# Chapter 12
# Numerical Procedure

For realistic material functions no analytic solutions are possible, so that one depends all the more on numerical solutions of the basic differential equations. Consequently the activity and the number of results in this field has increased with the numerical capabilities. The growth of computing facilities by leaps and bounds since the 1960s may be illustrated by a remark of Schwarzschild (1958): "A person can perform more than twenty integration steps per day", so that "for a typical single integration consisting of, say, forty steps, less than two days are needed". The situation has changed drastically since those days when the scientist's need for meals and sleep was an essential factor in the total computing time for one model. Nowadays one asks rather for the number of solutions produced per second. And these modern solutions are enormously more refined (numerically and physically) than those produced 40 years ago. This progress has been possible because of the introduction of large and fast electronic computers and the simultaneous development of an adequate numerical procedure connected with the name of L.G. Henyey. His method for calculating models in hydrostatic equilibrium is now generally used and will be described later. For more details and for further references see Kippenhahn et al. (1967). If inertia terms with $\ddot{r} \neq 0$ become important, one needs a so-called "hydrodynamic" procedure (see Sect. 12.3).

## 12.1 The Shooting Method

It is not difficult to see that the appropriate choice of a numerical procedure is anything but a trivial matter. Consider the simplest case, the calculation of a model in complete equilibrium at a given time, for given mass $M$ and given chemical composition $X_i(m)$. The "spatial problem" can then be separated and is described by the structure equations (10.1), (10.2), (10.4) and (10.16). The naïve attempt simply to integrate them from one boundary to the other would encounter the difficulty that the boundary conditions are split, one pair being given at the centre,

R. Kippenhahn et al., *Stellar Structure and Evolution*, Astronomy and Astrophysics Library, DOI 10.1007/978-3-642-30304-3_12, © Springer-Verlag Berlin Heidelberg 2012

the other at the surface. Moreover, a test calculation starting with trial values $P_c, T_c$ at the centre has little chance of meeting the correct surface conditions. Outward integrations differing only a little near the centre have the tendency to diverge strongly when approaching the surface (see Sect. 11.3). The reason is that for radiative transport (10.4) with (10.6) contains the factor $T^{-4}$. For inward integrations starting with trial values $R, L$ at the surface another divergence occurs near the centre owing to the singularity produced by the factor $r^{-4}$ in (10.2).

A compromise between these two possibilities is a fitting procedure often used in earlier, non-automized computations. Outward and inward integrations were both carried to an intermediate fitting point, where they were fitted smoothly to each other by a gradual variation of the trial values $P_c$, $T_c$ and $R, L$. The simultaneous fit of four variables $(r, P, T, l)$ is, in principle, possible, since one can vary four free parameters $(P_c, T_c, R, L)$ in the partial solutions. The fitting point is preferably chosen to be at the interface between physically different regions. For example, one takes the border between a convective central core and a radiative envelope, or between regions of different composition.

Fitting methods turned out to be unsuitable for calculating large series of complicated models. For these purposes they were generally replaced by the Henyey method. There are, however, certain applications where a fitting method is still unsurpassed, for example, if one wishes to find *all* possible solutions for given core and envelope parameters. Another application is the generation of the very first model for an evolutionary sequence, since the relaxation methods, which will be introduced in the next section, always need a trial model for finding a solution. For chemically homogeneous stars the shooting methods are well suited to construct such initial models.

## 12.2 The Henyey Method

This method is very practical, especially for solving boundary-value problems where the conditions are given at both ends of the interval. A trial solution for the whole interval is gradually improved upon in consecutive iterations until the required degree of accuracy is reached. In each iteration, corrections to *all* variables at *all* points are evaluated in such a way that the effect of each of them on the whole solution (including the boundaries) is taken into account. In a generalized Newton–Raphson method, corrections are obtained from linearized algebraic equations.

For spherical stars in hydrostatic equilibrium we have the partial differential equations (10.1)–(10.5) together with boundary conditions at the centre and at the surface. In addition the proper initial values have to be specified as well as the stellar mass $M$. The general structure of the system of equations suggests that one should treat two subsystems separately and alternately. First, the system (10.1)–(10.4) is solved for given $X_i(m)$, then (10.5) is applied to a small time step $\Delta t$, after which (10.1)–(10.4) is solved for the new values of $X_i(m)$, and so on. In modern language such an approach is called *operator splitting*. In this way one

can construct a whole evolutionary sequence of models (But one should be aware of the fundamental inconsistency inherent to this approach, which was discussed in Chap. 10.). We now describe in detail the first of these two steps, the solution of the "spatial system".

If there is complete equilibrium ($\ddot{r} = \dot{P} = \dot{T} = 0$), the initial values to be given are the $X_i(m)$, so that we can treat them as known parameters for any point. According to (10.7)–(10.14) the material functions $\varepsilon, \kappa, \varrho, \ldots$ on the right-hand sides of (10.1), (10.2), (10.4) and (10.16) can be replaced by their dependencies upon $P$ and $T$. Then we have to solve the four ordinary differential equations (10.1), (10.2), (10.4) and (10.16) for the four unknown variables $r, P, T, l$ in the interval $[0, M]$ (where $M$ is also thought to be given).

The case of hydrostatic equilibrium ($\ddot{r} = 0$) but thermal non-equilibrium ($\dot{P} \neq 0, \dot{T} \neq 0$) is almost equivalent, the only difference being the additional term $\varepsilon_g$ in (10.3), which contains the partial derivatives $\dot{P}$ and $\dot{T}$. This requires as initial values for the earlier time $t_0 - \Delta t$ not only the $X_i(m)$ but also $T(m)$ and $P(m)$ (See the remarks on possible initial values in Chap. 10.). Assume that we take them from a "foregoing" solution, calling these given functions $P^*(m), T^*(m)$. At any point $m = m_j$, we denote the variables by $P_j, T_j$ and replace the time derivatives $\dot{P}_j, \dot{T}_j$ by

$$\dot{P}_j = \frac{1}{\Delta t}(P_j - P_j^*), \quad \dot{T}_j = \frac{1}{\Delta t}(T_j - T_j^*) \,. \tag{12.1}$$

The given values of $\Delta t, P_j^*, T_j^*$ can now be considered known parameters. Then $\dot{P}_j, \dot{T}_j$ are functions of $P_j, T_j$ only, as is the case with all material functions, and therefore we can also consider $\varepsilon_g$ to be replaced by the function $\varepsilon_g(P, T)$, and the situation is as before with the complete equilibrium models: we again have the four ordinary differential equations (10.1)–(10.4) for the four unknown variables $r, P, T, l$, but with a somewhat different right-hand side of (10.3).

Let us write these four differential equations briefly as

$$\frac{dy_i}{dm} = f_i(y_1, \ldots, y_4), \quad i = 1, \ldots, 4 \,, \tag{12.2}$$

where we have used the abbreviations $y_1 = r, y_2 = P, y_3 = T, y_4 = l$. The next step is discretization, i.e. we proceed from the differential equations (12.2) to corresponding difference equations for a finite mass interval $[m^j, m^{j+1}]$. Let us denote the variables at both ends of this interval by upper indices, for example, $y_1^j, y_1^{j+1}, \ldots, y_4^j, y_4^{j+1}$. The functions $f_i$ on the right-hand sides of (12.2) have to be taken for some average arguments we call $y_i^{j+1/2}$; they are a combination of $y_i^j$ and $y_i^{j+1}$, for example, the arithmetic or the geometric mean. If we define the four functions

$$A_i^j := \frac{y_i^j - y_i^{j+1}}{m^j - m^{j+1}} - f_i\left(y_1^{j+1/2}, \ldots, y_4^{j+1/2}\right), \quad i = 1, \ldots, 4 \,, \tag{12.3}$$

then the difference equations replacing (12.2) for the mass interval between $m_j$ and $m_{j+1}$ are

$$A_i^j = 0, \quad i = 1, \ldots, 4 \,. \tag{12.4}$$

The difference equations (12.4) and (12.1) represent a linearization of the differential equations and are therefore an approximation, the accuracy of which has to be controlled. Obviously, the smaller $\Delta t$ and $\Delta m^j = m^j - m^{j-1}$, the better the approximation. In practical circumstances the spatial discretization is not constant throughout the stellar model, but depends on the changes of the physical variables. A good approach is to choose $\Delta m^j$ for each $j$ such, that all variables change by less than a predefined upper limit between points $j$ and $j-1$. That maximum change will differ between variables and has to be determined by numerical experiments reducing it to a limit from where on the numerical solution no longer depends on the $\Delta m^j$ significantly. Apart from this basic control algorithm there are more advanced methods, which, for example, take into account not only the slope but the curvature of the functions $T(m)$, $P(m)$ (Wagenhuber and Weiss 1994). The advantage of this method is that it is sensitive to deviations from linear behaviour. It places many grid points where the variables are a strongly non-linear function of $m$, while it uses very few in the opposite case. Wagenhuber called this the *curvature method*, as opposed to the simpler *gradient method*.

It is possible to exclude the outermost envelope of the star from the iteration procedure, since time-consuming computations may be necessary for this part (e.g. partial ionization and superadiabatic convection). With sufficient computing power this is no longer a necessity, however. Another situation where this would be advisable is when fully realistic atmospheres are to be connected to the interior of the star, since the diffusion approximation (10.6) is not valid at $m = M$ but at some deeper layer where the optical depth $\tau \gg 1$. The lower boundary of such an atmosphere then provides the upper boundary of the interior model. As described in Sect. 11.2 the outer boundary conditions are imposed at a fitting mass $m_\mathrm{F}$, which may have the special value $m = M$ and may have the upper index $j = 1$, and they are formulated by the two equations (11.18) that relate the variables $y_1^1, \ldots, y_4^1$ at $m^1 = m_\mathrm{F}$. These equations are specific choices and may differ. With the definitions

$$B_1 := y_2^1 - \pi\left(y_1^1, y_4^1\right), \quad B_2 := y_3^1 - \theta\left(y_1^1, y_4^1\right) \,, \tag{12.5}$$

equations (11.19) become

$$B_i = 0 \,, \quad i = 1, 2 \,. \tag{12.6}$$

As described in Sect. 11.2 the functions $\pi, \theta$ have to be derived by "downward" integrations starting with different trial values of $R, L$. In practice this may be greatly simplified if we content ourselves with a linear approximation for $\pi$ and $\theta$ (i.e. taking the tangential planes instead of the complicated surfaces in Fig. 11.1). Then only three trial integrations suffice to determine all coefficients in $B_1$ and $B_2$.

In the innermost interval of $m$, between the central point $m^K (= 0)$ and $m^{K-1}$, we apply series expansions for all four variables as given by (11.3), (11.4), (11.6) and (11.9). These four equations are written as

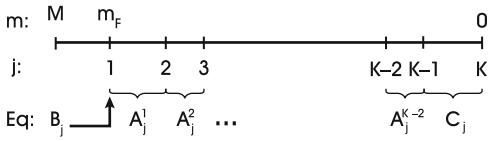


**Fig. 12.1** Sketch of the mesh points in the interior solution, from the fitting mass $m = m_{\rm F}$ (in this example $m_{\rm F} < M$) to the centre ($m = 0$). It is also indicated which of the equations (12.4), (12.6) and (12.7) have to be fulfilled at $m_{\rm F}$ or between two adjacent mesh points

$$C_i\left(y_1^{K-1}, \ldots, y_4^{K-1}, y_2^K, y_3^K\right) = 0, \quad i = 1, \ldots 4\,, \tag{12.7}$$

which already incorporates the central boundary conditions $y_1^K = y_4^K = 0$ (i.e. $r = l = 0$ at the centre).

Consider now the whole interval of $m$, between $m^K = 0$ and the fitting mass $m^1 = m_{\rm F}$, to be divided into $K - 1$ intervals (usually not equidistant) by $K$ mesh points as sketched in Fig. 12.1. At these $K$ mesh points we have $(4K - 2)$ unknown variables (since $y_1^K = y_4^K = 0$), and in order to have a solution, these unknowns have to fulfil the following equations: (12.6) for the outer boundary, (12.4) for each interval except the last one ($j = 1, \ldots, K-2$), and (12.7) for the central boundary; thus there are $2 + 4(K - 2) + 4 = 4K - 2$ equations, which may be written:

$$\begin{aligned} B_i &= 0, \quad i = 1, 2\,, \\ A_i^j &= 0, \quad i = 1, \ldots, 4, \quad j = 1, \ldots, K-2\,, \\ C_i &= 0, \quad i = 1, \ldots, 4\,. \end{aligned} \tag{12.8}$$

Suppose that we are looking for a solution for given values of $M, X_i(m)$, $P^*(m), T^*(m)$ (which all enter into these equations as parameters). And suppose, furthermore, that we have a first approximation to this solution, say, $(y_i^j)_1$ with $i = 1, \ldots, 4, j = 1, \ldots, K$ (This may be a rough first guess, e.g. obtained by an extrapolation of a foregoing solution or a solution for similar parameters. It may also be obtained from a shooting method.). Since the $(y_i^j)_1$ are only an approximation, they will not fulfil (12.8), i.e. when we use them as arguments in the functions $A_i^j$, $B_i$, and $C_i$, we find that

$$B_i(1) \neq 0, \quad A_i^j(1) \neq 0, \quad C_i(1) \neq 0\,, \tag{12.9}$$

where we indicate by (1) that the first approximation is used as arguments. Let us now look for corrections $\delta y_i^j$ for all variables at all mesh points such that the second approximation

$$\left(y_i^j\right)_2 = \left(y_i^j\right)_1 + \delta y_i^j \tag{12.10}$$

of the arguments makes the $B_i, A_i^j$, and $C_i$ vanish. The changes $\delta y_i^j$ of the arguments produce the changes $\delta B_i, \delta A_i^j$, and $\delta C_i$ of the functions, and we obviously have to require that

$$B_i(1) + \delta B_i = 0, \quad A_i^j(1) + \delta A_i^j = 0, \quad C_i(1) + \delta C_i = 0\,. \tag{12.11}$$

For small enough corrections, we may expand the $\delta B_i, \ldots$ in terms of increasing powers of the corrections $\delta y_i^j$, and keep only the linear terms in this expansion; for example,

$$\delta B_1 \approx \frac{\partial B_1}{\partial y_1^1}\delta y_1^1 + \frac{\partial B_1}{\partial y_2^1}\delta y_2^1 + \frac{\partial B_1}{\partial y_3^1}\delta y_3^1 + \frac{\partial B_1}{\partial y_4^1}\delta y_4^1\,. \tag{12.12}$$

For (12.5) the third term would vanish because in this special case $B_1$ is independent of $y_3$. With this linearization (12.11) can be written as

$$\begin{aligned}
\frac{\partial B_i}{\partial y_1^1}\delta y_1^1 + \cdots + \frac{\partial B_i}{\partial y_4^1}\delta y_4^1 &= -B_i\,, \\
i &= 1, 2\,, \\
\frac{\partial A_i^j}{\partial y_1^j}\delta y_1^j + \cdots + \frac{\partial A_i^j}{\partial y_4^j}\delta y_4^j + \frac{\partial A_i^j}{\partial y_1^{j+1}}\delta y_1^{j+1} + \cdots + \frac{\partial A_i^j}{\partial y_4^{j+1}}\delta y_4^{j+1} &= -A_i^j\,, \\
i &= 1, \ldots, 4, \quad j = 1, \ldots, K-2\,, \\
\frac{\partial C_i}{\partial y_1^{K-1}}\delta y_1^{K-1} + \cdots + \frac{\partial C_i}{\partial y_4^{K-1}}\delta y_4^{K-1} + \frac{\partial C_i}{\partial y_2^K}\delta y_2^K + \frac{\partial C_i}{\partial y_3^K}\delta y_3^K &= -C_i\,, \\
i &= 1, \ldots, 4\,.
\end{aligned} \tag{12.13}$$

(The $B_i, A_i^j, C_i$, and all derivatives have here to be evaluated using the first approximation as arguments.) This is a system of $2 + 4(K-2) + 4 = 4K - 2$ linear, inhomogeneous equations for the $4K - 2$ unknown corrections $\delta y_i^j$ ($i = 1, \ldots, 4$ and $j = 1, \ldots, K$; but $\delta y_1^K = \delta y_4^K = 0$ because of the central boundary conditions). Equation (12.13) may be written concisely in matrix form as

$$H \begin{pmatrix} \delta y_1^1 \\ . \\ . \\ . \\ \delta y_3^K \end{pmatrix} = - \begin{pmatrix} B_1 \\ . \\ . \\ . \\ C_4 \end{pmatrix}, \tag{12.14}$$

where the matrix $H$ of the coefficients is called the *Henyey matrix*; its elements are the derivatives on the left-hand sides of (12.13).

Usually $H$ has a non-vanishing determinant, $\det H \neq 0$, and we can solve these linear equations, obtaining the wanted corrections $\delta y_i^j$. These are applied as shown in (12.10) to obtain a second, better approximation $(y_i^j)_2$. When using these

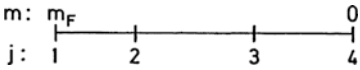


**Fig. 12.2** Mesh points in the "three-layer model"

second approximations as arguments, we will generally still find $B_i \neq 0$, $A_i^j \neq 0$, and $C_i \neq 0$, i.e. equations (12.8) are not yet fulfilled. This is because the corrections were calculated from the *linearized* equations (12.13), while equations (12.8) are non-linear (Even if we had linear equations instead of (12.8), the solution might require several iterations, since the numerical solution of (12.13) has only limited accuracy.). Therefore in a second iteration step we calculate new corrections by the same procedure to obtain a third approximation

$$\left(y_i^j\right)_3 = \left(y_i^j\right)_2 + \delta y_i^j \ , \tag{12.15}$$

and so on. In consecutive iterations of this type, the approximate solution can be improved until either the absolute values of all corrections $\delta y_i^j$, or the absolute values of all right-hand sides in (12.13), drop below a chosen limit. Then we have approached the solution with the required accuracy.

If a time sequence of models is to be produced, one can now change the parameters appropriately for a new small time step $\Delta t$ [by evaluating from (10.5) the change of the $X_i(m)$, and by redefining the just-calculated $P(m), T(m)$ as the new $P^*(m), T^*(m)$]. The new model for $t + \Delta t$ is then calculated by the Henyey method in the same manner as for the model for $t$.

Of course, there is no guarantee that the iteration procedure for improving the approximations really does converge. In fact often enough one finds divergence if the chosen approximation is too far from the solution; then the required corrections are so large that one cannot neglect the second-order terms when evaluating $\delta B_i, \delta A_i^j$, and $\delta C_i$ in (12.11), and the linearized equations (12.14) therefore yield wrong corrections.

What happens, on the other hand, if we take a given precise solution as the "first approximation"? It fulfils (12.8) such that the right-hand sides of (12.14) vanish. Equation (12.14) is then a system of *homogeneous* linear equations, which for det $H \neq 0$ has only the trivial solution $\delta y_i^j = 0$: in this (normal) case, there is no other solution ("local uniqueness" as mentioned in Sect. 12.6). If, however, det $H = 0$, then we obtain solutions $\delta y_i^j \neq 0$, i.e. other solutions for the same parameters. In this somewhat pathological situation the "local uniqueness" of the solution is violated.

The Henyey matrix and its determinant are obviously important quantities. This concerns also their connection with the stability properties (see Sect. 12.6). It is worthwhile noting the general structure of $H$, which turns out to be very simple. This is most easily demonstrated by considering the simple "three-layer model", which has only four mesh points from centre to fitting mass (Fig. 12.2). One interval is adjacent to $m_F$, one to the centre, while the intermediate interval

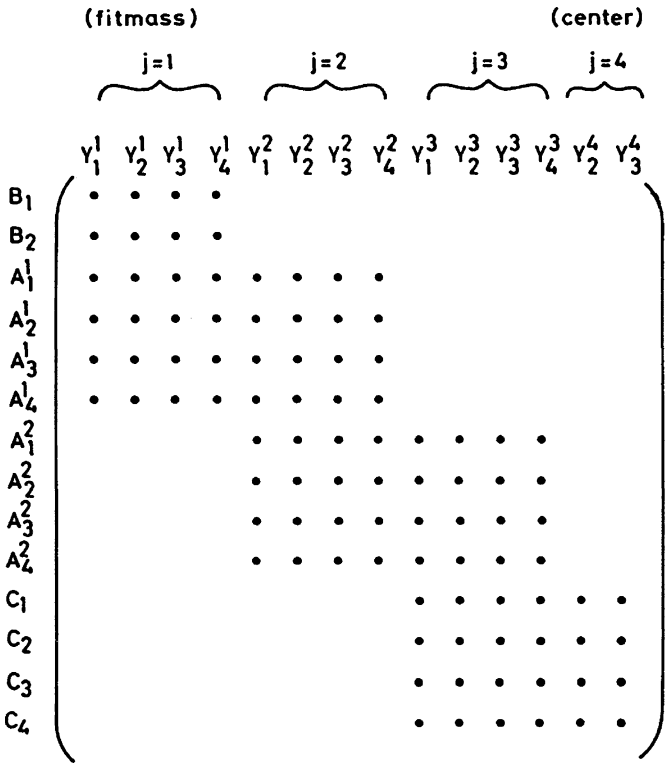


**Fig. 12.3** Structure of the Henyey matrix $H$ for the three-layer star sketched in Fig. 12.2. A *dot* in, for example, the column $y_i^j$ and the row $A_k^l$ represents the matrix element $\partial A_k^l / \partial y_i^j$. All matrix elements outside the *dotted area* are zero

borders on neither of these two boundaries, so that the full generality of possible cases is exhibited. Any further mesh point will only duplicate the situation of the intermediate interval. The Henyey matrix $H$ for this three-layer star is indicated in Fig. 12.3, where a dot in a column under $y_i^j$ and in a row denoted at the left-hand side by $A_k^l$ represents a matrix element $\partial A_k^l / \partial y_i^j$. Some of these derivatives will be zero, since some basic equations do not depend on all variables [e.g. (10.16) does not contain $y_1 = r$]. Outside the dotted area there are only zero elements, because the first-order scheme (12.13) connects only neighbouring points. The Henyey matrix therefore has non-vanishing elements only in overlapping blocks along the main diagonal, so that this can be easily used for devising simple and well-behaved algorithms for computing det $H$ and inverting the matrix through elimination processes. The most widely used method for solving such block matrices in stellar evolution codes is that by Henyey et al. (1964), which was described in all details by Kippenhahn et al. (1967). The basic idea is to express the corrections of the block matrix connecting points $(j, j + 1)$ in terms of the quantities of the next

block $(j+1, j+2)$, and so on. At the end there is a final block (usually the innermost one), for which the corrections are determined by matrix inversion, and from which on then all the other corrections can be calculated by going backwards again. The Henyey method has $K$ inversions of matrices of size $4 \times 8$ instead of straightforwardly inverting the Henyey matrix of size $K \times K$. It therefore grows only linearly–instead of quadratically–with increasing number of grid points.

## 12.3 Treatment of the First- and Second-Order Time Derivatives

When devising a numerical scheme for solving our partial differential equations one can choose many details more or less arbitrarily without greatly affecting the results. This concerns questions such as the prescription for averaging between spatial mesh points, and the definition of the variables; these can be, for example, the physical quantities themselves, their logarithms, or any other functions describing them properly.

Concerning the manner in which the time derivatives are approximated, one distinguishes between explicit and implicit schemes that are known to behave differently, in particular when one is dealing with second-order time derivatives. Forward integration in time, starting from given initial values, can require time steps of various length, and the results can be unstable with respect to small numerical errors. In Sect. 12.2 we encountered examples of both types of scheme:

An *explicit* scheme was indicated in the case of the chemical equations (10.5). Consider the time interval between $t^n$ (at which all variables $q^n$ are supposed to be known) and $t^{n+1}$ (for which the variables $q^{n+l}$ are to be calculated). We may use (10.5) simply in order to calculate time derivatives $\dot{X}_i^n$ of the chemical composition from the known reaction rates $r_{ik}^n$ and densities $\varrho^n$. The composition for $t^{n+l}$ is then evaluated as $X_i^{n+1} = X_i^n + \Delta t \dot{X}_i^n$ before the other variables for this time are derived. In fact the $X_i^{n+1}$ are used as fixed parameters when calculating the solution at $t^{n+1}$ by iteration. Such a procedure is relatively simple, and in general, the results can be sufficiently accurate if the time steps are kept small enough. However, there is no guarantee to prevent unphysical solutions in explicit methods. For example, if $\dot{X}_i^n$ is sufficiently negative even a small time step might lead to a negative $X_i^{n+1}$. To prevent this, an *implicit* treatment is indicated. If $\dot{X}_i$ depends on the chemical abundances itself, as is the case for the nuclear reactions (10.5), the abundance at $t^{n+1}$ is used on the right-hand side, too. This constitutes a set of implicit equations, which need to be solved by inversion methods, but which are numerically stable. An easy way is by writing $X_i^{n+1} = X_i^n + \Delta t \dot{X}_i^n = X_i^n + \Delta X_i^n$ and linearizing the equations in the $\Delta X_i^n$, neglecting all higher terms. The resulting system of equations is linear in $\Delta X_i^n$ and can be solved by one matrix inversion. However, the quality of the linearization again depends on the size of $\Delta t$. Such

implicit schemes are generally used to solve networks of nuclear reactions, where the terms $\dot{X}_i^n$ may vary by many order of magnitudes.

In the set of structure equations (10.1)–(10.4) to be solved at time $t_i^{n+1}$ for given $X_i^{n+1}$ the energy equation (10.3) contains the time derivatives of $\dot{P}$ and $T$. With respect to these an *implicit* scheme was used in Sect. 12.2. According to (12.1) the $\dot{P}$ and $\dot{T}$ are replaced by $(P^{n+1}-P^n)/\Delta t$ and $(T^{n+1}-T^n)/\Delta t$, respectively. These time derivatives are therefore considered to depend also on the variables at time $t^{n+1}$ and are evaluated together with them in the iteration procedure. In principle one could also have used an explicit method. For example, replace $\dot{P}$ and $\dot{T}$ in (10.3) by the time derivative of the entropy $s$ and use this equation only in order to evaluate $\dot{s}^n$ at time $t^n$. Then, as in the case of the chemical composition, the solution for $t^{n+1}$ is calculated for a given, fixed entropy $s^{n+1} = s^n + \Delta t \dot{s}^n$ from the other equations.

It is well known that, for differential equations that involve first-order derivatives in time and first- (or higher-) order spatial derivatives, implicit methods allow larger time steps for a given spacing in mass; for explicit difference schemes the time step has to be kept small to avoid numerical instability (For details see, for instance, Richtmyer and Morton 1967.).

Let us now turn to the so-called*hydrodynamical problem,* which arises when the inertial term in the equation of motion cannot be neglected. Then in addition to the first-order time derivatives in (10.3) there is a second-order time derivative in (10.2), as in (2.16). One usually introduces the radial velocity

$$v = \frac{\partial r}{\partial t} \tag{12.16}$$

of the mass elements as a new variable, with which (10.2) becomes

$$\frac{\partial P}{\partial m} = -\frac{Gm}{4\pi r^4} - \frac{1}{4\pi r^2}\frac{\partial v}{\partial t} \,. \tag{12.17}$$

When using (12.16) and (12.17) instead of (2.16) one has again to deal with first-order time derivatives only. These can be replaced by ratios of differences, and one can use an explicit or an implicit scheme as before, the explicit being simpler but demanding smaller time steps. However, this is not the only choice to be made. For example, within the framework of an explicit method, the different variables can be defined at different times (say, the radius values at $t^n, t^{n+1}, \ldots$, and the velocities at the intermediate times $t^{n-1/2}, t^{n+1/2}, \ldots$). Furthermore, one may devise a scheme which treats the mechanical equations explicitly but is implicit with respect to the time derivatives in the energy equation (10.3).

The presence of the second-order time derivatives changes the properties of the equations and the behaviour of the numerical procedure considerably. Whenever an explicit scheme is used, the time steps have to be kept small in order to fulfil the Courant condition, according to which the time step $\Delta t$ must not exceed $\Delta r/v_s$, where $\Delta r$ is the thickness of the smallest mass shell and $v_s$ is the local velocity of sound.

## 12.4 Treatment of the Diffusion Equation

The diffusion equation (8.25) contains first-order derivatives in time and second-order derivatives in space for the $N$ chemical species. It may be supplemented by the nuclear reactions of (10.5), and by the additional term for diffusive convective mixing (8.28) to achieve a consistent treatment of "burning and mixing", but these terms do not change the nature of the equations further.

The left-hand side of (8.25) can again be written as $(X_i(t + \Delta t) - X_i(t))/\Delta t$, and $X_i(t + \Delta t)$ is the quantity to be determined. As with the nuclear reactions (10.5) discussed in 12.3, an implicit scheme is to be preferred for sake of numerical stability, implying that on the right-hand side $X_i(t + \Delta t)$ is used, too. This constitutes at each grid point a set of $N$ implicit equations, which can be solved either through linearization or iteration. However, in contrast to the situation we found for the nuclear network, these sets of equations are now coupled between grid points due to the spatial derivatives of the diffusion equation.

These second-order spatial derivatives of, for example, $\ln T$, are calculated in two steps. First, the first-order derivative for grid point $j$ is approximated in the standard way by

$$\frac{\Delta \ln T^j}{\Delta r^j} = \frac{\ln T^j - \ln T^{j-1}}{r^j - r^{j-1}} \tag{12.18}$$

and similarly for $j + 1$. Then the second-order derivative at grid point $j$ can be calculated from

$$\left.\frac{\partial^2 \ln T}{\partial r^2}\right|_j \approx \left(\frac{\Delta \ln T^{j+1}}{\Delta r^{j+1}} - \frac{\Delta \ln T^j}{\Delta r^j}\right) / \left(\bar{r}^{j+1} - \bar{r}^j\right), \tag{12.19}$$

where $\bar{r}^j$ is a suitable mean value for $r$ in the interval $(j, j + 1)$. In the simplest case it is the arithmetic mean and thus the denominator in (12.19) reduces to $(r^{j+1} - r^{j-1})/2$. All other quantities in (8.25) appearing in front of the first-order derivatives, such as $A_T(i)$, also have to be taken as mean quantities for the second derivative in analogy to (12.19). We note that the spatial derivatives are defined here at each grid point, contrary to the system of equations (12.3), where the derivatives were defined for the shell between $j$ and $j + 1$. One may imagine that the shells now are centred at a grid point, extending halfway to the neighbouring ones. The advantage of this definition is that the diffusion equations are defined at the same location as the nuclear network equations.

In this way, the discretized equations for the $N$ elements at the $M - 2$ grid points from $j = 2, \ldots, M - 1$ contain values of the $X_i$ at three grid points $j - 1$, $j$, $j + 1$. As for the structure equations, they are solved by iterating for $X_i(t + \Delta t)$, starting with the initial trial values $X_i(t)$, which are already known. The iteration method can again be the standard Newton–Raphson method, which requires first-order derivatives of all quantities appearing in (8.25). The complete

system of equations is similar to (12.13) with the exception that three instead of two neighbouring grid points are connected. It is therefore obvious that the Henyey method will be applicable again, the only difference being that the block matrices are now of dimension $N \times 3N$.

The missing two equations to complete the system for the $N$ elements result from the boundary conditions at $j = 1$ and $j = M$, which follow from mass conservation. We follow here the formulation by Schlattl (1999), where also more technical details concerning the solution of (8.25) can be found.

Mass conservation leads to

$$\sum_{j=1}^{M} \left( X_i^j(t + \Delta t) - X_i^j(t) \right) \triangle m^j = 0, \qquad 1 \le i \le N \tag{12.20}$$

where $j$ again denotes the grid point ($1 \le j \le M$) and $i$ the element. Since (8.25) is formulated in Eulerian space, the mass intervals $\triangle m^j$ have to be defined appropriately, for example, by

$$\triangle m^j = \begin{cases} \frac{1}{2}\left(m^1 - m^2\right) & j = 1 \\ \bar{m}^j - \bar{m}^{j+1} & 2 \le j \le M-1 \\ \frac{1}{2} m^{M-1} & j = M. \end{cases} \tag{12.21}$$

Note that mean values for $m^j$ are used in the second line. This way $M$ mass intervals are created.

As an example we formulate the expression for the $\ln T$ term in (8.25), abbreviating $r^2 X_i(t + \Delta t) T^{5/2} A_T$ by $K_T$. With $\triangle r^j = \triangle m^j / (4\pi \varrho^j (r^j)^2)$ the boundary conditions translate into expressions like

$$\frac{1}{\varrho r^2} \frac{\partial}{\partial r} \left( K_T \frac{\partial \ln T}{\partial r} \right)_{r=R} \approx -\frac{2}{\varrho^{j=1} R^2} \frac{\bar{K}_t^{j=2} (\Delta \ln T / \Delta r)^{j=2}}{r^{j=1} - r^{j=2}} \tag{12.22}$$

and

$$\frac{1}{\varrho r^2} \frac{\partial}{\partial r} \left( K_T \frac{\partial \ln T}{\partial r} \right)_{r=0} \approx \frac{24}{\varrho^{j=M} (r^{j=M-1})^2} \frac{\bar{K}_T^{j=M} (\Delta \ln T / \Delta r)^{j=M}}{r^{j=M-1}} \tag{12.23}$$

To simplify reading we have written suffixes indicating grid numbers $j$ explicitly. In (12.23), the linear expansion (11.3) of $m$ at the centre was used to compute $\triangle r^{j=M}$ from $\triangle m^{j=M}$, which involves $m^{M-1}$.

We finally add that Schlattl (1999) justifies the Eulerian formulation for the diffusion equations, as opposed to our otherwise preferred Lagrangian one, with the necessity for very dense spatial resolution in situations of shallow convective layers.

## 12.5 Treatment of Mass Loss

The mass loss formulae (9.1)–(9.4) describe only how the stellar mass reduces with time due to stellar winds. Therefore, the treatment in stellar evolution calculations is very simple. Over a time step $\Delta t$, during which the chemical composition changes as described in Sect. 12.3, the stellar mass will change according to

$$M(t + \Delta t) = M(t) - \Delta M(t) = M(t) - \dot{M}(t)\Delta t\,, \tag{12.24}$$

where $\dot{M}(t)$ is the mass loss rate evaluated according to (9.1) or any other similar prescription, using the stellar parameters at time $t$.

In terms of the mass grid established in Sect. 12.2 a simple removal of all grid points $i$ with $m_i \geq M(t) - \Delta M(t)$ can be done. Such a procedure, of course, ignores all effects of accelerating matter and moving it out of the star's gravitational potential. To treat this correctly, however, a hydrodynamical method with an open outer boundary would be needed, which in most cases is not necessary. Consider the energy spent to remove mass from the stellar surface to infinity. This is, according to (1.13), $GM/R$ per mass unit of the stellar wind. Multiplying with the mass loss rate we obtain the result that $(\dot{M}GM)/R$ erg/s are needed. For the Sun this amounts to $1.2 \times 10^{27}$ erg/s, which is only $10^{-7}$ of the solar luminosity, and can therefore be safely ignored. For a very evolved red giant with very strong mass loss the energy spent for expelling mass can reach values up to 0.001 or even 1 % of the stellar luminosity.

While the simple removal of grid points is correct in terms of mass distribution and chemical composition, it is not taking into account thermal effects. Imaging a mass layer that was deep inside the stellar envelope now suddenly being the outermost one, since all overlying layers were expelled. It will be hotter than the surface layers have been before and temperature and pressure will not be that of a photosphere. Thermal relaxation will therefore set in. While the Sun is losing mass continuously, its surface temperature is constant. This is because the timescale for mass loss, $\tau_{\rm ML} \approx M/\dot{M}$ is of the order of $10^{14}$ years and therefore much longer than even the nuclear timescale. As long as $\tau_{\rm ML} \gg \tau_{\rm KH}$ the outermost layers will quickly expand and restore the previous photospheric conditions. The adjustment, of course, vanishes with increasing depth. Numerical schemes are therefore trying to take this into account: while grid points are removed due to mass loss, the thermal structure of the star remains almost unperturbed. In the opposite case, when $\tau_{\rm ML} \lesssim \tau_{\rm KH}$, the layers uncovered by mass loss indeed have no time to change their temperature (pressure can be adjusted, since $\tau_{\rm hydro}$ is still much shorter). This, however, may happen only in binary systems during extreme mass transfer episodes. In such cases, a hydrodynamical treatment of the complete system is indicated, anyhow.

## 12.6 Existence and Uniqueness

As every numerical scheme, the Henyey method sometimes does not converge easily to a solution, and there are cases when it seems to oscillate between two solutions. While in most cases this is a purely numerical issue, resulting for example from insufficiently accurate derivatives, one wonders whether there could also be deeper mathematical reasons. This relates to questions about the existence and uniqueness of the solution. It is closely connected to the determinant of the Henyey matrix, as $\det \mathrm{H} = 0$ obviously does not allow an inversion for determining the $\delta y_i$ and $\det \mathrm{H} \approx 0$ will lead to numerical problems during the inversion.

An old problem is whether, for stars in complete equilibrium and of given "parameters" (stellar mass $M$ and chemical composition $X_i$), there exists one, and only one, solution of the basic equations of stellar structure. From simple considerations concerning uncomplicated cases, answers to this question were given in the 1920s by Heinrich Vogt und Henry Norris Russell; however, there is no mathematical basis for this so-called Vogt–Russell theorem, and when by numerical experiments multiple solutions for the same parameters were found to exist it had to be abandoned. The conditions under which uniqueness is violated, and why, have therefore been investigated. A linearized treatment (concerning "local" uniqueness) is easier to understand, whereas non-linear results refer to the "global" behaviour of the solutions and require a more involved mathematical apparatus. Relevant work concerning these issues was done by Kähler (1972, 1975, 1978). For another representation, particularly of the linear problem, see Paczyński (1972).

The mathematical discussion is usually restricted to models in complete or at least hydrostatic equilibrium and analyses the behaviour of solutions under (infinitesimally) small changes of the parameters. Mathematical conditions can be formulated when a solution is locally unique, which can be translated into the statement that the evolution–considered as being a change of parameters (chemical composition and/or entropy) with time–follows a unique sequence of solutions. However, there is no general statement about when such conditions are fulfilled. The condition for having a locally unique solution is equivalent to $\det \mathrm{H} \neq 0$. But even if this condition is fulfilled, there still might be multiple, well-separated solutions. If one of them is unstable, the star switches to the stable one when perturbed. This is related to the general stability of stars.

Behind the mathematical question there is thus also interest concerning the predicted evolution of stars. For example, after learning that often more than one solution exists, that solutions can disappear, or that new solutions appear in pairs, one might begin to wonder whether the star really "knows" how to evolve. But we should keep in mind that normally the star will be brought into one particular state (corresponding to a certain solution) according to its history. And if the equations indicate that the evolution approaches a "critical point", then this means in general only that the approximation used breaks down. For example, if an evolutionary sequence calculated for complete equilibrium comes to a critical point beyond which continuation is not possible, then the difficulties are normally removed by

allowing for thermal non-equilibrium. Correspondingly if hydrostatic models that are not in thermal equilibrium evolve to a critical point, the difficulties are usually removed after the introduction of inertia terms. An example would be the reaction of a star when reaching the *Schönberg–Chandrasekhar limit* (Sect. 30.5), where two existing solutions of complete equilibrium merge. The star easily switches from one to the other by leaving thermal equilibrium.

Broadly speaking, it was found that indeed several solutions for the same set of parameters (stellar mass $M$ and chemical composition $X_i$) exist but that they are widely separated and a star's evolution proceeds along a well-defined sequence of locally unique solutions.