Stellar Atmosphere Modeling
ASP Conference Series, Vol. 288, 2003
I. Hubeny, D. Mihalas, & K. Werner, eds.

Model Photospheres with Accelerated Lambda Iteration

Ivan Hubeny

AURA/NOAO, Tucson, AZ 85726-6731, USA

Thierry Lanz

Dept of Astronomy, University of Maryland, College Park, MD 20742, & NASA Goddard Space Flight Center, Code 681, Greenbelt, MD 20771

Abstract. We review briefly efficient numerical methods for computing non-LTE model stellar atmospheres. We give emphasis to methods that use the Accelerated Lambda Iteration and related methods. Finally, the computer program TLUSTY is briefly described.

1. Introduction

There is hardly any need to stress that modeling stellar atmospheres has undergone dramatic improvements during the last decade. This was, in fact, the reason for organizing this Workshop. At the same time, recent advances in observational detectors have resulted in unprecedented quality (and quantity) of stellar spectra. Because of that, many astronomers, who would otherwise not be very interested in stellar atmospheres, are feeling a need to have better models than those previously used, in particular the broad grid of Kurucz (1994) models. Stellar atmosphere theory and modeling are thus enjoying a period of revival.

On the modeling side, the central role in the development of the efficient numerical techniques has been played by Accelerated Lambda Iteration (ALI) methods. The method was reviewed in these Proceedings by Hubeny. In this paper, we will discuss its specific application to modeling stellar atmospheres. Because the case of spherical expanding atmospheres will be covered by several other contributions in this volume (Hamann; Koesterke; Hillier, Hoeflich, Hauschildt), we will concentrate here on a simple problem, 1-D, plane-parallel, hydrostatic atmospheres, i.e. stellar photospheres. In particular, we will discuss the basic philosophy of various numerical schemes rather than describe details of any particular method.

2. Basic Structural Equations

Consider first the basic structural equations that describe a stellar atmosphere as a plane-parallel, horizontally-homogeneous (i.e., 1-D) slab, in hydrostatic and radiative equilibrium.

Radiative transfer equation. The most advantageous form of the transfer equation for the use in model atmosphere construction is the second-order form,

$$\frac{d^2(f_\nu J_\nu)}{d\tau_\nu^2} = J_\nu - S_\nu \,, \tag{1}$$

where τ_ν is the monochromatic optical depth and f_ν is the variable Eddington factor (Auer & Mihalas 1970). Equation (1) contains only the mean intensity of radiation, J_ν (a function of frequency and depth), not the specific intensity, $I_{\mu\nu}$, (which is also a function of angle μ). In fact, it is the mean intensity of radiation that enters other structural equations, and the rate equations, therefore mean intensities, not specific intensities, are the appropriate variables to take as atmospheric state parameters. An obvious numerical advantage is that instead of dealing with $NF \times NA$ quantities describing the radiation field per depth-point (NF and NA being the number of discretized frequency and angle points, respectively), we have only NF parameters. However, f_ν is not given a priori; it has to be computed by a separate set of formal solutions for the specific intensities, one frequency at a time, and gradually updated using current values of optical depth and the source function.

Hydrostatic equilibrium equation. This equation reads

$$\frac{dP}{dm} = g \,, \tag{2}$$

where P is the total pressure, and m the Lagrangian mass, $dm = -\rho\, dz$, with ρ being the density and z the geometrical distance measured along the normal to the surface from the bottom of the atmosphere to the top. g is the surface gravity, which is assumed constant throughout the atmosphere, and given by $g = GM_*/R_*^2$, where M_* and R_* are the stellar mass and radius, respectively; G is the gravitational constant. The total pressure is generally composed of three parts, the gas pressure, $P_{\rm gas}$, the radiation pressure, $P_{\rm rad}$, and the "turbulent pressure, $P_{\rm turb}$". The gas pressure is given, assuming an ideal gas equation of state, by $P_{\rm gas} = NkT$, where N is the total particle number density, T the electron temperature (we assume that all the particles have the same kinetic temperature), and k the Boltzmann constant. The hydrostatic equilibrium equation may then be written as (neglecting "turbulent pressure")

$$\frac{dP_{\rm gas}}{dm} = g - \frac{4\pi}{c}\int_0^\infty \frac{dK_\nu}{dm} = g - \frac{4\pi}{c}\int_0^\infty \frac{\chi_\nu}{\rho} H_\nu d\nu \,, \tag{3}$$

where H_ν and K_ν are the first and second angular moments of the specific intensity, respectively. The r.h.s. of this equation is usually called the *effective gravitational acceleration*, resulting from the action of a true gravitational acceleration (acting downward) minus a radiative acceleration (acting outward).

Radiative equilibrium equation. It can be written either as expressing conservation of the total radiation flux (the so-called differential form),

$$\int_0^\infty H_\nu d\nu = \int_0^\infty \frac{d(f_\nu J_\nu)}{dm}\, d\nu = \frac{\sigma}{4\pi}\, T_{\rm eff}^4 \,, \tag{4}$$

where $T_{\rm eff}$ is the effective temperature, and σ the Stefan-Boltzmann constant. The radiative equilibrium equation can also be written as an equality of the total absorbed and emitted energy (the integral form),

$$\int_0^\infty (\kappa_\nu J_\nu - \eta_\nu)\, d\nu = 0\,, \tag{5}$$

where κ_ν is the *thermal* absorption coefficient, and η_ν the thermal emission coefficient (the scattering contributions cancel out when one assumes coherent scattering, which is usually the case in stellar atmospheres).

Both equations are equivalent, but have different numerical properties. To improve a numerical accuracy of the solution, one either considers the integral form at the upper layers of the atmosphere and the differential one at deep layers (Gustafsson 1971; Hubeny 1988), or one considers a linear combination of both forms (Hubeny & Lanz 1995; Werner & Dreizler 1999).

An alternative expression of radiative equilibrium is the condition of thermal balance of electrons, stating that total energy added to the electron thermal pool by radiative and collisional processes is equal to the energy removed from the thermal pool. This condition is usually used in modeling photoionized media (Osterbrock 1989); it was used for constructing model stellar atmospheres by Kubát et al (1999; see also Kubát, this volume). It was also used by Hillier & Miller (1998) as a check of their radiative + statistical equilibrium solution.

Statistical equilibrium equations. They are also called rate equations, and are usually written as

$$n_i \sum_{j\neq i} (R_{ij} + C_{ij}) = \sum_{j\neq i} n_j\, (R_{ji} + C_{ji})\ , \tag{6}$$

where n_i is the population (occupation number) of level i, R_{ij} and C_{ij} is the radiative and collisional rate, respectively, for the transition from level i to level j, including continuum states. The radiative rates depend on the radiation intensity, while the collisional rates are assumed to be given functions of temperature and electron density.

The set of rate equations for all levels of an atom would form a linearly dependent system. Therefore, one equation of the set has to be replaced by another equation. Usually, this is the *total number conservation* equation (or abundance definition equation), $\sum_i n_i = N_{\rm atom}$, where the summation extends over all levels of all ions of a given species.

Charge conservation equation. This equation expresses the global electric neutrality of the medium,

$$\sum_i n_i Z_i - n_{\rm e} = 0\,, \tag{7}$$

where Z_i is the charge associated with level i (i.e. equal to 0 for levels of neutral atoms, 1 for levels for once ionized ions, etc.), and $n_{\rm e}$ is the electron density. The summation extends over all levels of all ions of all species.

Auxiliary equations. The above set of structural equations have to be complemented by equations defining the absorption and emission coefficients, expressions for atomic cross-sections, and other auxiliary expressions.

3. Solution by Complete Linearization

The basic structural equations (1) - (7) are discretized in frequency and depth, which yields a set of highly–coupled, non–linear algebraic equations. The fundamental problem of stellar atmosphere modeling is to find a robust and efficient method for a numerical solution of these equations. The decisive breakthrough, and in fact the beginning of the modern era of stellar atmosphere models, was the development of the *complete linearization* (CL) method by Auer & Mihalas (1969). This was the first scheme to treat all equations at the same footing, thus solving all structural equations simultaneously. Before that, the equations were typically solved one at a time, iterating between them. In many cases, iterations were slow, or a scheme failed to converge at all.

We shall first briefly outline the complete linearization method. The physical state of an atmosphere is fully described by the set of vectors ψ_d for every depth point d, $d = 1, \ldots, ND$, ND being the total number of depth points. The state vector ψ_d is given by

$$\psi_d = \{J_1, \ldots, J_{NF}, N, T, n_{\rm e}, n_1, \ldots, n_{NL}\}\,, \tag{8}$$

where J_i is the mean intensity of radiation in the i-th frequency point; we have omitted the depth subscript d. The dimension of the vector ψ_d is NN, $NN = NF + NL + NC$, where NF is the number of frequency points, NL the number of atomic energy levels for which the rate equations are solved, and NC is the number of constraint equations ($NC = 3$ in the present case; we leave NC as a general number since in some cases it may be larger than 3, for instance in the case of radiative + convective equilibrium it is advantageous to add an additional state parameter Δ, the logarithmic gradient of temperature, etc.).

The set of structural equations may be formally written formally as

$$\mathrm{P}(\mathbf{x}) = 0, \tag{9}$$

where $\mathbf{x}$ is a vector formed from all vectors ψ_d, $\mathbf{x} = \{\psi_1, \ldots, \psi_{ND}\}$. The original complete linearization is nothing else than the Newton–Raphson method of solving Eq.(9), namely

$$\mathbf{x}^{(n+1)} = \mathbf{x}^{(n)} - J\left(\mathbf{x}^{(n)}\right)^{-1} \mathrm{P}\left(\mathbf{x}^{(n)}\right), \tag{10}$$

where J is the Jacobi matrix (Jacobian), $J_{ij} = \partial P_i / \partial x_j$, i.e. the ij-element of the Jacobian is the derivative of the i-th equation with respect to the j-th unknown. Since the system (9) represents a *finite difference* solution of at most second–order differential equations the Jacobian J has a particularly simple structure, namely a block–tridiagonal form, and Eq. (9) reduces to,

$$-\mathsf{A}_d \delta\psi_{d-1} + \mathsf{B}_d \delta\psi_d - \mathsf{C}_d \delta\psi_{d+1} = \mathbf{L}_d\,. \tag{11}$$

Here A, B, C are $NN \times NN$ matrices, and $\mathbf{L}_d = \mathrm{P}_d\left(\mathbf{x}^{(n)}\right)$ is the residuum vector (of dimension NN) at depth d. Equation (11) is solved as a block-tridiagonal system, which, by the way, can be done in place (i.e. without additional storage beyond A, B, C.) We are left with inverting one $NN \times NN$ matrix per depth

point. Therefore, the total computer time for ordinary complete linearization scales roughly as

$$(NF + NL + NC)^3 \times ND \times N_{iter} \,. \tag{12}$$

It is immediately clear that the original complete linearization, despite its inherent power and robustness, cannot be used as a general numerical scheme because in realistic calculations one must use a very large number of frequency points *NF* to describe the radiation field sufficiently accurately – of the order of 10^5 to 10^6 points. Inverting matrices of this dimension is completely out of question. One obviously must seek less global, but much faster schemes.

4. The Quest for the Most Efficient Method of Simultaneous Solution of The Structural Equations

4.1. Overview

Any method should solve all the structural equations, although not necessarily simultaneously as the complete linearization does. The art is to find a procedure that allows one not to solve all the equations simultaneously (that is, some information is being lagged behind), while remaining sufficiently efficient and robust. Taking equation (12) as a guide, we list below a number of options that can be used to reduce the total computer time. We first outline all the options, then discuss some of the most important possibilities in detail.

1. Modifications within the framework of standard linearization:
   - Reducing *NF*
     - coarse frequency resolution; neglecting lines
     - Λ-iteration type of treatment of some frequency points
     - frequency blocks
     - applying ALI
   - Reducing *NL*
     - LTE
     - simplified atomic models
     - superlevels
     - level grouping
     - level zeroing
   - Reducing *NC*
     - solving hydrostatic equilibrium equation separately
     - solving radiative equilibrium equation separately
   - Reducing *ND*
     - coarse depth resolution
     - multigrid schemes
     - adaptive grids
   - Reducing N_{iter}

- improved formal solution
- acceleration of convergence (e.g., Ng)
- successive over-relaxation

2. Using linearization, but avoiding inverting the Jacobian
   - Broyden method
   - Kantorovich method
3. Approximate Newton-Raphson method
4. Not using linearization at all

4.2. Improvements of Standard Linearization

Here we discuss in more detail the first category of options, namely those in which the global framework of complete linearization remain intact, while the efficiency is gained by reducing the size of the state vector. i.e., by reducing the size of matrices to be inverted.

Reducing NF There are several ways to achieve this:

i) The obvious possibility is to consider only a low number of frequencies. This was actually approach that had to be adopted in applications of the original complete linearization method. However, as we aim at constructing a method that is able to provide an "exact" solution for the stellar atmosphere problem, we do not consider this possibility

ii) The next simplest approach is to linearize mean intensities only those frequencies that are "essential", while keeping the others fixed during linearization and updating them during a subsequent formal solution. Such an idea was implemented, for instance, in the original CL program by Mihalas et al. (1975), who actually used a more sophisticated equivalent-two-level-atom (ETLA) procedures, and in early versions of the program TLUSTY (Hubeny 1988).

By "formal solution" we mean a solution of *one equation at at time*, using current values of all other state parameters (for instance, solving the radiative transfer equation for given frequency for current values of the level populations, temperature, and electron density). The fixed–rates approach has proved to be useful in some cases, but if too many transitions are taken in this mode one basically recovers a Lambda-iteration type of behavior, which is a serious drawback. In particular, the convergence rate is slow, and the solution tends to stabilize rather than truly converge. Using ETLA procedure partly overcomes this problem, but it is much more time-consuming.

iii) Another way is to adopt the multi–frequency/multi–grey method (Anderson 1985; 1989). Basically, one substitutes $(J_1, \ldots, J_{NF}) \rightarrow (\tilde{J}_1, \ldots, \tilde{J}_{NB})$, where $\tilde{J}_i$ represents a mean intensity characteristic of i-th frequency block, and NB is the number of blocks. In order to achieve a substantial reduction of time we have $NB \ll NF$. Each block groups together all frequencies for which the radiation is formed in a similar way. It is not necessary that the block be composed of a continuous frequency interval (for instance, one block may represent all wings of weak lines, etc. – see Anderson 1985). The essence of the method consists of selecting appropriate frequency bands and the individual frequency

points belonging to them. However, this is also a drawback of the method, since the bands have to be set essentially by hand.

iv) Finally, the Accelerated Lambda Iteration (ALI) method (see, e.g., Hubeny, in this Proceedings and references there) reduces the number of unknowns even more, because it is able to eliminate *all* frequency points completely. This is achieved by expressing the mean intensity of radiation as

$$J_\nu^{(n)} = \Lambda_\nu^* S_\nu^{(n)} + (\Lambda_\nu - \Lambda_\nu^*)\, S_\nu^{(n-1)}, \tag{13}$$

where Λ_ν and Λ_ν^* are the exact and the approximate Lambda operator, and S_ν the source function, all at frequency ν. The superscript n indicates the iteration number. The mean intensity of radiation is thus represented by two terms. The second one – the "correction" term – is known from the previous iteration and is thus not linearized, while the first one represents an action of an approximate (and, therefore, simple) operator, Λ^*, on the source function, which is expressed as a function of temperature, density, and atomic level populations. The radiative transfer equations are thus eliminated from the coupled system of structural equations. The method was developed by Hubeny & Lanz (1995), who coined the term "hydrid complete linearization/accelerated Lambda iteration" (CL/ALI) method. We shall return to this method in § 5.3..

Reducing NL While the above described methods, in particular the hybrid CL/ALI method, are able to reduce the value of the dominant component of the state vector (*NF*); the second largest component, the total number of atomic energy levels for which the statistical equilibrium equations are solved (*NL*), may still be prohibitively large. For instance, the iron-peak elements have typically some 10,000 levels per ion, so the total number of levels may actually become comparable to the total number of frequency points. Again, there are several ways to cope with this problem:

i) A straightforward way is to assume Local Thermodynamic Equilibrium (LTE), in which case the atomic level populations are given by the Saha-Boltzmann distributions, and are thus given functions of T and $n_{\rm e}$. However, we intend to construct more realistic, non-LTE models, so we reject this option.

ii) Analogously to the case of *NF*, an obvious possibility is to use simplified model atoms, in which only several most important (usually low-lying) levels are considered explicitly. The remaining levels are either neglected, or treated in LTE. A variant of the latter approach is to express the total population of non-explicit, higher levels through the partition function (Hubeny 1988). Generally, such a strategy may work for light elements, but cannot be used for iron-peak elements because of an large number of levels in the whole energy range.

iii) A much better approach is to consider so-called *superlevels* (Anderson 1989; Dreizler & Werner 1993; Hubeny & Lanz 1995). The idea consists of grouping several (many) individual energy levels together, to form a "superlevel". The basic physical assumption is that all real levels j forming the superlevel J have a common NLTE departure coefficient, or, in other words, all components j are in Boltzmann equilibrium with respect to each other. There is a certain flexibility in choosing a particular partitioning of individual levels into superlevels. Various authors use different approaches to set up superlevels; we have used (Hubeny & Lanz 1995) an explicit criterion that all levels within a superlevel must have

the same parity. The actual choice is made by inspecting the distribution of the individual excitation energies for levels in the even and the odd parity system, looking for clustering of energies. One selects typically 10 to 60 superlevels per ion. The idea of superlevels may be used for any species, not only for iron-peak elements. In fact, this approach has always been used in astrophysical radiative transfer without being called such; for instance one typically works with hydrogenic levels that are specified only by the main quantum number n and the populations of the individual l states are summed up; or in most cases the individual J components of the multiplet systems are lumped to one LS level, etc.

iv) Another idea, which may significantly reduce the number of level populations to be linearized is the idea of *level grouping*. The level group is a set of one or more levels whose populations are assumed to vary in a coordinated way in the linearization. More precisely, instead of linearizing the individual level populations, one linearizes the total population of the groups, assuming that the ratios of the individual level populations within the group to the total population of the group is unchanged in linearization. In the formal solution step, one solves exactly for all the individual level populations. The concept of level groups should not be confused with the concept of superlevels; in the former case, the level groups are only a numerical trick to make the matrices of complete linearization smaller, while the level populations are determined exactly; the latter case – superlevels – approximates the individual populations of the components of the superlevel by assuming that they are in Boltzmann equilibrium with respect to each other. In fact, one may group the individual superlevels into level groups as well.

v) Finally, we mention a useful numerical trick that belongs to the category of reducing *NL* – *level zeroing*. We have implemented it in two steps. The first one is very simple: Whenever a local population of a level becomes lower than a prescribed fraction of the total population of the species (typically 10^{-20} to 10^{-30}), the population is set exactly as zero, and instead of considering an appropriate rate equation for such a level, one replaces it by a simple condition $n_i = 0$. This does not really decrease the number of state parameters, but improves numerical stability without compromising the final solution.

The second step is called a *global zeroing*, which simply rejects a level from the set of state parameters if the population of this level satisfies a zeroing condition for all depths in the atmosphere. This is a useful approach in all cases where the selection of explicit energy levels is data-oriented, and the same input data file is used for atmospheres of quite different basic parameters (e.g., $T_{\rm eff}$). The advantage is that one may use a universal data sets for the whole array of ions of a given species (for instance, Fe I - Fe XXVII), and let the program itself determine which levels of which ions make have non-negligible population for given conditions (see, e.g., AGN disk models computed by a variant of TLUSTY by Hubeny et al. 2001).

Reducing NC One possibility is to solve the hydrostatic equilibrium equation separately (Gustafsson, 1971; Werner 1986; Werner & Dreizler, this volume). Yet another possibility is to solve the radiative equilibrium equation separately by a *temperature-correction procedure*. These methods are discussed in detail in other papers in this volume (Dreizler; Hauschildt).

In the context of modified complete linearization, these possibilities usually do not offer any substantial advantage. In some cases they may lead to a somewhat more stable solution, but typically the iteration process converges slower or fails to converge at all. The reasons are analogous to those for the failure of the classical Lambda iteration, and have been explained many times in the literature (e.g. Mihalas 1978). However, these methods may be very helpful if one tries to avoid linearization completely. We shall return to this point in § 5.2.; see also contributions by Dreizler and by Hauschildt in this Proceedings.

Reducing ND Again, a trivial possibility is using a coarse depth mesh. This has an additional benefit that the ALI method converges faster when using a coarser depth resolution. However, many problems (e.g. model in which one encounters multiple ionization fronts; convective models; etc.) may require a large number of depths. But as the computation time is linear in number of depths, increasing *ND* is usually not a big problem. Nevertheless, convergence of ALI-based methods may deteriorate. In such cases one could use *multi-grid schemes* (e.g., Trujillo Bueno & Fabiani Bendicho 1995). To our knowledge, such an approach has not yet been used in the context of the full stellar atmosphere problem.

Another potentially promising possibility is to use *adaptive grid* techniques (e.g., Dorfi & Druri 1987), which, so far, have been used with great success in radiation-hydrodynamics calculations. Again, these have not yet been used in the context of stellar atmosphere models. In our opinion, it perhaps may be one of the most important remaining improvements of classical 1-D stellar atmosphere modeling techniques.

Reducing N_{iter} Such a reduction can certainly be achieved by an appropriate modification of the global iteration scheme; for instance the hybrid CL/ALI method decreases the number of iterations significantly with respect to the ALI method by itself. What we mean here, however, is a reduction of number of iterations for a given global scheme. There are different ways to achieve that:

i) The first possibility is to improve the "formal solution", i.e., all calculations that are done between two subsequent iterations of the linearization scheme. Typically, linearization provides new values of the components of the state vector; one may then keep some parameters fixed (e.g., temperature and density), and compute a more consistent values of others (that is, mean intensities and level populations, and possibly electron density), by a simultaneous solution of the transfer and the rate equations. Typically, most of the modeling programs offer some kind of "Lambda iteration" treatment to improve level populations In TLUSTY we use another ALI approach with preconditioning (see § 5.1.) to update the atomic level populations. The basic idea behind all such approaches is to determine values of all state parameters as consistently as possible before entering a new iteration of complete linearization, with the hope that this helps the overall iteration process. In many cases, it does indeed help significantly.

ii) One can use mathematical *acceleration of convergence* procedures. The most popular one is the Ng acceleration (Ng 1974; Auer 1987, 1991), which was first used in the context of accelerating a complete-linearization based scheme to calculate model stellar atmospheres by Hubeny & Lanz (1992), where the

reader is referred to for more details. Our experience showed that in a vast majority of cases the Ng acceleration improves convergence significantly; the acceleration is usually performed for the first time at or around 7-th iteration of CL, and is done typically every 4 iterations afterwards. In some cases, like in models with convection, or in AGN disk models with sharp ionization fronts, the Ng acceleration does not help, and may even lead to numerical problems and divergence.

iii) Another potential possibility is to use the *successive over-relaxation* (SOR) method. It consists in multiplying corrections $\delta\psi$ by a certain coefficient α. This coefficient can be either set up by an educated guess, or one can use a procedure suggested by Trujillo Bueno & Fabiani Bendicho (1995), namely to express α through the spectral radius of the appropriate iteration operator, which in turn is given as a ratio of maximum relative changes of the source function in two subsequent previous iterations.

4.3. Linearization without Inverting the Jacobian

The next class of improvements, which can be used in conjunction with any of the above methods, is to avoid somehow inverting the Jacobian of the system, which clearly leads to substantial computer-time savings. In these methods, one has to perform a few first iterations using the original Newton-Raphson scheme where the Jacobian is inverted; only when the current estimate of solution is "close" to the true solution, one may use a simplified treatment. There are essentially two possibilities:

Broyden inversion The first is an application of the Broyden method, also called "least change secant method", introduced by Koesterke, Hamann, & Kosmol (1992). Instead of using a Jacobi matrix (which is an analog of tangent in one dimension), one uses a Broyden matrix, whose analog in 1-D is a secant. The inverse Broyden matrix in the subsequent iteration can be calculated directly from the previous inverse Broyden matrix. This means that inverting matrices $NN \times NN$ is avoided. The method and its actual application to the stellar photosphere models is described by Werner & Dreizler in this volume.

Kantorovich method An even simpler method is called Kantorovich method, which keeps the Jacobian fixed after a certain iteration, so the subsequent iterations of the CL used the same Jacobian (better speaking, the inverse of Jacobian is kept fixed for future use); only the residuum vectors $\mathbf{L}$ are re-evaluated after each iteration. In the one-dimensional analog, the Newton method computes a new iterate using the current slope of the tangent, while the Kantorovich variant keeps the tangent fixed. The method was used by Hillier (1990); the properties of the method as applied to the CL scheme were studied in detail by Hubeny & Lanz (1992), who also coined the term Kantorovich method, since Kantorovich (1949) first proved exactly the convergence of the method. Our experience with the method showed that the method is surprisingly robust. Usually, one needs to perform 2-4 iterations of the full linearization scheme, depending on the problem at hand and the quality of the initial estimate. Also, it is sometimes very advantageous to "refresh" the Jacobian (i.e., set it up using current solution and invert it) after certain number of Kantorovich iterations.

4.4. Approximate Newton-Raphson Method

The method was first suggested by Hempe & Schönberg (1986) in the context of line formation with velocity fields, and subsequently elaborated and extended by Hillier (1990) and Hillier & Miller (1998) to treat full spherically expanding model atmosphere problem.

The method is conceptually very close to the ALI-type methods; it differs in implementation and in some subtle points. The idea is as follows. We start with the traditional CL method. Instead of linearizing all the components of the state vector, we use the transfer equations to eliminate the corrections δJ_i from the solution vector. From the linearized transfer equation we have (for $i = 1, \ldots, NF$ and $d = 1, \ldots, ND$)

$$\delta J_{id} = \sum_{d'=1}^{ND} \sum_{j=1}^{NL+NC} \frac{\partial J_{id}}{\partial x_{jd'}} \, \delta x_{jd'} \,, \tag{14}$$

where x_{jd} are the components of the state vector other than radiation intensities (i.e., level populations, T, N, and $n_{\rm e}$), and the $\partial J_{id}/\partial x_{jd'}$ terms can be derived from linearized transfer equations. The important point is that all components of this matrix are, in general, nonzero. This is because a discretized transfer equation (1) can be written in matrix form as

$$\mathrm{T}_i \, \mathbf{J}_i = \mathbf{S}_i \,, \tag{15}$$

where $\mathbf{J}_i = (J_1, \ldots, J_{ND})^T$ is the vector of mean intensities at frequency i, $\mathbf{S}_i$ the source vector, and T_i is a tridiagonal matrix. Consequently, in $\mathbf{J}_i = \mathrm{T}_i^{-1} \, \mathbf{S}_i$, T_i^{-1} is a full matrix.

Equation (14) by itself does not offer any advantage, because by using it the global tri-diagonality of the block system would be destroyed. However, the trick is to consider an approximate form of equation (14), namely

$$\delta J_{id} = \sum_{d'=d-a}^{d+a} \sum_{j=1}^{NL+NC} \frac{\partial J_{id}}{\partial x_{jd'}} \, \delta x_{jd'} \,, \tag{16}$$

where a is set either to $a = 0$ (a diagonal form); or $a = 1$ (a tridiagonal form); or possibly even $a = 2$ (pentadiagonal form). Using equation (16), δJ_i can be eliminated from linearized equations, so one is left with linearizing only $NC+NL$ quantities. Because of the method uses an approximate expression for δJ_i, it earned the name approximate Newton-Raphson (ANR) method.

The method is very similar to the hybrid CL/ALI method described earlier (if the latter uses a full ALI setup). The only difference is that in ALI methods one linearizes $\delta J_i = \Lambda^* \, \delta S_i$, where δS_i is expressed through corrections of other state parameters $(\delta n_j, \delta T, \delta n_{\rm e})$, but with Λ^* computed on a fixed optical depth scale. In other words, ANR takes automatically into account a response of radiation field to changes in source function as well as to the optical depth, whereas the standard variant of the CL/ALI method takes into account only the response to changes in the source function.

4.5. Avoiding Linearization

Finally, we come to the most interesting question: could linearization be avoided completely? As we mentioned earlier, using a Λ-iteration-type method (that is, solving one structural equation at a time and iterating between them), is not at all a sufficiently robust method. The fundamental reason is that the intimate coupling between radiation field and other structural parameters is essentially ignored.

On the other hand, we saw in the previous paper (Hubeny, this volume) that the ALI method provides an efficient, robust, and fast scheme to treat such a coupling. We may therefore adopt a different philosophy than that adopted in complete linearization. We view the transfer equation as a fundamental one, and all the other structural equations are viewed as "constraint" equations. The question whether one can avoid linearization in the global stellar atmosphere problem can then be approached in two consecutive steps: (i) how to formulate a multilevel problem (i.e. solving simultaneously radiative transfer and statistical equilibrium) efficiently within the ALI formalism, and without linearization; and (ii) can one add other structural equations (hydrostatic + radiative equilibrium), again without a need to linearize? We shall consider these two questions in the next Section.

5. Role of ALI

5.1. ALI for Multilevel Atoms

To illustrate the basic problem of applying ALI in multilevel problems, we first write down the expression for the radiative rates. For simplicity, we consider only lines; the treatment of continua is analogous. The net transition rate for any line $i \to j$ is

$$R_{ji}^{\rm net} \;=\; n_j A_{ji} \;-\; (n_i B_{ij} - n_j B_{ji}) \bar{J}_{ij} \,, \tag{17}$$

where $\bar{J}_{ij} = \int_0^\infty \phi_{ij}(\nu)\, d\nu$, with $\phi_{ij}(\nu)$ being the normalized absorption profile coefficient for transition $i \to j$. The basic ALI equation gives for $\bar{J}_{ij}$

$$\bar{J}_{ij} \;=\; \Lambda^*[S^{\rm new}] \;+\; (\Lambda - \Lambda^*)[S^{\rm old}] \,. \tag{18}$$

Here the second term, which may be written as $\Delta \bar{J}_{ij}^{\rm old}$, is known from the previous iteration. However, the first term contains $S^{\rm new}$, which is a complicated and generally non–linear function of the "new" populations. An application of ALI thus eliminates radiation intensity from the rate equations, but at the expense of ending with a set of non-linear equations for the populations. This can be dealt with in two different ways:

Linearization The usual way of solving the set of non-linear equations is by applying the Newton-Raphson method. If a diagonal Λ^* is used, the system has a block-diagonal form (i.e., no explicit depth coupling), so one needs to perform ND inversions of $NL \times NL$ matrices per iteration.

Preconditioning. Let us demonstrate the idea of preconditioning on a simple case, where the total source function is given by the line source function $S_{ij} = n_j A_{ji}/(n_i B_{ij} - n_j B_{ji})$ (i.e., the case of non-overlapping lines and no background continuum). Further, assume that Λ^* is a local (diagonal) approximate Λ^* operator so that each of its elements is simply an ordinary number. The net rate (17) may be written, after some simple algebra,

$$R_{ji}^{\rm net} \;=\; n_j A_{ji}(1 - \Lambda_{ji}^*) \;-\; (n_i B_{ij} - n_j B_{ji})\Delta\bar{J}_{ij}^{\rm old}\,, \qquad (19)$$

where terms $(n_i B_{ij} - n_j B_{ji})$ canceled in the first term of equation (19), and one recovered a linear expression. Consequently, the whole set of rate equations is linear in populations.

The term "preconditioning" comes from the fact that instead of having the net rate formulated as a subtraction of two terms that may be orders of magnitude larger than their difference [as in Eq. (17)], the new form is a subtraction of two much smaller quantities that are now of similar magnitude as their difference. Indeed, deep in the atmosphere, $\Lambda^* \to 1$, so that the first term is indeed very small. Similarly, the second term is also small because $\Delta\bar{J}_{ij}$ is small. In other words, the radiative rates are preconditioned. In the context of the ALI approach, this idea was first used by Werner & Husfeld (1985); a systematic study was presented by Rybicki & Hummer (1991, 1992), who have extended it to the case of general overlap of lines and continua.

From the physical point of view, preconditioning is related to an ingenious way to remove inactive (scattering) parts of radiative rates from the rate equations analytically, and, as a by-product, to recover linearity of the ALI form of the rate equations (see, e.g., Rybicki, 1984).

The two approaches outlined above seem, at first sight, to be quite different. Indeed, their original formulations differ quite substantially. However, in a very interesting paper, Socas-Navarro & Trujillo Bueno (1997) showed that both approaches, linearization and preconditioning, are essentially equivalent from the mathematical point of view. These authors showed that linearization takes into account (in an approximate way) the linear response of the radiation field to the perturbations in the source function and in opacity, while preconditioning takes into account the response to the source function perturbations. In a different variant of these approaches, they can actually both take into account (in an approximate way) the linear response of the radiation field to the perturbations in both the source function and in the opacity. The interested reader is referred to the paper cited for more details.

So, the answer the first question posed in the previous Section is that one can indeed avoid linearization, but doing so actually does not offer any substantial benefits.

5.2. Other Constraint Equations

If one adopts a linearization treatment of statistical equilibrium equations with ALI, there is no fundamental problem to include two or three more state parameters (T, $n_{\rm e}$, and N), and to linearize two or three more equations. This was indeed the approach adopted in the pioneering work of Werner (1986), and in

the subsequent developments of the Kiel/Tuebingen code (Werner 1989; Dreizler & Werner 1993; Werner & Dreizler, this volume).

There is a subtle, but important issue here. An ALI treatment of solving simultaneously radiative transfer and statistical equilibrium (sometimes called the restricted non-LTE problem) can be solved separately for individual species, because there is no direct interaction between species (or if there is, like for instance in the case of charge transfer reactions, it may be treated iteratively). This means that instead of solving simultaneously $NL = \sum_a NL_a$ equations (NL is the total number of levels, and NL_a the number of levels for species a), we solve NA independent problems of NL_a equations (NA being the number of species). Obviously, $NL^3 \gg \sum_a NL_a^3$, so the savings of computer time are considerable.

However, other structural equations (hydrostatic and radiative equilibrium, charge conservation) couple all species (either through the total opacity and emissivity, or through the total charge). Because of that, one has to linearize the whole set, and thus to invert matrices $(NL + NC) \times (NL + NC)$. This shows that being able to treat the other constraint equations separately would lead to a large reduction of computer time per iteration.

Can this be done? Treating the charge conservation separately is easy, in particular for hot atmospheres where the material is almost completely ionized, so $n_{\rm e}$ is nearly linearly proportional to density. As shown by Gustafsson (1971), Werner (1986) and others, treating hydrostatic equilibrium equation separately is also easily possible. The only obstacle is then the radiative equilibrium equation. Methods that solve it separately are traditionally called "temperature correction" methods. They were being used extensively before the advent of complete linearization, and are still being used for LTE models. However, it was amply demonstrated that such methods may converge very slowly or to fail to converge altogether in the case of non-LTE models. But, because of the possibility of achieving a significant reduction of computer time in ALI-based methods, there has been a recent revival of interest in them. Temperature corrections were used successfully in code PHOENIX (Hauschildt, this volume, and reference therein), in particular for rapidly expanding atmospheres. This is not surprising because expansion leads to desaturation of the lines, and a localization of state parameters. It was also tested in the context of static plane-parallel atmospheres (Werner & Dreizler 1999). The question whether there is a temperature correction procedure that can provide a sufficiently robust scheme for all applications is, however, not yet settled. Because of its potential importance, the topic of temperature correction is discussed in other papers in this volume (Dreizler; Hauschildt; Koesterke).

5.3. Hydrid CL/ALI Method

We now return to the hybrid CL/ALI method mentioned in § 4.2., and put it in a different perspective. The ALI-based methods discussed above start with the multi-level formulation of ALI, and linearize the resulting set, together with other structural equations. In contrast, the hybrid CL/ALI scheme (and the ANR method too), start with the linearized structural equations and then eliminate δJ_ν from the linearized set using ALI. Although both approaches are very similar, the hybrid scheme offers a significant benefit, as pointed out by Hubeny & Lanz (1995): While the ALI treatment is used for most frequency

points, the radiation intensity at a few selected frequency points may still be linearized. The method thus offers a wide spectrum of options, ranging from the full CL to the full ALI method. It was shown that by selecting a few (typically 10-30) frequency points judiciously (typically, at the head of the most opaque continua, like the hydrogen and He II Lyman continua; and in the centers of strongest lines), the computer time per iteration is essentially the same as in the case of full ALI, while the number of iterations is essentially the same as in the case of full CL, i.e. much lower than in the case of true ALI. The method thus combines two major advantages of its two constituents, namely the convergence rate being virtually as high as for the standard CL method, while the computer time per iteration is almost as low as for the standard ALI method.

6. Metal Line Blanketing

By the term line-blanketed model atmospheres we understand models that take into account effects of "all' lines of all important species. There are literally millions of lines that contribute to the opacity; their number is even a few orders of magnitude higher when considering molecular lines (see papers by Jorgensen and by Alexander in these Proceedings).

To be able to treat metal line blanketing numerically, a useful method must be able to work efficiently with a large number of frequency points, and a large number of energy levels (populations). As discussed in § 4.2., the latter problem is effectively solved by using the concept of superlevels, possibly together with level-grouping if one works within the framework of a linearization method.

The problem of a large number of frequencies is dealt with by the an application of the ALI method, or by the ANR method. However, although the frequency points are effectively eliminated from the state vector and thus from the linearization process, their number may still be too large to prevent handling all necessary frequencies even in the formal solution of the transfer equation.

To estimate the number of frequency points ideally needed, let us assume that metal lines are practically everywhere, so one should cover all the frequency range by frequency points that are spaced proportionally to the value of a fiducial Doppler width; $\Delta\nu_D^* = \nu\, v^*/c$, where ν is the frequency, v^* a characteristic velocity (given either by a thermal velocity of a characteristic species, e.g. Fe), or a characteristic turbulent velocity, whichever is larger, and c the light speed. We require the frequency resolution of $a\,\Delta\nu_D^*$, where a is an adjustable parameter (ideally, a value below 1).

The total number of frequency points would then be

$$NF \approx \int_{\nu_{\min}}^{\nu_{\max}} \frac{1}{a\,\Delta\nu_D^*}\, d\nu = \frac{c}{a\,v^*}\, \ln(\nu_{\max}/\nu_{\min})\,, \tag{20}$$

Assuming for example that we need to cover about three decades in frequency, and taking $a = 0.75$, we would need about 10^5 frequency points for hot models ($v^* \approx 30$ km/s), while we need about 10^6 points for cool models ($v^* \approx 3$ km/s), or even more for cooler and low-microturbulence models.

It is now becoming possible to accommodate such numbers in modeling codes, thanks to a dramatic increase of available computer memory and speed. Only few years ago, however, this was not possible, and therefore one had to

use approximations to reduce number of needed formal solutions of the transfer equation. There are essentially two possibilities:

- *Opacity Distribution Functions* (ODF). In NLTE, this method is used in conjunction with the concept of superlevels. The transitions between superlevels are called superlines. The idea is to resample a complicated frequency dependence of the superline cross-section to form a monotonic function of frequency; this function is then represented by a small number of frequency quadrature points (Anderson 1989; Hubeny & Lanz 1995).
- *Opacity Sampling* (OS). The idea is a simple Monte Carlo-like sampling of frequency points of the superline cross-sections (Anderson 1989; Dreizler & Werner 1993). The advantage of this approach is that it can easily treat line blends and overlaps; the disadvantage is that considering too few frequency points may easily lead to missing many important line cores. On the other hand, the "exact" method is in fact a variant of the OS with a sufficiently high resolution, as discussed above.

An explicit comparison between results using the ODF and the OS approaches, and with various frequency resolutions in the latter, is presented e.g. in Lanz & Hubeny (this volume).

7. Program TLUSTY

7.1. General Characteristics

The computer program TLUSTY has been described in several papers; Hubeny (1988) – the original CL version; Hubeny & Lanz (1992) – implementation of Ng and Kantorovich accelerations; Hubeny, Hummer & Lanz (1994) – treatment of level dissolution and occupation probabilities; Hubeny & Lanz (1995) – hybrid CL/ALI method, concept of superlevels and superlines; and Lanz & Hubeny (2001) – opacity sampling method.

The program solves the basic equations (radiative transfer, hydrostatic equilibrium, radiative equilibrium, statistical equilibrium, charge and particle conservation). However, not all of the basic equations actually have to be solved. The program has options for omitting some of the equations, while keeping the corresponding quantities fixed. For instance, one may keep the temperature fixed and skip the radiative equilibrium equation (all the other equations being solved exactly); this corresponds to calculating so-called semi-empirical models.

Recently, the previously separate variant called TLUSDISK was combined into one universal TLUSTY, which thus allows one to compute either a model stellar atmosphere, or the vertical structure of a given annulus in an accretion disks. Accretion disk models are described in detail by Hubeny & Hubeny (1998), and Hubeny et al (2001). Recent upgrades contain an improved treatment of convection (with several variants of the mixing-length formalism); external irradiation; Compton scattering (described in Hubeny et al. 2001); dielectronic recombination; and X-ray opacities, including the inner-shell (Auger) ionization (described in Hubeny et al. 2001). On the low-temperature side, we have recently developed a variant called COOLTLUSTY (described briefly in Burrows et al. 2002) to work with pre-calculated opacity tables. We currently use opacity tables appropriate for brown, T, and L dwarfs, and giant planets.

The program is fully data-oriented as far as the choice of atomic species, ions, energy levels, transitions, and opacity sources is concerned. We stress that there are no default opacities built in (only some default formulae for various cross-sections are included – see paper by Lanz & Hubeny in this Proceedings).

7.2. Numerical Methods

TLUSTY uses the hybrid CL/ALI scheme. There are several formal solvers of the transfer equation available: Feautrier (2-nd or 4-th order), or Discontinuous Finite Element scheme (DFE – Castor et al 1992). The Λ^*-operator can be either diagonal or tri-diagonal, computed as the corresponding part of the exact Λ (see Hubeny, this volume). If the Feautrier method is used for a formal solver, Λ^* is evaluated using Rybicki-Hummer (1991) procedure; if DFE is used, Λ^* is evaluated as described in Hubeny (this volume).

To further reduce the number of quantities to be linearized, we have implemented the concepts of superlevels, level grouping, and level zeroing, as described in §4.2. The corresponding setups are all data-oriented. To reduce the overall number of iterations, we implemented the Ng acceleration, the Kantorovich method, as well as the successive over-relaxation method, again with all setups being driven by input data.

To improved convergence speed, one can perform an optional number of iterations solving simultaneously the transfer and the statistical equilibrium equations (together with the charge conservation) between two successive iterations of the global linearization scheme, keeping temperature and density fixed. The iterations can be either ordinary Lambda iterations, or iterations of the multilevel atom problem with preconditioning, as described in the § 5.1., for all species separately.

Both options, ODF and Opacity Sampling, are offered for a treatment of metal line blanketing. Again, all details of the actual setup are driven by input data. Until recently, most of our blanketed models were computed using the ODF approach, which required only about 30,000 - 50,000 frequency points (e.g., Lanz et al. 1996a, 1996b; Lanz, Hubeny, & Heap 1987); however our recent models (including the grid described by Lanz & Hubeny in this volume) use the Opacity Sampling Method with a sampling of 0.75 fiducial Doppler-widths. Hot (OB) star models need typically about 200,000 frequency points.

Acknowledgments. We thank Dimitri Mihalas for a careful reading of the manuscript and for useful suggestions.

References

Anderson, L. S. 1985, ApJ, 298, 848

Anderson, L. S. 1989, ApJ, 339, 588

Auer, L.H. 1987, in Numerical Radiative Transfer, ed. W. Kalkofen (Cambridge Univ. Press, Cambridge), 101

Auer, L.H. 1991, in NATO ASI Series C 341, Stellar Atmospheres: Beyond Classical Models, ed. L. Crivellari, I. Hubeny, & D.G.Hummer, (Dordrecht: Kluwer), 9

Auer, L.H., & Mihalas, D. 1969, ApJ, 158, 641

Auer, L.H., & Mihalas, D. 1970, MNRAS, 149, 60

Burrows, A., Burgasser, A.J.; Kirkpatrick, J.D., Liebert, J., Milsom, J.A., Sudarsky, D., & Hubeny, I. 2002, ApJ, 573, 394

Castor, J.I., Dykema, P.G., & Klein, R.I. 1992, ApJ, 387, 561

Dorfi, E. & Druri, L. 1987, J. Comp. Phys., 69, 175

Dreizler, S., & Werner, K. 1993, A&A, 278, 199

Gustafsson, B. 1971, A&A, 10, 187

Hempe, K., & Schönberg, K. 1986, A&A, 160, 141

Hillier, D.J. 1990, A&A, 231, 116

Hillier, D.J. & Miller, D.L. 1998, ApJ, 496, 407

Hubeny, I. 1988, Computer Physics Comm., 52, 103

Hubeny, I. 1992, in Lecture Notes in Phys. 401, The Atmospheres of Early-Type Stars, ed. U. Heber & C.J. Jeffery, (Berlin: Springer), 377

Hubeny, I., Agol, E., Blaes, O., & Krolik, J.H., 2001, ApJ, 559, 680

Hubeny, I., & Hubeny, V. 1998, ApJ, 505, 558

Hubeny, I., Hummer, D.G., & Lanz, T. 1994, A&A, 282, 151

Hubeny, I., & Lanz, T. 1992, A&A, 262, 501

Hubeny, I., & Lanz, T. 1995, ApJ, 439, 875

Kantorovich, L.V. 1949, Trudy Math. Inst. Steklova, XXVIII, 104

Koesterke, L., Hamann W.-R., Kosmol, P. 1992, A&A, 255, 490

Kubát, J., Puls, J., & Pauldrach, A.W.A. 1999, A&A, 341, 587

Kurucz, R.L. 1994, Kurucz CD ROM No. 1 (Smithsonian Astrophys. Obs.)

Lanz, T., Barstow, M.A., Hubeny, I., & Holberg, J.B. 1996a, ApJ 473, 1089

Lanz, T., de Koter, A., Hubeny, I., & Heap, S.R. 1996b, ApJ, 465, 359

Lanz, T., Hubeny, I., & Heap, S.R. 1997, ApJ, 485, 843

Mihalas, D. 1978, Stellar Atmospheres, 2nd Ed., (San Francisco: Freeman)

Mihalas, D. , Heasley, J.N., & Auer, L.H. 1975, A non-LTE model stellar atmospheres computer program, NCAR-TN/STR-104

Ng, K.C. 1974, J. Chem. Phys., 61, 2680.

Osterbrock, D.E. 1989, Astrophysics of Gaseous Nebulae and Active Galactic Nuclei, (Mill Valley: University Science Books)

Rybicki, G.B. 1984, in Methods in Radiative Transfer, ed. W. Kalkofen, (Cambridge: Cambridge Univ. Press), 21

Rybicki, G.B., & Hummer, D.G. 1991, A&A, 245, 171

Rybicki, G.B., & Hummer, D.G. 1992, A&A, 262, 209

Socas-Navarro, H., & Trujillo Bueno, J. 1997, ApJ, 490, 383

Trujillo Bueno, J., & Fabiani Bendicho, P. 1995, ApJ, 455, 646

Werner, K. 1986, A&A, 161, 177

Werner, K. 1989, A&A, 226, 265

Werner, K., & Dreizler, S. 1999, J. Comp. Appl. Math., 109, 65

Werner, K., & Husfeld, D. 1985, A&A, 148, 417