Numerical performance of stability enhancing and speed increasing steps in radiative transfer solution methods☆

Per Edström

Department of Natural Sciences, Engineering and Mathematics, Mid Sweden University, SE-87188 Härnösand, Sweden

ABSTRACT

Methods for solving the radiative transfer problem, which is crucial for a number of sectors of industry, involve several numerical challenges. This paper gives a systematic presentation of *the effect* of the steps that are needed or possible to make *any* discrete ordinate radiative transfer solution method numerically efficient. This is done through studies of the numerical performance of the stability enhancing and speed increasing steps used in modern tools like DISORT or DORT2002.

Performance tests illustrate the effect of steps that are taken to improve the stability and speed. It is shown how the steps together give a stable solution procedure to a problem previously considered numerically intractable, and how they together decrease the computation time compared to a naive implementation with a factor 1000 in typical cases and far beyond in extreme cases. It is also shown that the speed increasing steps are not introduced at the cost of reduced accuracy. Further studies and developments, which can have a positive impact on computation time, are suggested.

© 2008 Elsevier B.V. All rights reserved.

1. Introduction

Radiative transfer solution methods are important tools for modelling the interaction of radiation with turbid (scattering and absorbing) media. Applications range from stellar atmospheres and infrared and visible light in space and in the atmosphere, to optical tomography and diffusion of neutrons. An industrially important application is light scattering in textile, paint, pigment films, paper and print, and accurate calculation methods are crucial for these sectors of industry.

Discrete ordinate solution methods for radiative transfer problems have been studied throughout the last century. In the beginning most radiative transfer problems were considered intractable because of numerical difficulties. Therefore coarse approximations were used, and methods developed slowly due to the lack of mathematical tools. The first approximate solution to the radiative transfer problem was presented by Schuster [11], and Wick [20] gave the first general treatment of discrete ordinate methods. Chandrasekhar described a method using spherical harmonics [1], but having read Wick's article, he adopted the discrete ordinate method, and further refined it [2]. Later, he wrote a classic exposition on radiative transfer theory in book form [3], and since then the area has expanded tremendously. Mudgett and Richards [8,9] described a discrete ordinate method for use in technology, and reported on numerical difficulties, as have many before and after them.

Through a great effort, ranging over several years, Stamnes and coworkers [17,15] presented in a series of papers the successive development of a stable discrete ordinate algorithm, and provided a software package, DISORT. Thomas and

---

Stamnes [19] also wrote a textbook on radiative transfer in the atmosphere. In a recent paper, Edström [4] presented a systematic review of the stability enhancing and speed increasing steps used in modern discrete ordinate radiative transfer algorithms. Edström also presented the solution method DORT2002, which is adapted to light scattering simulations in paper and print, but which is also designed for methodical numerical experiments through its modularized design and ability to give any kind of intermediate results and performance data.

The point of this paper is to give a systematic presentation of *the effect* of the steps that are needed or possible to make *any* discrete ordinate radiative transfer solution method numerically efficient, and in particular the effect of the most important steps used in modern tools like DISORT or DORT2002. To the author's knowledge, this has not been summarized in one single publication before, in particular not with the focus on quantifying the effect of the steps.

First, a short overview of a generic solution method is given. Then the resulting improvements, quantified in terms of reduced condition number and increased speed compared to a naive implementation, are illustrated. The speed increasing steps are also analyzed to verify that speed is not introduced at the cost of reduced accuracy. Finally, some studies and developments that can have a positive impact on computation time are suggested.

2. Solution method overview

This section gives a short introduction to the radiative transfer problem, and the structure of a modern generic discrete ordinate solution method.

Edström [4] states the equation of radiative transfer as

$$u\frac{\mathrm{d}I(\tau, u, \varphi)}{\mathrm{d}\tau} = I(\tau, u, \varphi) - \frac{a}{4\pi}\int_0^{2\pi}\int_{-1}^1 p(u', \varphi'; u, \varphi)I(\tau, u', \varphi')\mathrm{d}u'\mathrm{d}\varphi'. \tag{1}$$

The unknown intensity, I, at optical depth τ is considered as non-interacting beams of radiation in all directions. The phase function, p, specifies the probability distribution of scattering from incident direction (u', φ') to direction (u, φ), where u is cosine of polar angle, and φ is azimuthal angle. The shape of the phase function may be controlled by a parameter called the asymmetry factor, g, ranging from complete forward scattering $(g = 1)$ over isotropic scattering $(g = 0)$ to complete backward scattering $(g = -1)$, or it may be defined by any number of discrete phase space moments. The single scattering albedo, a, is the probability for scattering given an extinction event, and is defined as $a = \sigma_s/(\sigma_a + \sigma_s)$, where σ_s and σ_a are the scattering and absorption coefficients of the medium. The first term on the right-hand side in the radiative transfer equation (1) thus gives intensity absorbed when traversing a thickness $\mathrm{d}\tau$, while the integral term gives the intensity scattered from all incoming directions at a point to a specified direction.

The common procedure in discrete ordinate methods is to use Fourier analysis on the azimuthal angle to turn the integro-differential equation (1) into a number of uncoupled equations, one for each Fourier component of the unknown intensity, which are then discretized using numerical quadrature. By introducing the function

$$p^m(u', u) = \sum_{l=m}^{2N-1}(2l+1)\chi_l\Lambda_l^m(u')\Lambda_l^m(u),$$

where the χ_l are Legendre expansion coefficients and the $\Lambda_l^m(u)$ are normalized associated Legendre functions, the phase function can be expressed in a Fourier cosine series as

$$p(u', \varphi'; u, \varphi) = \sum_{m=0}^{2N-1}(2 - \delta_{0m})p^m(u', u)\cos(m(\varphi' - \varphi)),$$

and the intensity can be expanded in a similar way as

$$I(\tau, u, \varphi) = \sum_{m=0}^{2N-1}I^m(\tau, u)\cos(m(\varphi_0 - \varphi)),$$

where $2N$ is the number of quadrature points. The Double-Gauss quadrature formula, proposed by Sykes [18], approximates an integral over the two hemispheres separately,

$$\int_{-1}^1 f(u)\mathrm{d}u = \int_0^1 f^+(\mu)\mathrm{d}\mu + \int_0^1 f^-(\mu)\mathrm{d}\mu \approx \sum_{j=1}^N \omega_j f^+(\mu_j) + \sum_{j=1}^N \omega_j f^-(\mu_j),$$

where the quadrature points μ_j and weights ω_j are chosen for the "half interval" [0, 1] according to ordinary Gaussian quadrature on the interval $0 \le \mu \le 1$ (the plus and minus signs designate quantities in the upper and lower hemispheres). Application of the Double-Gauss quadrature rule is what is called the discrete ordinate approximation, and this gives for each Fourier component (where the superscript m has been dropped)

$$\begin{cases} \mu_i \dfrac{\mathrm{d}I^+(\tau,\mu_i)}{\mathrm{d}\tau} = I^+(\tau,\mu_i) - \dfrac{a}{2}\sum_{j=1}^N \omega_j p(\mu_j,\mu_i)I^+(\tau,\mu_j) \\[2mm] \qquad - \dfrac{a}{2}\sum_{j=1}^N \omega_j p(-\mu_j,\mu_i)I^-(\tau,\mu_j) - X_{0i}^+ \mathrm{e}^{-\tau/\mu_0} \\[2mm] -\mu_i \dfrac{\mathrm{d}I^-(\tau,\mu_i)}{\mathrm{d}\tau} = I^-(\tau,\mu_i) - \dfrac{a}{2}\sum_{j=1}^N \omega_j p(\mu_j,-\mu_i)I^+(\tau,\mu_j) \\[2mm] \qquad - \dfrac{a}{2}\sum_{j=1}^N \omega_j p(-\mu_j,-\mu_i)I^-(\tau,\mu_j) - X_{0i}^- \mathrm{e}^{-\tau/\mu_0}, \end{cases} \quad i = 1,\dots,N,$$

which is a system of first-order linear differential equations. Here, $X^\pm$ pertains to the particular solution. This system can be put into block matrix form as

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\begin{bmatrix}\mathbf{I}^+ \\ \mathbf{I}^-\end{bmatrix} = \begin{bmatrix}-\alpha & -\beta \\ \beta & \alpha\end{bmatrix}\begin{bmatrix}\mathbf{I}^+ \\ \mathbf{I}^-\end{bmatrix} - \begin{bmatrix}\mathbf{Q}^+ \\ \mathbf{Q}^-\end{bmatrix}, \tag{2}$$

where $\mathbf{I}^\pm = \{I^\pm(\tau,\mu_i)\}$, and where $\mathbf{Q}^\pm$ pertains to the particular solution. The block matrices α and β are given as the solutions to $\mathbf{M}\alpha = \frac{a}{2}\mathbf{D}^+\mathbf{W} - \mathbf{1}$ and $\mathbf{M}\beta = \frac{a}{2}\mathbf{D}^-\mathbf{W}$, respectively, where $\mathbf{M} = \{\mu_i\delta_{ij}\}$, $\mathbf{W} = \{\omega_i\delta_{ij}\}$, and $\mathbf{D}^\pm = \{p^m(\pm\mu_j,+\mu_i)\} = \{p^m(\mp\mu_j,-\mu_i)\}$. The identity matrix is denoted by $\mathbf{1}$, δ_{ij} is the Kronecker delta, and $i,j = 1,\dots,N$.

It is well known that the homogeneous solutions to systems of coupled ordinary differential equations such as (2) are of the form $\mathbf{I}^\pm = \mathbf{G}^\pm \mathrm{e}^{-k\tau}$. This gives the eigenvalue problem

$$\begin{bmatrix}\alpha & \beta \\ -\beta & -\alpha\end{bmatrix}\begin{bmatrix}\mathbf{G}^+ \\ \mathbf{G}^-\end{bmatrix} = k\begin{bmatrix}\mathbf{G}^+ \\ \mathbf{G}^-\end{bmatrix} \tag{3}$$

to solve for the eigenvalues k and the eigenvectors $\mathbf{G}^\pm$. Then boundary and continuity conditions need to be treated, as well as the problem of extending the computed intensity from the quadrature points to the entire interval through interpolation formulas.

Two variables determine the size of the problem. The number of layers in a multilayer medium is denoted by L, and the number of terms in the numerical quadrature is denoted by N. The underlying physical problem usually gives L, while N can be freely chosen, larger N giving higher accuracy. The N needed to achieve a given accuracy depends primarily on the phase function. A sharply peaked phase function needs a large number of terms in its Legendre function expansion [21], and a comparable number of terms are needed in the numerical quadrature. In some places the quantity $2N$, which corresponds to the notion of 'streams' or 'channels' in many applications, is used. The flowchart below describes the overall structure of the solution method.

Algorithm 1.

for all Fourier components, $m = 0\dots 2N-1$
    for all layers, $q = 1\dots L$
        Solve system of ODEs through an eigenvalue problem.
        Represent homogeneous solution by a linear combination of the
        eigensolutions with coefficients C_{jq}, where $j = \pm 1,\dots,\pm N$
        is an enumeration of the eigensolutions.
        Compute particular solution.
    end
    Apply boundary and continuity conditions to obtain C_{jq}.
    Assemble mth Fourier component of the intensity.
    if convergence criterion is met
        Break loop over Fourier components.
    end
end
Assemble total intensity as sum of Fourier components.
Apply interpolation formulas.

3. Effect of stability enhancing steps

There are a lot of numerical difficulties in radiative transfer problems, and therefore a solution method needs to include several steps that improve stability. Some of them have an obvious effect in a limited part of the method and need no further investigation, while others have a more profound influence on the stability of the overall method. Investigations of

condition number and related issues are relevant for the two core problems. The conditioning of the eigenvalue problem is treated in Section 5, while the following subsection covers the system of equations corresponding to the boundary and continuity conditions (after the *layer* loop in *Algorithm 1*). These sub-problems are solved independently for each of the azimuthal Fourier components of the unknown intensity, since they are entirely uncoupled. When investigating the properties mentioned, it is therefore relevant to do so for different Fourier component numbers, as well as for different N and L and different media properties.

3.1. Conditioning of the system of equations for the boundary and continuity conditions

The system of ODEs for each Fourier component (2) is solved through the eigenvalue problem (3) for each layer (in the *layer* loop in *Algorithm 1*), and the homogeneous solution is represented by a linear combination of the eigensolutions. The unknown coefficients C_{jq}, $j = \pm 1, \ldots, \pm N$, $q = 1, \ldots, L$ of the linear combination in the multilayer solution are given by boundary and continuity conditions (after the *layer* loop in *Algorithm 1*). They constitute a $(2N \times L) \times (2N \times L)$ system of equations, given by

$$
\begin{cases}
\displaystyle\sum_{j=1}^{N} \left(C_{j1} G_{j1}(-\mu_i) + C_{-j1} G_{-j1}(-\mu_i) \right) = \mathcal{L}(-\mu_i) - U_1(0, -\mu_i), \quad i = 1, \ldots, N \\[2ex]
\displaystyle\sum_{j=1}^{N} \left\{ \left(C_{jq} G_{jq}(\mu_i) e^{-k_{jq}\tau_q} + C_{-jq} G_{-jq}(\mu_i) e^{+k_{jq}\tau_q} \right) - \left(C_{j,q+1} G_{j,q+1}(\mu_i) e^{-k_{j,q+1}\tau_q} + C_{-j,q+1} G_{-j,q+1}(\mu_i) e^{+k_{j,q+1}\tau_q} \right) \right\} \\[1ex]
\qquad = U_{q+1}(\tau_q, \mu_i) - U_q(\tau_q, \mu_i), \quad i = \pm 1, \ldots, \pm N, \; q = 1, \ldots, L-1 \\[2ex]
\displaystyle\sum_{j=1}^{N} \left(C_{jL} r_j(\mu_i) e^{-k_{jL}\tau_L} + C_{-jL} r_{-j}(\mu_i) e^{+k_{jL}\tau_L} \right) = \Gamma(\tau_L, \mu_i), \quad i = 1, \ldots, N,
\end{cases}
\tag{4}
$$

where $\mathcal{L}$ and Γ constitute the boundary conditions, the U_q are the particular solutions of the ODE system (2), and the r_j are a variant of the eigenvectors G_j at the lower boundary. The coefficient matrix is sparse and block diagonal, with $6N - 1$ diagonals (see Fig. 4). These equations are ill-conditioned, which is why the method was considered to be numerically intractable in the past, and consequently discarded. The ill-conditioning is removed by using a preconditioner suggested in [12], the scaling transformation $C_{+jq} = C'_{+jq} e^{k_{jq}\tau_{q-1}}$, and $C_{-jq} = C'_{-jq} e^{-k_{jq}\tau_q}$, where k_{jq} are the eigenvalues (known from the core eigenvalue problems) and τ_q is the optical depth at the bottom of layer q. The problem of solving for the scaled coefficients C'_{jq} is unconditionally stable [12]. It should be pointed out that it is essential to use the scaled coefficients in the rest of the solution procedure. Otherwise, the use of some re-scaling transformation would introduce a risk of enlarging errors later.

3.2. Numerical experiments

The DISORT package includes a suite of published test problems [22], which cover most variations of normal cases as well as the most interesting extreme cases, and which include high-dimensional discrete phase spaces. The parameters from test problems 1–9 were used in the tests.

The sensitivity of the solution of the system of equations corresponding to boundary and continuity conditions (4) was measured using the 2-norm condition number, defined as $\|A\|_2 \|A^{-1}\|_2$ for a matrix A (see e.g. [5, Section 2.7.2]). The condition number and the sparsity pattern of the coefficient matrix for this system were studied with DORT2002 for a wide range of parameter sets (using the test problems [22]), with and without the preconditioner, and with varying numbers of layers and Fourier component numbers. All different parameter sets gave very similar results, so only the results of one set, which is representative for any of the other sets, are presented here. The parameters of the presented test case were the following. The illumination was a combination of diffuse light of intensity 0.3, and a beam of intensity 1.0 with polar angle cosine of 0.5 and azimuthal angle of $\pi/2$. The depth at the upper boundary was 0, and layer number q (using 1–10 layers) had a thickness of $0.02/q$. The underlying surface was diffuse, with a reflectance of 0.5. The medium had a scattering coefficient of 100 and an absorption coefficient of 10, and had a Henyey–Greenstein [6] phase function with an asymmetry factor of 0.5. The calculations used $2N = 40$.

3.3. Results

Fig. 1 shows the condition number of the coefficient matrix for different Fourier component numbers and different numbers of layers after applying the preconditioner. Fig. 2 is a slice from Fig. 1 for the single layer case. The plots show that the preconditioner works very well, giving a condition number close to 1 in most cases, and around 30 in the worst case. Fig. 3 is the same as Fig. 2, but without the preconditioner (note the factor 10^{277} on the condition number axis scale), and shows that the problem is very ill-conditioned without the preconditioner, having a condition number near the largest positive floating-point number for the system.
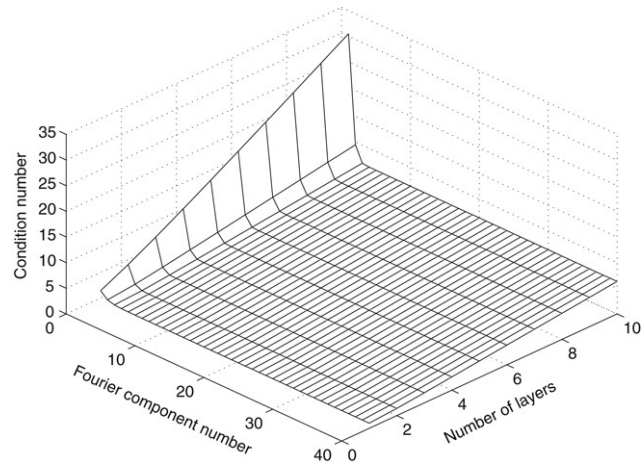
Fig. 1. Condition number of the coefficient matrix after applying the preconditioner. The condition number is close to 1 in most cases, and moderate for the 0th and 1st Fourier components. The condition number increases slowly with increasing number of layers.
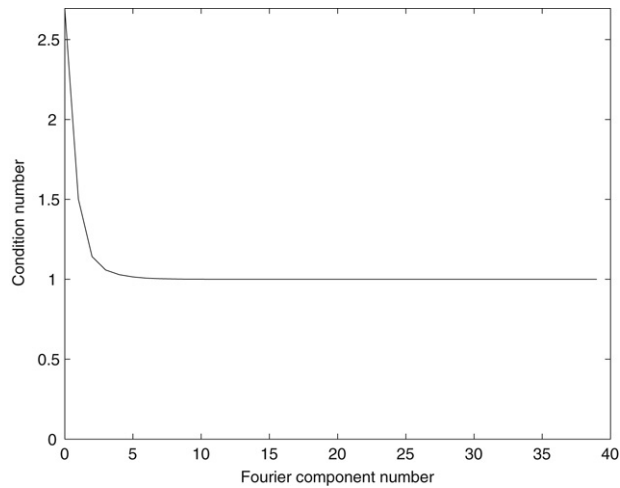

Fig. 2. Condition number of the coefficient matrix in the single layer case after applying the preconditioner. The condition number rapidly decreases to 1 with increasing Fourier component number.
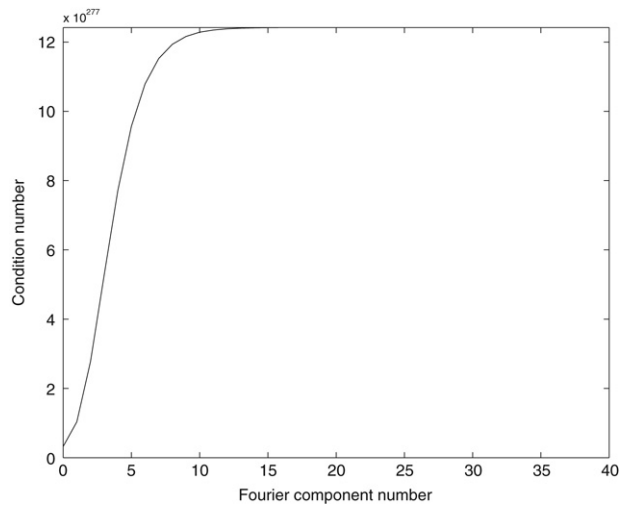

Fig. 3. Condition number of the coefficient matrix in the single layer case without the preconditioner. Note the factor 10^{277} on the condition number axis scale.
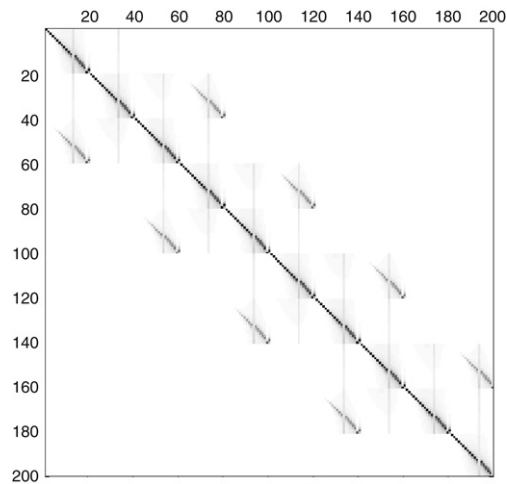
Fig. 4. Sparsity pattern of the coefficient matrix for the 0th Fourier component with $2N = 40$ and $L = 5$ after applying the preconditioner. Note the sparse block diagonal structure with the large elements (black in the gray-scale) along the diagonal, indicating a well-conditioned system of equations.

The sparsity pattern for the coefficient matrix after applying the preconditioner was generated with $2N = 40$ and $L = 5$. This gives a 200×200 coefficient matrix, and the single layer case can be obtained by extracting the top left 40×40 sub-matrix. Fig. 4 shows the coefficient matrix for the 0th Fourier component, which has a special structure since it is the azimuthally averaged case. For all other Fourier component numbers, the structure is far sparser, although the bandwidth $6N - 1$ is the same.

4. Effect of speed increasing steps

Several steps are needed to increase the speed of the method. Among the obvious steps are code optimization, the use of efficient solvers for the eigenvalue and system of equations problems, and the correct handling and exploitation of the sparse structure of the systems of equations corresponding to the boundary and continuity conditions. Other steps have a more profound influence on the speed of the overall method. This chapter covers investigations of speed increase from eigenvalue problem size reduction, from methods that maintain accuracy for a lower number of terms in the quadrature, and from methods that terminate calculations on earlier convergence.

4.1. Eigenvalue problem size reduction

The eigenvalue problem is an important part of the core of the method, and it is solved in the innermost loop (in the *layer* loop in *Algorithm 1*). Any improvement in speed there will have a large effect on the overall performance. Two deliberate choices concerning the properties of the phase function and the numerical quadrature give the eigenvalue problem (3) its structure, which is then exploited. As can be seen, the $2N \times 2N$ block matrix is composed by the $N \times N$ matrices α and β, and its structure comes from the choice that the phase function depends on the scattering angle only (which makes it possible to get uncoupled equations through the Fourier analysis), and from the choice of numerical quadrature where the nodes come in pairs and the corresponding weights are equal. This structure ensures that the eigenvalues occur in positive/negative pairs, and it allows the size of the eigenvalue problem to be reduced by a factor of 2 by rearranging it to the eigenvalue problem

$$(\alpha - \beta)(\alpha + \beta)(\mathbf{G}^+ + \mathbf{G}^-) = k^2(\mathbf{G}^+ + \mathbf{G}^-) \tag{5}$$

of size $N \times N$ for the eigenvectors $(\mathbf{G}^+ + \mathbf{G}^-)$ and the eigenvalues $\pm k^2$. Since the computation time for an eigenvalue problem grows approximately by the third power of the size, the eigenvalue computation time is thus reduced by a factor of 8. This was noted already by Chandrasekhar [3], and Stamnes and Swanson [14] proposed the formulation above.

4.2. Maintaining accuracy at low computational cost

In the cases of strongly forward peaked scattering, the phase function must be expanded in several hundreds or thousands of terms, and the numerical quadrature therefore needs a comparable number of terms. This quickly gives very large eigenvalue problems and systems of equations, which rapidly increases both computation time and memory requirements, and the problem soon becomes intractable. A transformation procedure, the δ–N method of Wiscombe [21], allows handling of such cases with maintained accuracy for significantly lower N than otherwise needed (in the setup just before *Algorithm 1*).
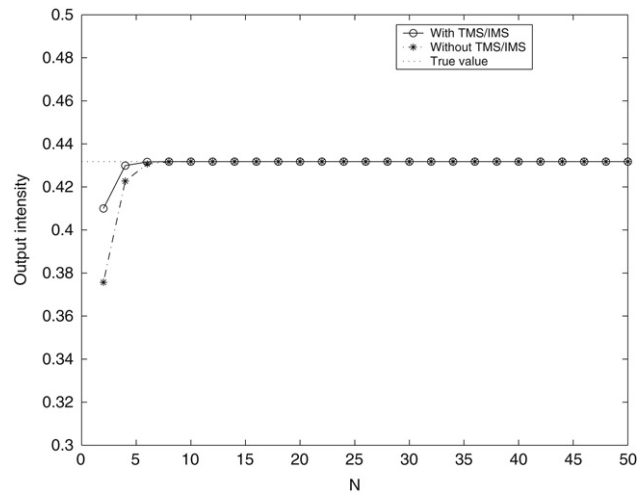
Fig. 5. Convergence of the method for an asymmetry factor of 0.5. Convergence to the true value is already achieved at a low number of terms, N, in the quadrature, but somewhat later without the intensity correction procedures (TMS/IMS).

The phase function is separated into the sum of a Dirac delta function in the forward direction and a truncated phase function, which is expanded in a much smaller number of terms. The structure of the equations remains unchanged under this operation. Intensity correction procedures (the TMS and IMS methods of Nakajima and Tanaka [10]) handle cases beyond the capabilities of the $\delta-N$ method. Through exact computation of low orders of scattering (for which closed expressions can be achieved after a substantial amount of algebra), these procedures help us to achieve high accuracy with small N, and thus help to speed up the calculations. These procedures involve a large amount of algebraic details, and the interested reader is referred to the original papers [21,10], or to the review of Edström [4].

4.2.1. Numerical experiments

The convergence behavior of the overall solution method was studied as in the conditioning studies in Section 3.2, but the test case presented here used 5 layers of constant thickness 0.01. The resulting intensity was studied with DORT2002, with and without the intensity correction procedures, in the middle of the medium in the incident direction, which for forward peaked scattering converges most slowly and is most sensitive to the asymmetry factor. All other directions converge much more rapidly. The 'true' intensity was calculated with $2N = 200$ using the intensity correction procedures.

4.2.2. Results

Figs. 5 and 6 show the convergence of the method with increasing N, for different asymmetry factors. As can be seen, the method converges monotonically to the true value when N increases. It is also evident from the plots that a larger N is needed to maintain accuracy for larger asymmetry factors, and that a larger N is needed for the same given accuracy if the intensity correction procedures are not used. It should be noted that if computation time is to be compared both with and without the intensity correction procedures, this should be done using the same accuracy, not the same N. A typical improvement with these methods is that for a medium with strongly forward peaked scattering (asymmetry factor larger than, say, 0.9) the N required for a reasonable accuracy decreases with a factor of 10 [10]. Since the overall computation time grows by $\sim N^3$, this decreases the computation time by a factor of ~ 1000.

4.3. Utilizing early convergence

Through Fourier analysis on the azimuthal angle, the integro-differential equation (1) is turned into a number of uncoupled equations, one for each Fourier component of the unknown intensity. These $2N$ sub-problems are solved independently one by one, since they are entirely uncoupled, starting with the 0th Fourier component and going up to number $2N - 1$. The complete azimuthal dependence is then assembled afterwards. Since this is done in an outer loop, much is gained if the loop can be terminated before the prescribed $2N$ times (the convergence criterion in *Algorithm 1*). In many cases the azimuthal dependence of the intensity converges well before this, as studied in [13,7]. Methods that break the azimuthal loop when a convergence criterion has been met can thus save large amounts of computation time. The formulation of the convergence criterion can vary; different approaches have been used in, e.g., DORT2002 [4] and DISORT [15] with similar results.
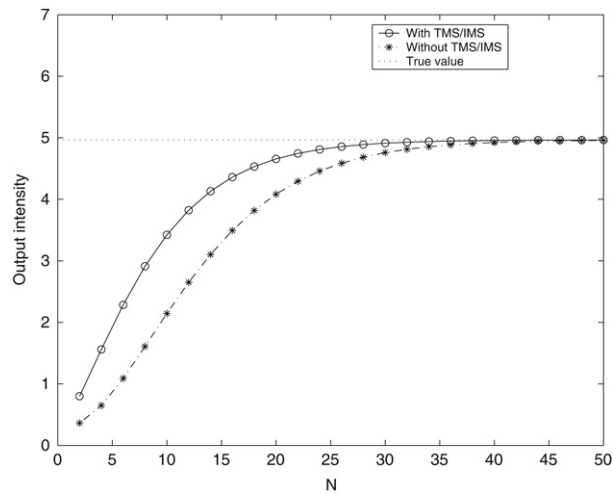
Fig. 6. Convergence of the method for an asymmetry factor of 0.9, which means rather forward peaked scattering. Convergence to a reasonable accuracy requires around twice as many terms, N, in the quadrature without, rather than with, the intensity correction procedures (TMS/IMS). This effect increases as the asymmetry factor approaches 1.
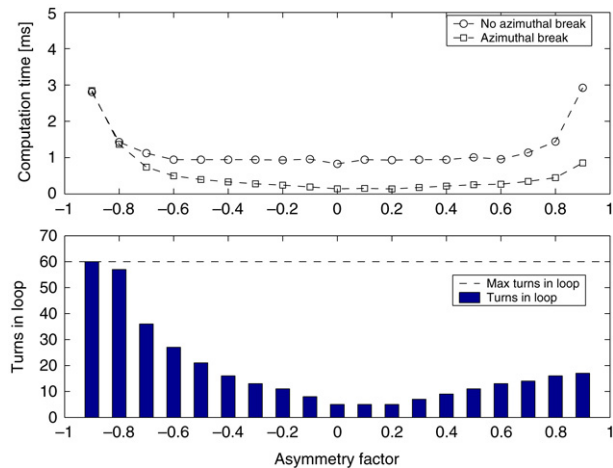

Fig. 7. Performance of the loop-breaking algorithm with no diffuse incident light (only beam), measured at the top of the medium. For an asymmetry factor of -0.9, the loop was correctly not broken, since a larger N is needed to achieve the specified accuracy. For all other asymmetry factors the loop could be broken well before its natural ending point, and the loop-breaking algorithm decreased the computation time by a factor of 10 on average.

4.3.1. Numerical experiments

The performance of the loop-breaking algorithm in DORT2002 was studied as in the conditioning studies in Section 3.2, but the test case presented here used a single layer thickness of 0.005, and $2N = 60$. The specified accuracy was a relative error less than 0.01. To study different interesting or extreme cases, one or more parameters were changed for four groups of tests, as indicated in the next section.

4.3.2. Results

The results are shown in plots that are divided into two parts. The first part shows the computation time with and without the loop-breaking algorithm, and thus shows the saving in computation time. The second part shows the number of turns used in the azimuthal loop, where the goal was to make it as far below $2N = 60$ (dashed line) as possible. The first group of tests used no diffuse incident light, only a beam. See Figs. 7 and 8, where the results are commented.

The second group of tests used only diffuse incident light and no beam, and the absence of a beam source gave extremely good performance at all depths and for all asymmetry factors, approximately as in Fig. 8 or better.

The third group of tests used the same parameter set as the first group, but with no underlying surface. This produced similar behavior to the first group at the top since the circumstances were hardly changed there, see Fig. 7. As could be expected, the absence of an underlying surface made the behavior at the bottom of the medium even simpler.
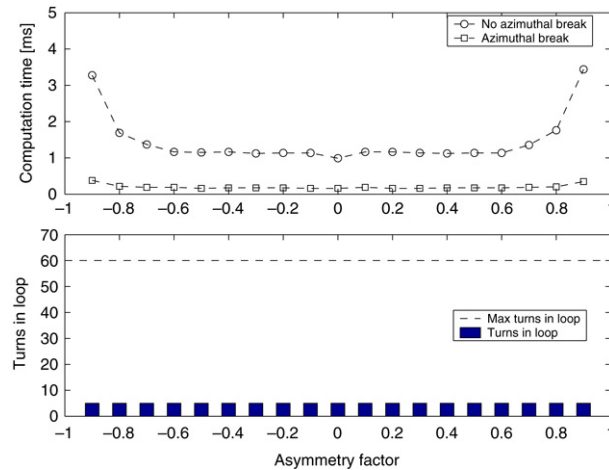
Fig. 8. Performance of the loop-breaking algorithm with no diffuse incident light (only beam), measured at the bottom of the medium. The performance was extremely good, which is always the case for a diffuse underlying surface, and the loop was already broken after the minimum number of turns that are needed to investigate the initial behavior of the Fourier components. The loop-breaking algorithm decreased the computation time by a factor of at least 10 in all cases.

The fourth group of tests used the same parameter set as the third group, but with a scattering coefficient of 0.01. This produced similar behavior to the first and third groups of tests at the top of the medium, while the performance was far better at the bottom.

The overall results show that the algorithm always works better for positive rather than negative asymmetry factors, that asymmetry factors closer to 0 give better performance, and that the loop-breaking algorithm decreased the computation time by a factor of 10 on average (this factor obviously depends on the properties of the medium, and the gain grows with larger N).

4.4. Computational shortcut for azimuthally averaged results

The azimuthally averaged intensity is given by the 0th Fourier component. Among the variables that depend only on the azimuthally averaged intensity are total reflectance, total transmittance, total absorptance and flux. Most solution methods break the azimuthal loop after the first time (at the convergence criterion in *Algorithm 1*) instead of fulfilling the prescribed $2N$ times when such variables are all that is required. This gives a significant reduction in computation time. The performance of the method was studied with DORT2002 for a wide range of parameter sets (using the test problems [22] mentioned in Section 3), and the shortcut decreased the computation time by a factor of 20–100 in these cases. Obviously, the gain increases with larger N.

5. Verification of stability and accuracy after speed increasing steps

5.1. Conditioning of the eigenvalue problem

Certain choices concerning the properties of the phase function and the numerical quadrature give the eigenvalue problem (3) its structure, which can be exploited to increase the speed of the calculations. This is treated in Section 4.1. However, since the eigenvalue problem is an important part of the core of the method (in the *layer* loop in *Algorithm 1*), it is important that it is well conditioned. Speed increasing steps must not be introduced at the cost of reduced stability.

5.1.1. Numerical experiments

The condition number of an individual eigenvalue is defined as the reciprocal of the cosine of the angle between the corresponding left and right eigenvectors (see e.g. [5, Section 7.2.2]). The largest of the eigenvalue condition numbers of the eigenvalue problem was studied with DORT2002 as in the conditioning studies in Section 3.2.

5.1.2. Results

Fig. 9 shows that the eigenvalue problem is very well conditioned, giving a condition number close to 1 in all cases. Thus, the speed increasing steps that affect the structure of the eigenvalue problem do not give it a poor stability. Therefore, no stability increasing steps are needed for the eigenvalue problem. As can be seen from the plot, the 0th Fourier component is the worst case, although still very good, with the condition number rapidly decreasing as the Fourier component number increases. The eigenvalue problem is independent of the number of layers, since it is solved for each layer separately.
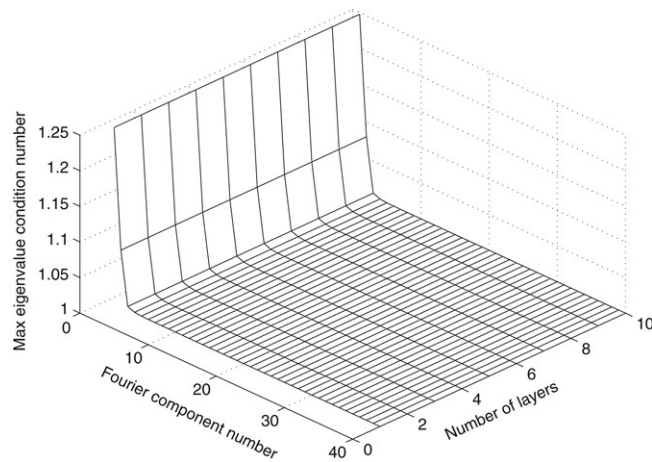
Fig. 9. The largest of the eigenvalue condition numbers of the eigenvalue problem. The condition number is close to 1 in all cases. The condition number is independent of the number of layers.

5.2. Overall accuracy

It is essential that the steps included to increase the speed of a method do not compromise the accuracy. To verify this, a series of tests were performed. Results with the speed increasing steps were compared to reference values obtained without the speed increasing steps and with large N for accuracy.

5.2.1. Numerical experiments

The accuracy of the speed increasing steps were studied for a wide range of parameter sets (using the test problems [22] mentioned in Section 3). The test problems come with both input and output values, and most problems can also be compared with published results that are referred to in the test suite. The different test cases include accuracy and consistency checks, variation of all parameters, and cases with a risk of breakdown due to extreme values of the input parameters.

5.2.2. Results

The comparisons gave very good agreement in all test cases without exception. The deviations from reference values of results obtained with the speed increasing steps were never larger than the round-off error in the given data. This provides substantial support for the accuracy of the tested speed increasing steps, and thus indicates that they are not introduced at the cost of reduced accuracy.

6. Suggestions for future work

The intensity correction procedures (the TMS and IMS methods) take almost no extra time in themselves. However, for very forward scattering media (asymmetry factor close to 1), they utilize values of the normalized associated Legendre functions, Λ_l^m, for indices l and m from 1 to fairly large numbers, and the computation time of all these evaluations becomes noticeable. Any studies that result in faster ways of evaluation of Λ_l^m for indices l and m from 1 up to a few hundreds would be welcome.

Another improvement that will have a significant effect on the overall performance is to speed up the computation of the eigenvalue problem. As discussed in [16], the reduced matrix is real and non-symmetric, but in spite of this is known to have real eigenvalues. They showed that it is possible to make the matrix symmetrical, but their method introduced extra computational cost as well as round-off errors due to the matrix multiplications involved in the transformations. Since the numerical solution of the eigenvalue problem is a critical and time-consuming part of the discrete ordinate method, a faster method that avoids the use of complex arithmetic – without loss of accuracy or efficiency – would be a welcome improvement.

7. Discussion

This paper gives a systematic presentation of *the effect* of the steps that are needed or possible to make *any* discrete ordinate radiative transfer solution method numerically efficient. To the author's knowledge, this has not been summarized in one single publication before. This is done through studies of the numerical performance of the stability enhancing and speed increasing steps used in modern discrete ordinate radiative transfer algorithms. The solution method DORT2002 in [4]

is used in the tests, since it is designed for methodical numerical experiments through its modularized design and ability to give any kind of intermediate results and performance data.

The system of equations corresponding to boundary and continuity conditions is very ill-conditioned, with a condition number near the largest positive floating-point number for the system. It is shown that after applying the preconditioner, the condition number is close to 1 in most cases, and around 30 in the worst test case. It is also shown that the preconditioner preserves the sparsity pattern, which is used to solve the system of equations efficiently. This indicates that the preconditioner works very well, and yields a system of equations well suited for numerical solution.

The structure of the eigenvalue problem allows reducing its size by a factor of 2, and thus the eigenvalue computation time by a factor of 8. It is shown that the reduced eigenvalue problem is very well conditioned, giving a condition number close to 1, so this reduction is not introduced at the cost of reduced stability.

The convergence behavior of the method is illustrated. The intensity correction procedures make it possible to maintain accuracy for significantly lower N than otherwise needed. It is shown that this typically decreases the computation time by a factor of $\sim$1000 in cases with strongly forward peaked scattering.

It is shown that algorithms for breaking the azimuthal loop, based on a convergence criterion, on the average decrease the computation time by a factor of 10. The gain grows with larger N, and also depends on the properties of the medium.

A computational shortcut, implemented to allow for much faster calculation of variables that depend only on the azimuthally averaged intensity, makes it possible to break the azimuthal loop after the first time instead of fulfilling the prescribed $2N$ times. It is shown that the shortcut decreases the computation time by a factor of 20–100 when $2N = 40$. Obviously, the gain increases with larger N.

Different sets of relevant test problems are solved with and without the speed increasing steps. It is shown, by comparing the results, that the agreement is invariably very good. This provides substantial support for the accuracy of the tested speed increasing steps, and thus shows that they are not introduced at the cost of reduced accuracy.

The performance tests illustrate the effect of the steps that are taken to improve the stability and speed of the method. It is shown how the steps together give a stable solution procedure to a problem previously considered numerically intractable, and how they together decrease the computation time compared to a naive implementation by a factor 1000 in typical cases and far beyond in extreme cases. Further studies and developments, which can have a positive impact on computation time, are suggested.

Acknowledgements

References

[1] S. Chandrasekhar, On the radiative equilibrium of a stellar atmosphere, Astrophys. J. 99 (1944) 180–190.
[2] S. Chandrasekhar, On the radiative equilibrium of a stellar atmosphere II, Astrophys. J. 100 (1944) 76–86.
[3] S. Chandrasekhar, Radiative Transfer, Dover, New York, 1960.
[4] P. Edström, A fast and stable solution method for the radiative transfer problem, SIAM Rev. 47 (2005) 447–468.
[5] G.H. Golub, C.F. Van Loan, Matrix Computations, 3rd ed., Johns Hopkins University Press, Baltimore, 1996.
[6] L.G. Henyey, J.L. Greenstein, Diffuse radiation in the galaxy, Astrophys. J. 93 (1941) 70–83.
[7] M.D. King, Number of terms required in the fourier expansion of the reflection function for optically thick atmospheres, J. Quant. Spectrosc. Radiat. Transfer 30 (1983) 143–161.
[8] P.S. Mudgett, L.W. Richards, Multiple scattering calculations for technology, Appl. Opt. 10 (1971) 1485–1502.
[9] P.S. Mudgett, L.W. Richards, Multiple scattering calculations for technology II, J. Colloid Interf. Sci. 39 (1972) 551–567.
[10] T. Nakajima, M. Tanaka, Algorithms for radiative intensity calculations in moderately thick atmospheres using a truncation approach, J. Quant. Spectrosc. Radiat. Transfer 40 (1988) 51–69.
[11] A. Schuster, Radiation trough a foggy atmosphere, Astrophys. J. 21 (1905) 1–22 (Reprinted in D.H. Menzel (Ed.), Selected Papers on the Transfer of Radiation, Dover, New York, 1966, pp. 3–24).
[12] K. Stamnes, P. Conklin, A new multi-layer discrete ordinate approach to radiative transfer in vertically inhomogeneous atmospheres, J. Quant. Spectrosc. Radiat. Transfer 31 (1984) 273–282.
[13] K. Stamnes, H. Dale, A new look at the discrete ordinate method for radiative transfer calculation in anisotropically scattering atmospheres. II: Intensity computations, J. Atmos. Sci. 38 (1981) 2696–2706.
[14] K. Stamnes, R.A. Swanson, A new look at the discrete ordinate method for radiative transfer calculations in anisotropically scattering atmospheres, J. Atmos. Sci. 38 (1981) 387–399.
[15] K. Stamnes, S.-C. Tsay, I. Laszlo, DISORT, a general-purpose fortran program for discrete-ordinate-method radiative transfer in scattering and emitting layered media, NASA Report, 2000.
[16] K. Stamnes, S.-C. Tsay, T. Nakajima, Computation of eigenvalues and eigenvectors for discrete ordinate and matrix operator method radiative transfer, J. Quant. Spectrosc. Radiat. Transfer 31 (1984) 273–282.
[17] K. Stamnes, S.-C. Tsay, W. Wiscombe, K. Jayaweera, Numerically stable algorithm for discrete-ordinate-method radiative transfer in multiple scattering and emitting layered media, Appl. Opt. 27 (1988) 2502–2509.
[18] J.B. Sykes, Approximate integration of the equation of transfer, Mon. Not. Roy. Astron. Soc. 111 (1951) 377–386.
[19] G.E. Thomas, K. Stamnes, Radiative Transfer in the Atmosphere and Ocean, Cambridge University Press, 1999.
[20] G.C. Wick, Über ebene Diffusionsprobleme, Z. Phys. 120 (1943) 702–718.
[21] W.J. Wiscombe, The delta-M method: Rapid yet accurate radiative flux calculations for strongly asymmetric phase functions, J. Atmos. Sci. 34 (1977) 1408–1422.
[22] The DISOTEST files in the DISORT2.0beta folder at the DISORT web site: ftp://climate1.gsfc.nasa.gov/wiscombe/Multiple_Scatt/.