Astron. Astrophys. 284, 675–686 (1994)

ASTRONOMY AND ASTROPHYSICS

Two–dimensional radiative transfer with partial frequency redistribution

I. General method

L.H. Auer[1] **and F. Paletou**[2]*

[1] Los Alamos National Laboratory, Los Alamos, NM 87545, USA
[2] Institut d'Astrophysique Spatiale, Université Paris XI, Bâtiment 121, F-91405 Orsay Cedex, France

Received 29 July 1993 / Accepted 9 November 1993

Abstract. We present a new method for the solution of non–LTE scattering problems in two dimensions. It is based on Accelerated Lambda Iteration and an improved short characteristic method. It is more than an order of magnitude faster than a direct approach for Complete Redistribution. We, further, have extended the method to the solution of Partial Redistribution problems. The computational cost of treating PRD with the new method is only a small factor larger than CRD. Results with the new iterative approach are in complete agreement with previously published results.

Key words: radiative transfer – methods: numerical

1. Introduction

In order to predict the emergent radiation from astronomical objects of finite size it is clear that we must be able to include the effects of losses through the multiple boundaries. The study of solar prominence resonance lines (Vial 1982) showed that two–dimensional (2D) radiative transfer improves the agreement between emergent profile computations and observations. The non–LTE modeling of spectral lines observed in isolated structures in radiative interaction with neighboring regions, such as solar prominences, needs to take into account the effects of lateral radiation transfer, which in turn requires a multi–dimensional geometric description of these structures.

To advance this two–dimensional modeling effort, however, new algorithms are needed, both to reduce the computational cost as well as to improve the level of physical approximations used. The computations of Vial (1982), for example, were performed with the two–dimensional radiative transfer code developed by Mihalas et al. (1978; MAM). This code suffers from two limitations: even on the fastest available computers it is so computationally intensive as to preclude its routine application, and it can only treat complete frequency redistribution (CRD) in a two–level atom. However, it is now well established that partial redistribution (PRD) effects are important in the wings and near wings of the strongest lines. The computations of Heinzel et al. (1987) show, for example, theoretical Lyman α profiles with strong symmetrical peaks which are caused by the partially coherent penetration of the incident solar chromospheric profile.

We present here a set of techniques to be used for the efficient solution of 2D radiative transfer problems both in CRD and PRD. The numerical methods used to solve the transfer problem in the presence of scattering are based on the work of Olson et al. (1986; OAB). That approach is reviewed in Sect. 2, and the Short Characteristic (SC) method for the efficient solution of the two–dimensional transfer problem is then presented in Sect. 3 and Sect. 4. The material on 2D transfer is an extension of the work of Kunasz & Auer (1988) with an important improvement with respect to the ability to treat sharp edged beams. In Sect. 5 the OAB technique is combined with the short characteristic approach and applied to the two–dimensional complete redistribution case. The results are in agreement with those obtained with the MAM code, but are obtained with more than a magnitude less computing time. In Sect. 6 the one–point approximate operator of OAB is extended to the solution of the two–dimensional partial redistribution problem. The resulting method is somewhat similar to that proposed by Scharmer (1983) and can solve for the full frequency dependent source function in only a few times as much computing as required for the CRD case. Finally, we compare PRD to CRD effects on a theoretical case in Sect. 7.

2. Iterative solution of scattering problems

In Olson et al. (1986, OAB hereafter) it was demonstrated that the basic problem with the convergence of Λ–iteration could

Send offprint requests to: L.H. Auer
* Present address: High Altitude Observatory, National Center for Atmospheric Research, P.O. Box 3000, Boulder, CO 80307, USA

be overcome simply by using the diagonal of the Λ operator as an approximation to the full operator. This, in effect, permits optically thick layers to remain in equilibrium with themselves when the source function is corrected. Roughly speaking, information is propagated one layer rather than one optical depth per iteration. This is a dramatic improvement when dealing with cases with large total optical depths.

Application of the OAB approach is extremely simple. The exact scattering problem may be formally stated, for both coherent scattering and complete redistribution, as the solution of

$$S(\tau) = (1-\epsilon)\Lambda(S) + \epsilon B \tag{1}$$

In order to solve this system numerically, the continuous run of S(τ) is replaced by a set of discrete values, $S_d \approx S(\tau_d)$, using implicit interpolation to "fill in" the missing values between the grid points. The resulting system, the solution of which we are addressing here, is

$$S_d = (1-\epsilon)\sum_{d'} \Lambda_{d,d'} S_{d'} + (1-\epsilon)\bar{J}_d^{BC} + \epsilon_d B_d \tag{2}$$

The BC term is due to radiation incident on the boundaries. Equation (2) is a set of linear equations which, in principle, may be solved directly; however, if there is a large number of grid points, such a direct approach is impractical. First, because it requires the construction and storage of the full Λ matrix. Second, because the time required to solve the equations scales as the cube of the system size.

In order to obtain an iterative method for solving Eq. (2), we introduce a one–point approximation to the full Λ operator

$$\sum_{d'} \Lambda_{d,d'}(S_{d'} + \delta S_{d'}) \approx \bar{J}_d + \Lambda_{d,d}\delta S_d \tag{3}$$

where $\bar{J}$ is the average intensity defined by

$$\bar{J}_d = \sum_{d'} \Lambda_{d,d'} S_{d'} + \bar{J}_d^{BC} \tag{4}$$

Because the effect of the δS is approximated by a one–point operator, which is the diagonal of the true Λ operator, the perturbation of S at depth point d affects only that same depth point: $\delta\bar{J}_d \approx \Lambda_{d,d}\delta S_d$. The discrete representation of the Λ operator is a full matrix. The approximate operator is a diagonal matrix. Combining Eqs. (2) and (3) yields the scalar equation for the computation of δS

$$\delta S_d = \frac{(1-\epsilon_d)\bar{J}_d + \epsilon_d B_d - S_d}{1-(1-\epsilon_d)\Lambda_{d,d}} \tag{5}$$

Here S is used for the current estimate of the source function, and $\bar{J}$ the average intensity evaluated by applying the Λ operator to those S. The corrected value of the source function is given by

$$S'_d = S_d + \delta S_d \tag{6}$$

As proven in OAB, the choice of the diagonal yields a scheme which is absolutely guaranteed to be convergent.

The iterative cycle consists of the (a) evaluation of $\bar{J} = \Lambda(S)$, (b) computation of δS, and (c) correction of the source function. Although we have written in Eq. (4) the contribution of the source function to the scattered radiation as an explicit sum, operationally this term needs not to be evaluated in this manner. All we require is the value of $\bar{J}$. This can be found by whatever means is most advantageous.

The critical point about this iterative scheme is that it is local. That is, the changes at a point involve only values at that same point in the grid. The subscript d points to some location in the spatial grid. There is no restriction on the structure of that grid. This correction scheme is as valid in multiple dimensions as it is in one dimension.

In order to extend the OAB iterative technique to implicit scattering transfer problems in two dimensions we must have two things. The first is a method for the efficient solution of the explicit transfer equation, i.e. the evaluation of $\bar{J} = \Lambda(S)$ at all grid points for a known set of S_d. The second is the exact diagonal of the effective Λ operator corresponding to that method of solution. Even though it may not be necessary to know the values of the elements of the Λ operator explicitly in order to find $\bar{J}$, it is important to use the exact values of the diagonal of this operator in the iterative cycle. (In OAB an approximation was suggested for estimating the diagonal; however, such an approximation may impair convergence and should be avoided if possible. It is much better to evaluate the diagonal exactly using the fact that $\Lambda_{dd} = \Lambda(\delta_{d,d'})$; that is, it is equal to the result of applying the Λ operator in the case where only one value of the source function is non–zero). Given these the solution of the implicit scattering problem is straightforward.

3. Solution of the two–dimensional transfer problem

In order to implement our iterative technique, it is imperative that the solution to the formal transfer problem be obtained with the greatest speed. The obvious idea, long characteristics, involves using a set of rays, one for each of the angle quadrature directions, going through each point in the grid and extending across the entire grid. Values of quantities along the rays are obtained by interpolation. The intensity at the one point is found by solving the radiative transfer equation along the entire ray. This is obviously inefficient. If N is a measure of the number of points in both the y and z coordinates, the time for solving the one–dimensional transfer equation along each long characteristic ray scales as N. The time to evaluate the average intensity at all the grid points, therefore, scales as N^3.

Fortunately as demonstrated in Kunasz & Auer (1988) the efficiency of the solution of the two–dimensional radiative transfer problem is dramatically improved by the use of short characteristics. Instead of traversing the entire space, the short characteristics start at a grid point and extend in the "upwind" direction until they hit one of the cell boundaries, as shown in Fig. 1. In this figure the ray is propagating from the upper right and, therefore, these are the boundaries along which the intensity for this

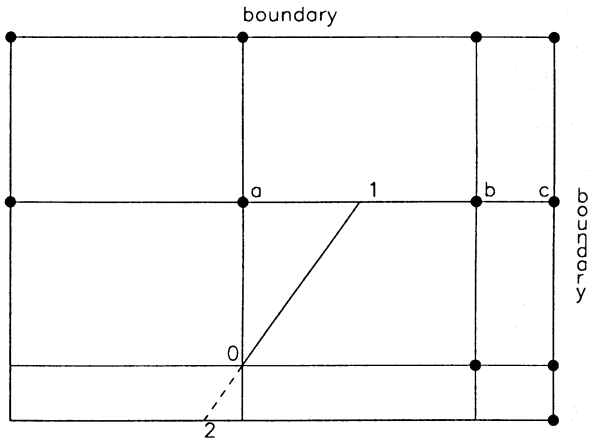


Fig. 1. Short characteristics at grid point 0 for a ray propagating from the upper right. The upwind point is 1. The radiation for this direction is known at all the darkened grid points

direction is a priori known. We step through the mesh propagating this information by using short characteristics to calculate the "downwind", unknown intensities. For example, in Fig. 1, we need to calculate the intensity at 0. We can evaluate this quantity simply by integrating across the cell from 1 to 0. The intensity at 0 for this direction is

$$I_0 = I_1 e^{-\Delta\tau_1} + \int_0^{\Delta\tau_1} S(\tau) e^{-\tau} d\tau \tag{7}$$

Evaluation of the integral in Eq. (7) is straightforward once we have estimated the values of the source function at the points 1 and 2 of Fig. 1. Because these do not lie at grid points they must be interpolated.

In order to recover the diffusion approximation, it is necessary to use parabolic or higher order interpolation in the evaluation of the integral. Not doing so would lead to serious errors in situations with high scattering. We may find the desired formula by implicit interpolation over (S_2,S_0,S_1) and then analytically integrating the interpolant $S(\tau) = S_0 + c_1\tau + c_2\tau^2$ which implies

$$\int_0^{\Delta\tau_1} S(\tau) e^{-\tau} d\tau = S_0 w_0 + c_1 w_1 + c_2 w_2 \tag{8}$$

The coefficients of which may be efficiently expressed in terms of the differences of the source functions

$$\begin{cases} d_1 = (S_0 - S_1)/\Delta\tau_1 \\ d_2 = (S_2 - S_0)/\Delta\tau_2 \\ c_1 = (d_2\Delta\tau_1 + d_1\Delta\tau_2)/(\Delta\tau_1 + \Delta\tau_2) \\ c_2 = (d_2 - d_1)/(\Delta\tau_1 + \Delta\tau_2) \end{cases} \tag{9}$$

with the weights

$$\begin{cases} w_0 = 1 - e^{-\Delta\tau_1} \\ w_1 = w_0 - \Delta\tau_1 e^{-\Delta\tau_1} \\ w_2 = 2w_1 - \Delta\tau_1^2 e^{-\Delta\tau_1} \end{cases} \tag{10}$$

This definition of the weights shows that they may be recursively generated. In the first interior row or column we do not use quadratic interpolation. No change in the quadrature formula is needed, however. One simply sets $c_1 = d_1$ and $c_2 = 0$, which forces linear interpolation at the boundary. (Note that since the boundary layer is normally optically thin, the use of linear interpolation there does not introduce significant error).

Once we have the weights in Eqs. (9–10), exact evaluation of the diagonal of the Λ operator is simple. At each point one sums over all directions with the d's in Eq. (9) defined by setting $S_0 = 1$, and taking the other S as 0. With the exception of the inclusion of the contribution from the off–diagonal source functions, the short characteristic formal solution of the transfer equation for $\bar{J}$, i.e. Eq. (4), and the computation of the diagonal of the Λ operator are formally identical and code for performing these tasks can be shared.

The problem with using Eq. (7) is that I_1 is not explicitly known. Only the intensities at the grid points – marked as dark spots – are known yet. Its value, however, is easily determined by interpolation along the upwind grid line. As long as we sweep the grid moving away from one or both of the boundaries, the values of the intensity in the upwind direction will have been determined and interpolation of the upwind boundary values for the short characteristics will be possible. The time to evaluate the radiation is directly proportional to the number of grid points; thus, the time to calculate the radiation field should vary as N^2 which is a factor of N better than with long characteristics.

The first step in the computation is the specification of the geometric weights. These depend only on the grid and the angle quadrature set. Once these have been fixed, the interpolation weights for each direction may be found. They do not depend on either the transition or frequency. As the algorithm for the solution of the scattering problem is iterative, from the standpoint of efficiency it is desirable to store these data either internally, or on external storage, buffering the input to minimize the time spent in their recovery.

The grid must be swept in an order which moves away from the one of the upwind boundaries. Let the cosines with respect to the axes of the direction into which the photon is moving be (c_y,c_z). Then, for example, if $c_y > 0$, we must evaluate the points in order of increasing y, or if $c_z < 0$, we must start at the largest z and step toward the other boundary. In order to minimize the storage, one evaluates all the intensities $I_{ij}(c_y, c_z)$ along the line $z = z_j$ before advancing to $j = j \pm 1$, where the sign of the increment is the sign of c_z. The storage of upwind values is discussed further below.

The other problem is to determine on which axis, y or z, the upwind point lies. This is done for Cartesian coordinates by a simple test. If Δy is the interval in y back to the previous y–grid line, and Δz is the interval back to the previous z–grid line, then if $\Delta y/c_y < \Delta z/c_z$ the ray hits the y–axis and the length of the step is $\Delta y/c_y$; otherwise, the upwind point lies on the z–axis, at the distance $\Delta z/c_z$ along the ray. The interpolation weights follow once these quantities are known. For interpolation in y,

if the upwind point is at $y_a < y < y_b < y_c$ or the reverse order but with y still in the interval (y_a, y_b),

$$\begin{cases} w_a = (y_b - y)(y_c - y)/(y_b - y_a)(y_c - y_a) \\ w_b = (y_a - y)(y_c - y)/(y_a - y_b)(y_c - y_b) \\ w_c = (y_a - y)(y_b - y)/(y_a - y_c)(y_b - y_c) \end{cases} \quad (11)$$

For interpolation in z, the formulae have identical forms.

It does not matter whether one makes the y or the z–axis the outer loop. It is only necessary to step through the grids in the order just discussed, away from one of the upwind boundaries. The requirement is that the intensities at the upwind points be known before the short characteristic is evaluated. This is guaranteed by the following, assuming that z is the more slowly advancing axis, for all the y_i points along the z_j line, given all the intensities for the direction under consideration on the upwind grid line, on the downwind line

(1) evaluate those points which interpolate in y then,

(2) evaluate those points which interpolate in z.

After all the y–grid values have been calculated, advance to the next z–line. The operational reason for splitting the points into the sets used in steps (1) and (2) is that the transfer for the points in step (1) is completely independent and, therefore, can be made in a fully parallel manner, but unless the intensities at these points have been determined, it will not be possible to interpolate all the upwind values needed in step (2).

To conclude the remarks about the interpolation, we should note that we have felt it important to retain symmetry between the axes with respect to the order of the interpolation. On the interior lines high order symmetric interpolation is possible in y, but only asymmetric interpolation is possible in z; in order to maintain equivalence between the two directions we, therefore, use parabolic interpolation for both directions. Along the first interior lines, however, sufficient data for parabolic interpolation is not always available, and here we use linear upwind interpolation.

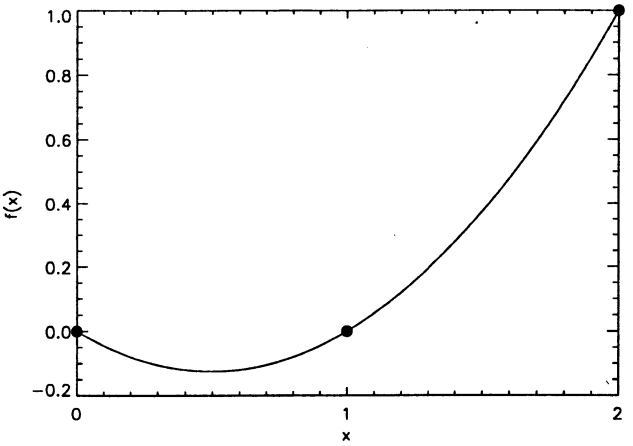


Fig. 2. Even though the values at x=0, 1, 2 are monotonic and greater than or equal to 0, parabolic interpolation generates a new extremum and "spurious" negative values

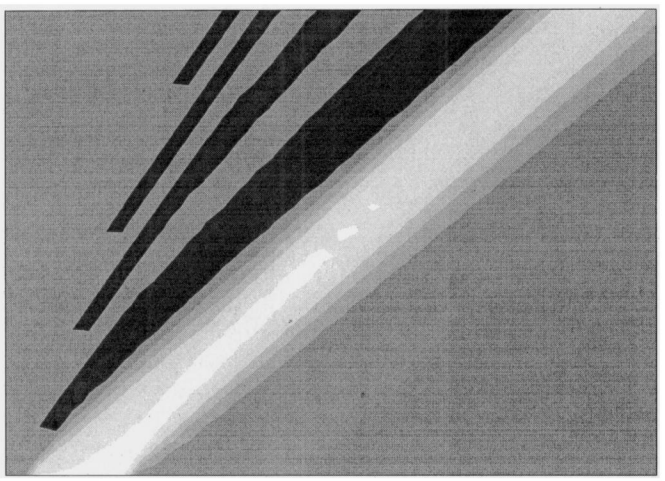


Fig. 3. Parabolic interpolation causes negative intensities in a sharp edged beam, making zero intensities appear gray in this half–tone plot

4. Notes on implementation of the formal solution

The efficiency obtained by the use of short characteristics to solve the two–dimensional radiative transfer problem was demonstrated in Kunasz & Auer (1988), but in order for this technique to become a fully practical method, two details in the implementation must be addressed: (1) the generation of "spurious" upwind intensities by high order interpolation, and (2) the efficient use of both vectorizing capabilities and memory. Both of these will be discussed here.

Note that for simplicity we will present only the example of interpolating $I(y)$ at an y in the interval (y_a, y_b) using the upwind values I_a, I_b, I_c and parabolic interpolation. Cases for different intervals or the z–direction are formally identical except for the indices.

4.1. Monotonic interpolation

As shown in Fig. 2, even if the values at a succession of points has monotonic order, e.g. $I_a \leq I_b \leq I_c$, or the reverse, parabolic interpolation of the upwind intensities can generate a new extremum, that is a value greater or less than the maximum or minimum of I_a and I_b. Although the values at the grid points are monotonic, the interpolated values may not be. The potential seriousness of this phenomenon is seen by noting that if $I_a = I_b = 0$ but $I_c = 1$, which corresponds to the edge of a sharp beam, $I(y)$ is negative for all $y_a < y < y_b$, which is physically impossible. Similarly, if $I_a = I_b = 1$ but $I_c = 0$, the interpolated $I(y)$ is larger than 1, which means that parabolic interpolation can potentially introduce spurious sinks and sources, even if all the upwind values are "reasonable".

The disastrous consequences of this negative overshoot are demonstrated in Fig. 3. Here we present a half–tone picture of the propagation of a sharp edged beam in a vacuum computed using short characteristics. The interpolation error produces a trough of negative intensities alongside the beam. The general gray background is caused by the fact that although these regions have zero intensity, they are *mathematically* not as dark as the

"darkest" values which have been incorrectly set less than zero! It is the rectification of this error that we now address.

The intensity is, of course, an integral over the upwind sources, and incident intensity. To the extent that the spatial variation of the sources is resolved by the gridding, the variation of the intensities will likewise be smooth on the grid, and there will be no interpolation difficulties. Problems arise when there are unresolved variations in the upwind sources and/or sinks, in particular when there is a step function variation in these quantities. This corresponds to an infinite derivative and invalidates the use of smooth interpolation. The problem is to recognize automatically when one has encountered such a situation. A partial answer is provided by the fact that if the structure has been resolved, and the intensity is monotonic at the grid points, then the intensity must also be monotonic at intermediate positions. In order for there to be subgrid structure in the radiation there must be subgrid structure in the material, which is contrary to the hypothesis that the material variations have been resolved. If, in fact, the interpolated values are not monotonic, the "existence" of a new extremum is due to errors introduced by the interpolation and should be discarded.

Mathematically, we insist that if (I_a, I_b, I_c) are monotonically ordered,

$$\min(I_a, I_b) \leq I(y) \leq \max(I_a, I_b)$$

i.e. that the interpolation does not introduce non–monotonic behavior in the upwind radiation. All we may a priori assert is that $I(y)$ is bounded by the interval (I_a,I_b). Accordingly, we might use linear interpolation in order to guarantee the ordering, as was done by van Leer (1977). We argue that while this will avoid the generation of spurious extrema, it will not improve the accuracy of the calculation relative to a simpler approach. The origin of the breakdown of the interpolation is usually that we have near step function variation in the upwind quantities. This, of course, is hardly better interpolated by a linear formula, than by a high order formula. The effect of linear interpolation is to introduce a spurious dispersion of a "hard edged" beam. Instead we suggest setting $I(y)$ equal to the minimum or maximum of I_a and I_b if the parabolic interpolant lies outside the permitted interval. The formal accuracy of this scheme is only zeroth order but in fact it is "exact" for the case of greatest interest, the step function. As stressed by Boris & Book (1973) in the presence of near singularities, the "formal order" of the interpolation is irrelevant, because the behavior simply cannot be approximated by polynomial interpolation. As shown in Fig. 4, the results from the linear and min–max approaches are essentially identical. Although both suppress the spurious interpolation values, and give a reasonable representation of a propagating beam, neither is entirely successful in suppressing the numerical dispersion of the ray. The origin of that dispersion is the inherent coarseness of the information which is being interpolated. That data do not give information on where the edge of beam lies exactly between the grid points. No matter what order of interpolation is used, it is impossible to estimate where the edge of a square beam lies, only that it is "somewhere between these two grid points".

4.2. Minimization of storage

In order to interpolate the upwind intensities, the frequency dependence of the intensity must be available. Given the potentially large number of frequency quadrature points, it is important to note that the full set of information has to be stored only for the current and two previous rows. That is, for the computation of the profile weighted average intensity, $\bar{J}$, used in CRD scattering problems, frequency dependent information is only required for three rows. This storage is simply cycled as one moves through the grid for a given direction. The contribution from the various frequencies are added to average intensity integral, $\bar{J}$, as they become available.

4.3. Vectorization

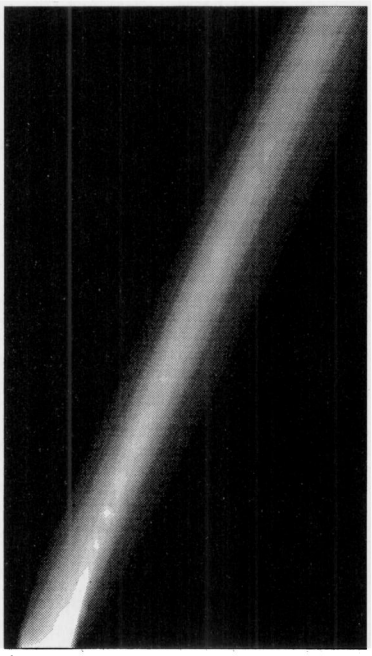
It is trivially guaranteed by making the inner loop run over frequency. In fact, from the standpoint of efficiency it is optimal to solve for the radiation field in all frequencies simultaneously, i.e. in all transitions, if more than one is being treated, as this maximizes the vector length being worked on. All geometric factors, such as the physical length of the short characteristics, are the same for each transition. With the exception of the opacity and emission there is no difference in the way different frequencies are treated. Further, because the directions are absolutely independent of each other, a potential major increase in speed can be obtained by parallel processing the various directions.

4.4. Acceleration of convergence

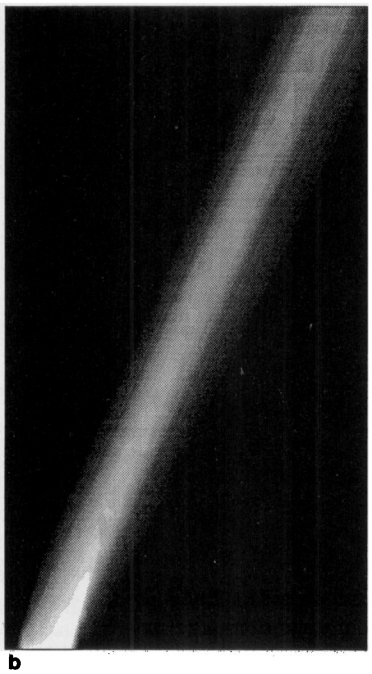
While it is easy to show by the use of Greshgorin's theorem as was done in OAB, that the diagonal of the two–dimensional Λ operator significantly improves the rate of convergence, the resulting convergence is still relatively slow. It is, therefore, highly desirable to improve further the speed with which the solution is obtained. Fortunately, this can be achieved in a simple manner, at very little computational cost.

Two general classes of methods are available for the acceleration of convergence: residual minimization and orthogonal vector acceleration. They have been described in Auer (1991) and are extremely effective for the multi–dimensional scattering problem. The Ng (1974) extrapolation, and the ORTHDX (Auer 1991) method both add less than 1% to the computational cost per iteration, but more than double the rate of convergence. Although the convergence is still linear, the overall cost is reduced by more than a factor of two by either of these techniques.

The effect of acceleration is demonstrated by the example shown in Fig. 5. The problem being solved is for complete redistribution over a Doppler profile with $\epsilon = 10^{-3}$, and optical depths 2×10^3 by 10^4. The results are typical of all cases. The unmodified convergence, using only the diagonal of the Λ operator, is shown by $\square$. While this is clearly "working", sixty iterations will be needed to achieve an accuracy of 10^{-6} in the solution. The convergence obtained with second–order Ng acceleration is shown by the filled circles. In this method, every fourth iteration a least squares extrapolation has been made in order to reduce the residuals (Auer 1987). The third sequence



a



b

Fig. 4a and b. Either **a** hard limiting, left or **b** linear interpolation, right, may be used to avoid overshooting

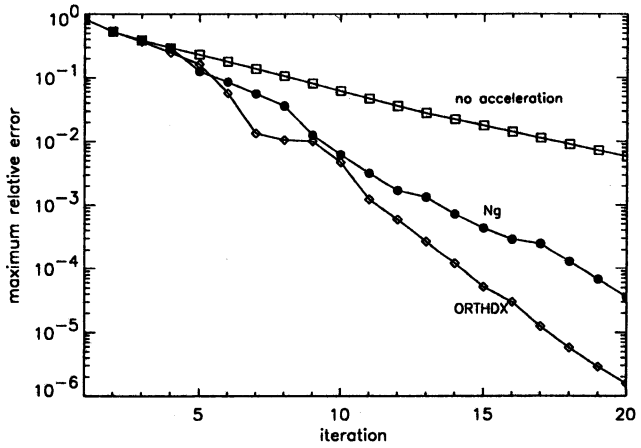


Fig. 5. The convergence rate is doubled by either residual (Ng) or orthogonal (ORTHDX) minimization

shown is that obtained by ORTHDX of order 2. The convergence is smooth and almost exactly double that of the original method. Higher order versions of ORTHDX did not achieve appreciably higher rates of convergence and are, accordingly, not recommended.

The choice between ORTHDX and Ng acceleration is determined by the availability of storage. ORTHDX in its second order form requires two more scratch matrices, the size of the matrix being solved, than the residual minimization algorithm. If this is available, then ORTHDX can and should be used. If not, Ng is still an effective alternative. In order to achieve the same accuracy with Ng requires only about 5 more iterations than ORTHDX.

As a final note, we should discuss the criterion for stopping the iteration. Because even with acceleration the convergence of the overall scheme remains linear, this can be treated in a relatively general manner. If $\Delta S^{(i)}$ is the maximum change in the i–th iteration, then a upper bound to the error of the i–th iterate is $\Delta S^{(i)}/[1-(\Delta S^{(i-1)}/\Delta S^{(i)})]$. This can thus be used to test for the termination of iteration. If the estimated error is better than the desired maximum error, you may stop the iteration loop. We have used the requirement that $\Delta S/S < 10^{-2} \times \epsilon$ as our stopping criterion. The only requirement for validity of this formula is that the ratio of ΔS between successive iterations has become approximately constant. As we see in Fig. 5, this requirement is relatively rapidly fulfilled.

5. Performance of the short characteristic method for the two–level atom

The key concerns with respect to the utility of the short characteristic method are its accuracy and speed. The results for both are remarkably positive and confirm that this should be the method of choice for the solution of two–dimensional radiative transfer problems.

The accuracy of the new method may be established by comparison with the results of the MAM code using the same angle quadrature and spatial grids. The short characteristic approach, which is based on iterative solution of an approximation to the integral equations for the source function, is fundamentally different from MAM, which is based on direct solution of multi–dimensional difference equations. Thus, comparison affords a direct check of the accuracy. We present, in Fig. 6, such comparisons for Doppler profile cases with $\epsilon = 10^{-4}$, $B = 1$ and optical depths $\tau = 10^4$ along both sides of the slab. Radiation, equal to B, is incident on both sides and the bottom. Five points per decade of optical depth, which is 63 grid points on each side, were used. In the figure, for each point in the

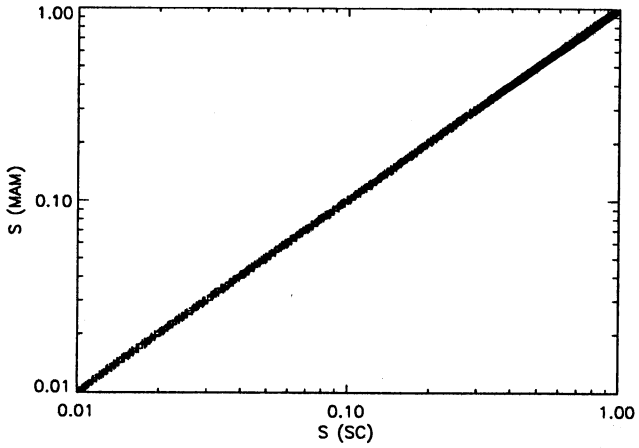


Fig. 6. Agreement between source function calculated using short characteristics and MAM for $\epsilon = 10^{-4}$ and $\tau = 10^4$

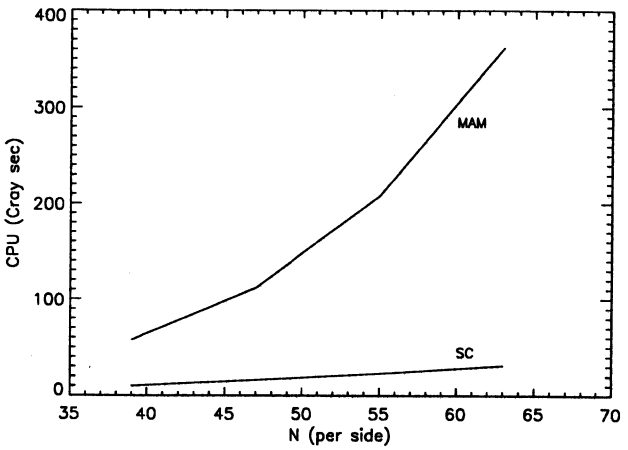


Fig. 8. Timings for direct and iterative methods as function of grid resolution

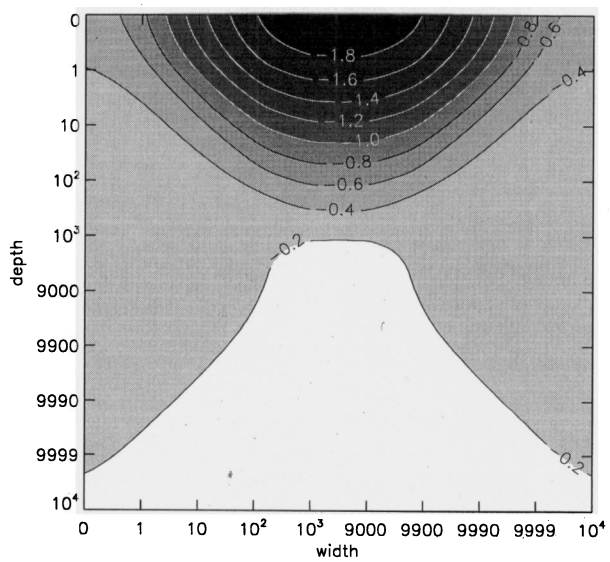


Fig. 7. Contour diagram of $\log(S(y,z))$. Note that, as in MAM, the same double logarithmic grid used in the computation has been used in this plot. This resolves the behavior at the boundary, but compresses the inner optically thick region

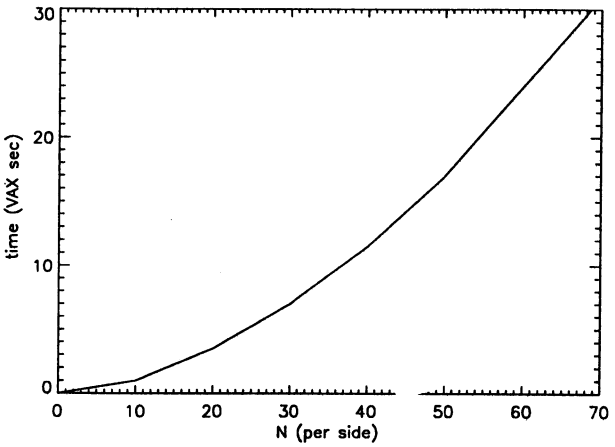


Fig. 9. Timing for converged solutions on a VAX 8650 computer. The dependence on N is clearly quadratic as shown by the fitted curve

two–dimensional grid we have plotted the values of the source functions computed using the two methods against each other. There is a nearly perfect 45° slope over the factor of 100 range in the source function. The worst disagreements are $\Delta S/S$ = 2.6% and occur in the optically thin boundary.

The results for $S(y,z)$ itself are shown in Fig. 7. We note, in particular, that the SC method recovers the correct asymptotic behavior with $S = B$ at high optical depths and $S_{\rm surface} = \sqrt{\epsilon}B$ in the middle of the sides where loss through the other sides becomes negligible.

The relative timing of the MAM code and the SC methods for the solution of non–LTE scattering problems are shown in Fig. 8. For typical grid sizes, with N $\geq$ 50, the short characteristic method is more than an order of magnitude faster than the direct solution of MAM. The data used here is the reported CPU time for runs on a Cray–2 computer (Paletou 1992). Square grids, with optical depths 10^4 on a side were used, as were $\epsilon = 10^{-4}$ and Doppler profiles. Convergence was accelerated using Ng's method for residual minimization. The solutions were iterated until accuracies better than $10^{-2} \times \epsilon$ had been obtained after, roughly, 25 times.

It is striking to note how far away both timing curves lie from the theoretical prediction for a dense system of linear equations. The size of the set of linear equations for the two–dimensional scattering problem, Eq. (2), scales as N^2. If the matrices were dense, the solution time should scale as N^6. In fact the two–dimensional transfer systems are not dense and, both the MAM and SC approaches take this into account. The timing of MAM is, according to this data, best represented as an exponential in N. The timing of the SC method is roughly linear in N. This later result is particularly perplexing if one considers that the number of unknowns is increasing as N^2. For N $\leq$ 100 the time for solution increases slower than the number of unknowns. Of course, this is simply a reflection of the relative sizes of terms in the polynomial as a function of N and that we have

used the full run time in the plots. For larger values of N or with more careful timing the curve for the timing of the SC solution is clearly non–linear, as is shown in Fig. 9 for a VAX computer. The efficiency of the SC method is so great that two–dimensional transfer computations, even for multi–level atoms, are entirely feasible using modest computers (Auer et al. 1993, in preparation for publication).

6. Solution including the effects of partial frequency redistribution

Our approach to the solution of the transfer equation including the effect of partial redistribution is, like Scharmer (1983), based on the use of an approximate operator technique. We, however, both use a different approximate operator and apply iterative acceleration (Auer 1991) to achieve a technique which solves the two–level atom scattering problem with partial redistribution and is only a small factor more computationally expensive than the complete redistribution case. Further, our approach to the PRD problem is completely compatible with the multi–dimensional short characteristic method described above, so we can take full advantage of the efficiency of this method for the treatment of two–dimensional cases. Remarkably, we can now solve the PRD scattering problem faster than the MAM approach could solve the CRD scattering problem.

The effective two–level atom source function in the presence of an overlapping continuous opacity may be written

$$S_\nu = \frac{\chi_L \phi_\nu S^L_\nu + \kappa B}{\chi_L \phi_\nu + \kappa} \tag{12}$$

where χ_L is the line opacity, κ the continuous opacity, and ϕ_ν the line profile. The line source function including partial redistribution may be written

$$S^L_\nu = (1-\epsilon)\bar{J}_\nu + \epsilon B \tag{13}$$

where the interchange among frequencies has been included in the definition of

$$\bar{J}_\nu = \phi_\nu^{-1} \int R(\nu,\nu') J_{\nu'} d\nu' \tag{14}$$

which integrates the average intensity, J_ν, against the angle averaged redistribution function. The frequency and angle averaged intensity which appear in the complete redistribution source function

$$S^L = (1-\epsilon)\bar{J} + \epsilon B \tag{15}$$

is the familiar

$$\bar{J} = \int \phi_{\nu'} J_{\nu'} d\nu' \tag{16}$$

although one must be careful to remember that the J_ν appearing in Eq. (16) refers to the result obtained with the true PRD source function, as given by Eqs. (12–16). Formally, we may write

$$\bar{J} = \int \phi_{\nu'} \Lambda_{\nu'}(S_{\nu'}) d\nu' = \int \phi_{\nu'} \Lambda_{\nu'} \left(\frac{\chi_L \phi_{\nu'} S^L_{\nu'} + \kappa B}{\chi_L \phi_{\nu'} + \kappa}\right) d\nu' \tag{17}$$

or

$$\bar{J} = L(S^L_{\nu'}) + \bar{J}_c \tag{18}$$

where we have introduced, for notational convenience, the new operator $L(S^L_\nu)$, which is just the ordinary Λ operator with the appropriate ratio of the line to total opacity as a weighting factor. We may further define the difference between the CRD and PRD line source functions as

$$S^L_\nu = S^L + \Delta S^L_\nu \tag{19}$$

with

$$\Delta S^L_\nu = (1-\epsilon)(\bar{J}_\nu - \bar{J}) \tag{20}$$

Note that, the ΔS^L_ν is not a correction. It is used to stand for the difference between the two average intensity integrals.

The iteration for the frequency dependent partial redistribution scattering source function consists of two nested loops. In the outer loop, using the current estimate of S^L_ν, we evaluate ΔS^L_ν using Eq. (20). We hold this quantity fixed during the "inner" iteration cycle which tries to solve the system

$$S^L = (1-\epsilon)(L(S^L) + \bar{J}_c) + (1-\epsilon)L(\Delta S^L_\nu) + \epsilon B \tag{21}$$

or

$$S^L = (1-\epsilon)L(S^L) + r \tag{22}$$

with

$$r = (1-\epsilon)(\bar{J}_c + L(\Delta S^L_\nu)) + \epsilon B \tag{23}$$

During the solution of Eq. (22) the spatially variable residual vector, r, defined by Eq. (23) is held fixed and we seek only to find a S^L which satisfies Eq. (22). In his presentation, Scharmer advocated the solution of the system as if it were a real two–level CRD problem because the mathematics underlying Eq. (22) and the CRD case are analogous.

We use, instead, an extension of the one–point approximate operator technique proposed in Olson et al. (1986). In that technique, one assumes the distribution of S has been discretized and the spatially continuous Λ operator has been replaced by a discrete matrix operator, i.e $L(S) \approx \sum L_{d,d'} S_{d'}$. When iterating, the full matrix is simply approximated at each frequency by the diagonal of the full matrix. Thus, in order to find a better solution to Eq. (23), we correct the current estimate of S^L by letting $S^L \rightarrow S^L + \delta S^L$, approximating the full L operator by its diagonal and then solving for the correction to S^L. If this is done, we will have *at each spatial grid point*

$$S^L + \delta S^L = (1-\epsilon)L(S^L) + (1-\epsilon)L_*(\delta S^L) + r \tag{24}$$

which gives for each point the correction to the current estimate of the S^L

$$\delta S^L = \frac{(1-\epsilon)L(S^L) - S^L + r}{1 - (1-\epsilon)L_*} \tag{25}$$

The diagonal one–point approximation to $L()$ is derived by using the diagonal of the frequency dependent Λ_ν operator appearing in Eq. (17), weighting it by the value of the appropriate combination of the profile function and opacities and then integrating over frequency

$$L_* = \int \frac{\chi_L \phi_\nu^2}{\chi_L \phi_\nu + \kappa} \Lambda^*_{\nu'} d\nu' \tag{26}$$

where we have designated the diagonal of the monochromatic Λ operator by Λ^*_ν.

The "inner iteration loop" consists of using the approximate operator from Eq. (26) in Eq. (25) in order to find an adequately corrected S^L satisfying Eq. (22) for the current r, defined by Eq. (23). The run of r is held fixed in this inner loop. Once the new S^L has been determined, it is substituted into Eq. (19) to get an estimate for the frequency dependent source function, S^L_ν, which in turn is used to compute the values of $\bar{J}_\nu$ and ΔS^L_ν. This constitutes the "outer iteration". The convergence of the predicted values of the ΔS^L_ν is accelerated by exactly the same residual minimization techniques as are described in Auer (1991) for the CRD problem although here they are applied to the full frequency dependent quantity.

For maximum efficiency it is essential to note that the "inner loop" is computing a correction to a correction, and, therefore, does not need to be computed to a high degree of accuracy. In fact, we have found that one needs only about 3 accelerated Λ–iterations, Eq. (25), in the inner loop, in order to achieve the best possible results for the overall convergence. That is, even if one were to iterate the inner loop to "perfect" convergence, the outer loop would not converge any faster. The convergence rate for the PRD source function, which is being found in the outer loop, is roughly the same as for the CRD case. Although the increase in the number of Λ–iterations is non–negligible, the factor is relatively small, on the order of three times as many. The speed of the PRD calculation becomes even more impressive when one notes that one is now solving for $N_{frequency}$ times as many values. That is, rather than a single frequency independent source function, we are determining the full frequency dependent PRD source function for only a small factor more computation time.

It is important to note that there is a serious potential difficulty with the PRD iteration technique we have presented here. In a low density plasma, a resonance line like Lyman α scatters coherently in the atomic frame. The only source of frequency redistribution in the laboratory frame is the physical motion of the atom and the change in Doppler shift between absorption and emission. The frequency shifts are on the order of a ± 1 Doppler width. If the medium is still optically thick many Doppler widths from line center, the scattering in the far wings becomes approximately coherent in the laboratory frame. The L_* operator which is based on complete redistribution is obviously a poor approximation for the far wing. Fortunately if one is dealing with slabs of finite optical thickness, this potential problem may not actually cause difficulty. In Fig. 10 the degree of coherence is shown for a (Voigt) line profile with $a = 10^{-4}$. The abscissa, ϕ_ν^{-1}, is the ratio of the optical depth at frequency

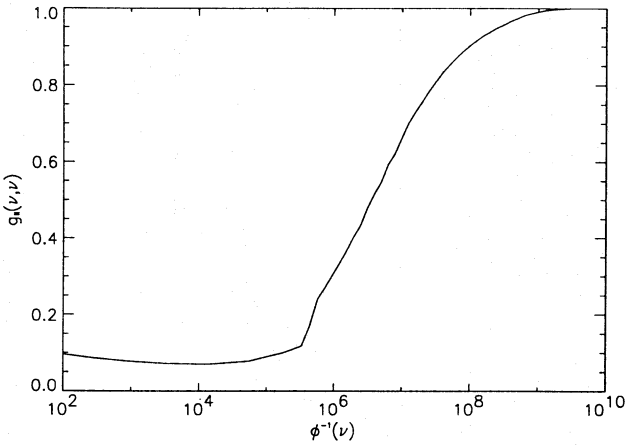


Fig. 10. If a slab is optically thick enough, scattering in the far wing becomes nearly coherent. In this figure, we show the degree of coherence at the optically thick wing frequency as a function of line center optical depth

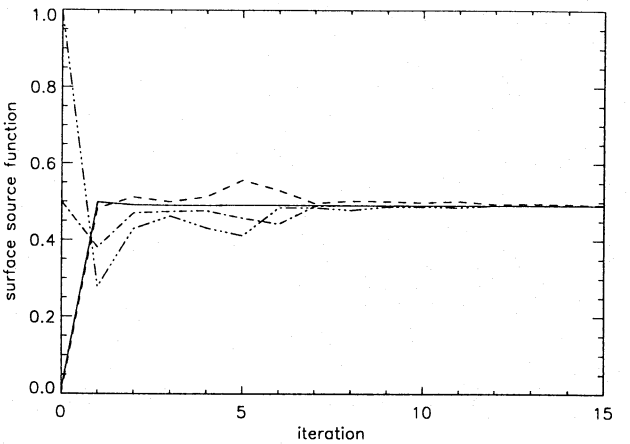


Fig. 11. Even 40 Doppler units from line core where the scattering is very coherent, there is no difficulty in convergence for optical depths like 10^6

ν to the total optical depth. If a finite slab becomes optically thin at a small enough frequency, the asymptotic coherence of the redistribution does not present any difficulty. Similar remarks apply to the semi–infinite case with either an overlapping continuum or a non–zero collision rate. Here the relevant quantity is the coherence at frequencies which become optically thick before they thermalize or are dominated by the continuum opacity. In all these cases the behavior in the far wing will not cause any convergence problem.

We can demonstrate the validity of these remarks by showing the convergence in the far wing. If the wings were optically thick at frequencies where the scattering is nearly coherent, the values would remain near the original estimates. In fact as we show in Fig. 11, this may not occur even if core optical depth seems to be large. The example shown is appropriate for solar prominences, with $\epsilon = 10^{-4}$, a Voigt profile with $a = 10^{-4}$ and optical depth at line center $\tau_0 = 10^6$. As starting approximations we have used $S_d(\nu) = S_{CRD}, 0, 0.5, 1 \times B$. No matter what

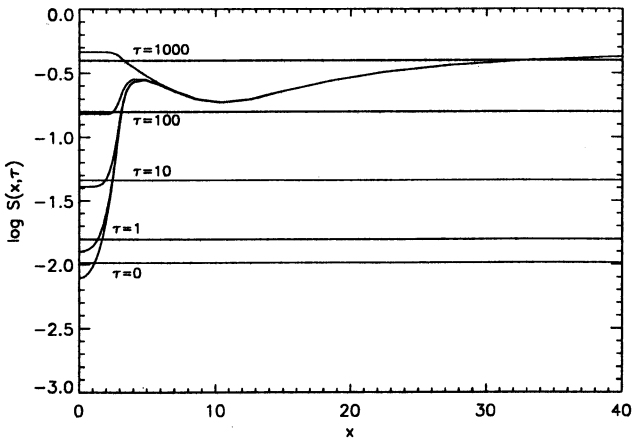


Fig. 12. Frequency–depth dependence of the PRD and CRD source functions for 1D, $\tau = 10^6$ and $\epsilon = 10^{-4}$ case solved using the new iterative technique

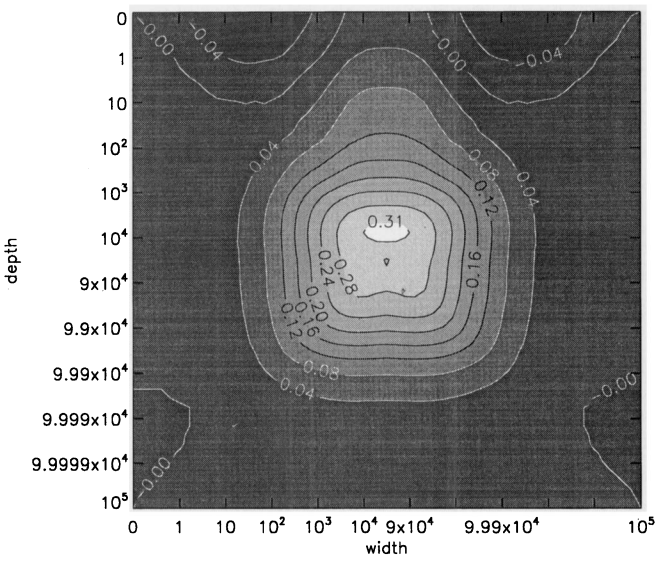


Fig. 13. Contour diagram of $\log(S_{\rm PRD}(x)/S_{\rm CRD})$ at line center, for two–dimensional transfer with $\epsilon = 10^{-4}$

the initial value, the correct solution is obtained. The point is that even for this large core optical depth, the wings at frequencies where the scattering is nearly coherent are optically thin (for $x = 40$, $\tau_x \approx 0.01$) and Λ–iteration is entirely adequate to solve the scattering problem. Equally good convergence is obtained nearer line core. Here there is substantial frequency redistribution and the OAB approximate operator takes care of any convergence difficulties. On the other hand, if the line is, in fact, optically thick at frequencies where the scattering is nearly coherent, convergence difficulties will occur. For example, if we had used $\tau_0 = 10^8$ and the above parameters there are difficulties. To treat such extreme cases a more robust approach, which we present in future work, is required.

In order to confirm our method for treating PRD effects, we have made extensive comparisons with the results of P. Gouttebroze (1991, private communication) who used the discrete ordinate method and finds results in complete agreement with our own. In order to permit comparison with published results, we show in Fig. 12 the 1D source functions for the cases of R_{II} redistribution and complete redistribution for a finite slab of optical thickness $\tau = 10^6$ and $\epsilon = 10^{-4}$. The case presented here is one treated by Hummer (1969; Fig. 3c). The PRD source functions appear as curved full lines and, because of its frequency independence, the CRD solution appears as horizontal lines. Although our frequency grid spans 40 Doppler widths instead of less than 21 as in Hummer's study, we recover the same major effects. The PRD source function at line center is roughly equal to the CRD value throughout the slab. In greater detail, however, the PRD solution at line center is smaller at the surface than the CRD value $\sqrt{\epsilon}B$; as also appears in Hummer's results. At line center, the PRD solution thermalizes "sooner" than in CRD. This is because the thermalization length scales as ϵ^{-1} for PRD–$R_{\rm II}$ rather than as $a\epsilon^{-2}$ for Voigt–CRD. In the present case, these values are respectively 10^4 for PRD and 10^5 for CRD. In the wings, completely unlike CRD, the PRD source function, which is dominated by the incident radiation, is not depth dependent.

7. Comparison of PRD and CRD effects in a sample 2D case

In this section, we apply the new iterative technique for the solution of radiative transfer problems in two dimensions to an example characteristic of a solar prominence and compare the results obtained under the assumptions of partial and complete redistribution. We assume an isothermal square slab with total optical depth equal to 10^5 on both sides. The spatial grid has 55 points in each directions and the frequency grid spans 20 Doppler widths from line center. The Planck function B is set to unity throughout the slab. For the line, we adopt a thermalization parameter $\epsilon = 10^{-4}$, a Voigt parameter $a = 10^{-3}$ and a continuum to line opacity ratio $r = 10^{-7}$. The slab is irradiated on both sides and the bottom, by an incident profile with $I(x) = B = 1$ for $\mid x \mid \leq 4$ and 0 farther in the wings. This roughly mimics the interaction between a solar prominence and the Lyman α radiation incident from the chromosphere. The PRD case is solved using $R_{\rm II}$ redistribution which applies when there is coherent scattering in the atom's frame (Hummer 1963).

In order to display the 2D–PRD effects, we have plotted $\log\,(S_{\rm PRD}(x)/S_{\rm CRD})$ at various frequencies from line core to line wing in Figs. 13 to 15. Even at line center (Fig. 13), we can see that PRD effects are non–negligible. At this frequency, deviations from CRD are largest at slab center. In fact, the PRD source function may be twice as large as the CRD one. This PRD effect of enhancement of the source function at slab center has already been noticed in 1D–PRD modeling of prominences (Heinzel et al. 1987) and persists, in a modified form, in these 2D calculations. (As a consequence, the excitation of hydrogen is predicted higher in the central part of a prominence by PRD than by CRD). The effect of quasi–coherent penetration of near wing photons is larger in illuminated 2D slabs than 1D slabs, since they have a greater area receiving incident radiation (Paletou et al. 1993). Near the non–illuminated top boundary, we predict small regions where the PRD solution is less than CRD (the

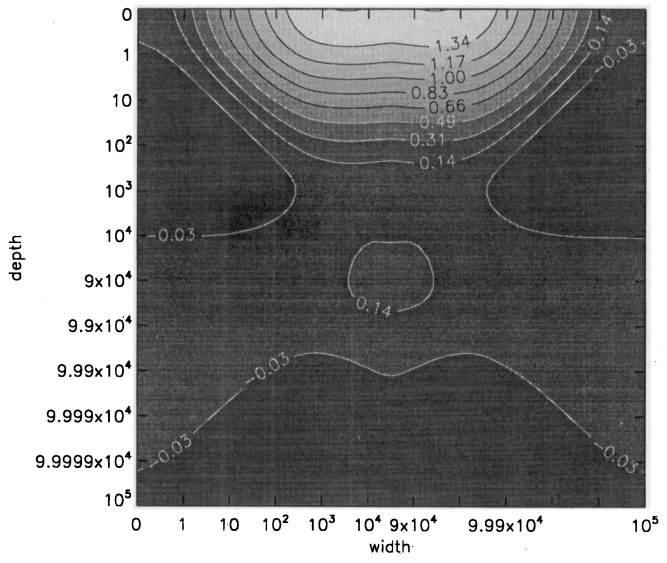


Fig. 14. Contour diagram of $\log(S_{\rm PRD}(x)/S_{\rm CRD})$, 4 Doppler widths from line center

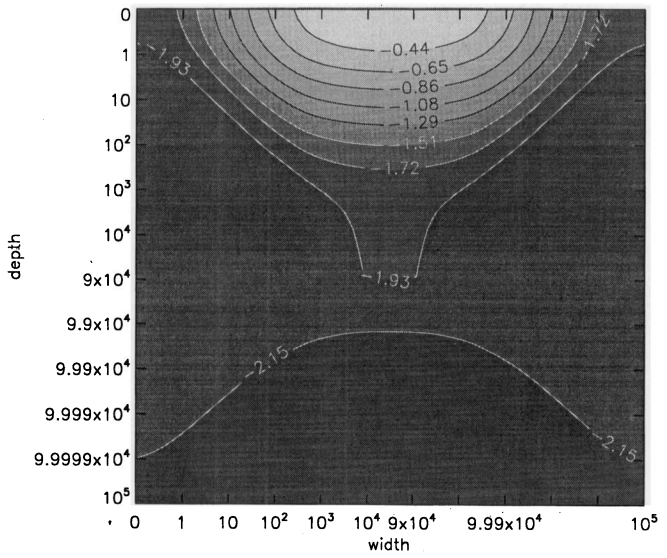
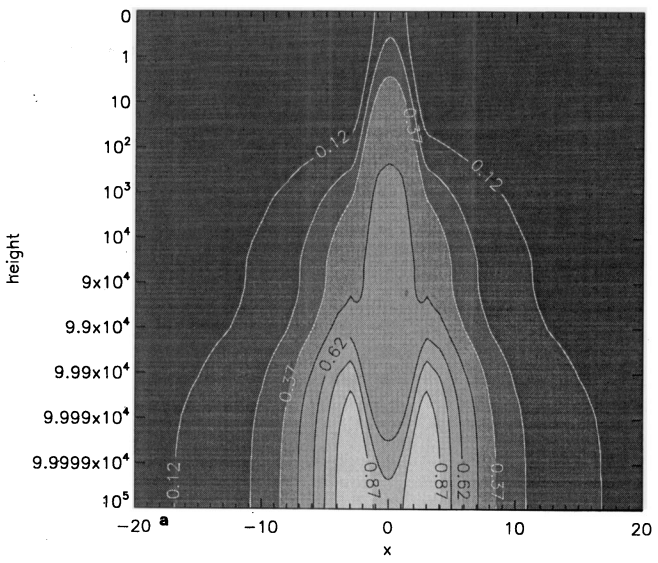


Fig. 15. Contour diagram of $\log(S_{\rm PRD}(x)/S_{\rm CRD})$, 20 Doppler widths from line center



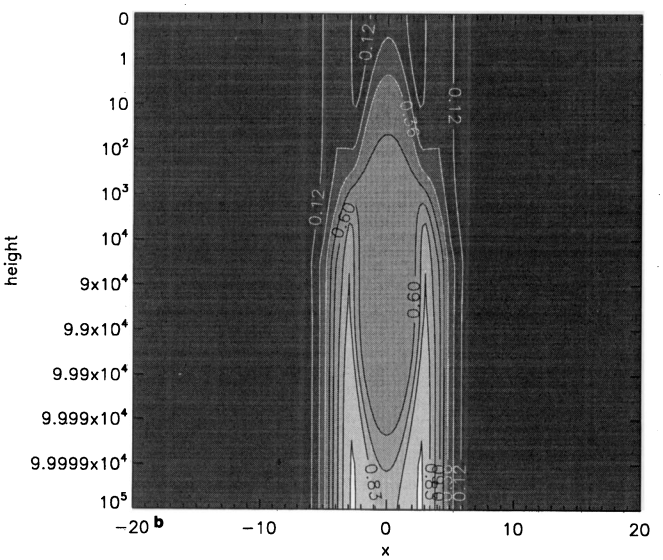


Fig. 16a and b. Contour diagrams of the emergent intensities for the slab as seen on the limb **a** in CRD and **b** in PRD (below)

maximum deviation is smaller than 20%). For frequencies with $|\ x\ | \leq 4$, deviations from the CRD solution are very small near the illuminated boundaries ($S_{\rm PRD}(x)/S_{\rm CRD} \approx 1$), because the incident radiation dominates the radiation for both CRD and PRD in these layers. Further, when observing the slab "on the limb" (i.e. along the horizontal axis), line center emergent intensities for both kinds of frequency redistribution are almost identical (see Fig. 16). This is because the regions where the PRD line center source function differs significantly from CRD are too optically thick to affect the emergent intensity.

Farther in the line wing, Fig. 14 shows the same source functions ratio at a characteristic frequency, 4 Doppler widths from line center, where the $R_{\rm II}$ redistribution function begins to show the quasi–coherent nature of wing photon scattering (i.e. the redistribution function shows a single peak, centered at the absorption frequency). Needless to say, the mathematics valid for this frequency regime is far different from that underlying CRD and the discrepancies between the PRD and CRD solutions are rather large. This is especially true in the uppermost layers of the slab, along the vertical axis of symmetry. The PRD source function can be a factor of 30 more than CRD. Indeed, at this frequency we have both a preferential re–emission of photons at the absorption frequency, instead of at line center as described by CRD, and the external incident radiation is still non–zero. The difference between PRD and CRD is particularly apparent when one compares the emergent profiles for lines of sight hitting the top layers of the slab, i.e. from above, as displayed in Fig. 16. The CRD profiles are Gaussian while, in PRD, secondary peaks are predicted with maxima 4 Doppler widths from line center.

The $\log(S_{\rm PRD}(x)/S_{\rm CRD})$ ratio, 20 Doppler widths from line center is plotted in Fig. 15. There are major differences in the calculated values of the source functions. The PRD source function is lower than the CRD one at all depths in the slab. The ratio is on the order of 10^{-2} over practically the entire slab, and the smallest deviation is still larger than about 40%. We recover here a well–known consequence of the PRD–$R_{\rm II}$ which predicts rapidly decreasing line wings while CRD predicts broad and intense ones (cf. Milkey et al. 1979 and Fig. 16).

The predicted emergent profiles from the slab, observed on the limb, for the two kinds of redistribution are shown in Fig. 16. The difference in variation with height is striking and cannot be reproduced by 2D– or PRD–only models. It is caused by an interaction of both of these effects. For the 2D–CRD case, the finite thickness of the slab produces a double peaked emergent profile, which disappears as losses through the upper boundary reduce the source function. PRD aggravates both the size and persistence of the self–reversal as well as suppressing the wings at all heights. Comparing the two figures, we can see that, in PRD, the central reversal of the line is predicted from the bottom all the way into the central part of the slab. Further, the peak–to–peak distance in the emergent profile is comparable to the width of the incident profile (± 4 Doppler widths). This is unlike CRD, for which the central reversal occurs only in the lowest layers of the slab. Further as mentioned above, in the description of the PRD effects on the source function, secondary emission peaks in the PRD profiles clearly appear in the top layers of the slab where the CRD ones are simply Gaussian.

The global height variations of the CRD emergent profile are indeed as described by Vial (1982) in his study of the 2D Lyman α profile emitted from prominences; however, from this example we can see that there are important combined 2D and PRD effects which should be treated simultaneously and consistently in the modeling of this optically thick line in solar prominences. The nature and the importance of the PRD effects is enhanced for 2D slabs which may be dominated by incident radiation and provide more escape opportunities than 1D.

8. Conclusions

The short characteristic approach provides a rapid and accurate method for the computation of the radiation field in the presence of scattering in a two–dimensional medium. The approach has been extended to include the effects of partial frequency redistribution at a modest computational cost. While the PRD approach is not valid for extremely large optical depths, it is applicable to many important astrophysical problems. In fact, convergence difficulties do not occur unless the optical depths are much larger than 10^7 at line center in one of the finite coordinates. Further, techniques which overcome this potential difficulty are under development at the current time.

Our first application of the method is an extension of the work of Vial (1982) to realistic simulations of resonance lines like H I Lyman α, Mg II h and k, Ca II H and K in prominences. For all these lines, both 2D and PRD effects may be expected

Acknowledgements. Most of the 2D computations were performed using the facilities of the *"Centre de Calcul Vectoriel pour la Recherche"* in Palaiseau (France). This work has been supported by the Los Alamos National Laboratory which is operated by the University of California for the U.S. Department of Energy. In addition, we would like to thank J.–C. Vial and P. Gouttebroze for helpful discussions.

References

Auer, L.H.: 1987, in *Numerical Radiative Transfer*, ed. W. Kalkofen, Cambridge University Press
Auer, L.H.: 1991, in *Stellar Atmospheres: Beyond Classical Models*, eds. L. Crivellari, I. Hubený and D.G. Hummer, NATO ASI Series, Kluwer
Auer, L.H., Fabiani, P., Trujillo Bueno, J.: 1993, *in preparation*
Boris, J.P., Book, D.L.: 1973, J. Comp. Phys. 11, 38
Heinzel, P., Gouttebroze, P., Vial, J.–C.: 1987, A&A 183, 351
Hummer, D.G.: 1963, MNRAS 125, 21
Hummer, D.G.: 1969, MNRAS 145, 95
Kunasz, P.B., Auer, L.H.: 1988, J. Quant. Spectrosc. Radiat. Transfer 39, 67
Mihalas, D., Auer, L.H., Mihalas, B.R.: 1978, ApJ 220, 1001
Milkey, R.W., Heasley, J.N., Schmahl, E.J., Engvold, O.: 1979, in *Physics of Solar Prominences*, IAU Colloquium 44, eds. E. Jensen, P. Maltby and F.Q. Orall, Dordrecht, Reidel, p.53
Ng, K.C.: 1974, J. Chem. Phys. 61, 2680
Olson, G.L., Auer, L.H., Buchler, J.R.: 1986, J. Quant. Spectrosc. Radiat. Transfer 35, 431
Paletou, F.: 1992, Ph.D. thesis, Institut d'Astrophysique Spatiale, Université de Paris XI
Paletou, F., Vial, J.–C., Auer, L.H.: 1993, A&A 274, 571
Scharmer, G.B.: 1983, A&A 117, 83
van Leer, : 1977, J. Comp. Phys. 23, 276
Vial, J.–C.: 1982, ApJ 254, 780

This article was processed by the author using Springer-Verlag LaTeX A&A style file version 3.