org/10.1017/S0252921100151267 Published online by Cambridge University Press