Chapter 3

Problem Solving Methodologies: Numerical Methods

Most numerical methods for solving Fokker-Planck equations fall into two broads classes: Monte Carlo simulations (Valleau and Whittington, 1976) and finite difference methods (the basic concept of the finite difference will be explained in more detail in Appendix C). Actually, the finite difference schemes are computationally very efficient and produce accurate solutions. They are, however, more difficult to implement because they have various stability constraints (Park and Petrosian, 1996 and references therein). Although Monte Carlo simulations are relatively easy to implement, the computational cost can be very expensive, because a very large sample size may needed in order to obtain an acceptably low statistical error. To obtain good accuracy without needing to use a very high performance computer system, the appropriate method for us to use is the finite

difference method.

We numerically simulate the transport of energetic charged particles (e.g., cosmic rays) in space, momentum, and pitch angle using a Fokker-Planck equation that includes various transport processes (Equation (2.12)) and the effect of a variable fluid speed u to first order in (u/c) (Ruffolo, 1995).

In this chapter, we will describe the essential numerical methods and the basic ideas relevant to the finite difference schemes we use in our simulations. Most of all, we would like to introduce the concept used to simplify a complex multi-dimensional partial differential equation to be solvable. After that, we present descriptive details about a generalized scheme that we had developed for the hyperbolic (convection) problems solving and the shock treatment will be described finally at the end of this chapter.

3.1 The Operator Splitting Technique

As we have seen in the previous chapter, the kind of a multi-dimensional transport equation we are concerned with is so complicated that one is unable to solve it analytically (for general boundary conditions). An appropriate numerical method is the only possible way to solve this. This section introduces a powerful technique to handle a complex multi-dimensional linear partial differential equation.

Our simulations deal with solving Equation (2.12) by means of a finite difference method. We exploited the concept of the "operator splitting" tech-

nique, or "fractional steps" as it is also called (Park and Petrosian, 1996; and the references therein: Richtmyer and Morton, 1967, Chap. 8; Press *et al.* 1992, Chap. 19). Operator splitting is not a distinct method for solving the Fokker-Planck equation itself but it is only a way of reducing a larger problem into a series of smaller solvable ones. It solves a partial differential equation with a J *differential* operator ($\mathcal{L}_j$)

$$\frac{\partial u}{\partial t} = \mathcal{L}_1 u + \mathcal{L}_2 u + ... + \mathcal{L}_{J-1} u + \mathcal{L}_J u, \tag{3.1}$$

using a sequence of J finite *difference* operators ($\mathcal{L}_j$) to get

$$\frac{\partial u}{\partial t} = L_J L_{J-1} ... L_2 L_1 u. \tag{3.2}$$

Each finite difference operator L_j solves the differential equation $\partial u/\partial t = \mathcal{L}_j u$ by advancing the solution Δt in time from u^n to u^{n+1} using $u^{n+1} = \mathcal{L}_j u^n$.

We rely on the operator splitting technique to implement a finite difference method over our rectangular simulation domain. The numerical method is a substantially modified version of that of Ruffolo (1995). That is, in a small enough time step, we group the right hand side of the transport equation into 3 groups, involving derivatives with respect to r, μ, and p, and then update $F(t, \mu, r, p)$ for each part consecutively.

In practice, the sequence of steps we used is as follows:

1. Update F for μ-changing processes over a time $\Delta t/2$.

2. Update F for p-changing processes (deceleration) over a time Δt.

3. Update F for r-changing processes (streaming and convection) over a time Δt. Crossing of a shock is also treated in this step.

4. Update F for μ-changing processes over another $\Delta t/2$.

Note that μ-changing processes are treated twice for $\Delta t/2$ each at the beginning and end. The reason why their treatment is split into two parts is because their symmetric treatment in time improves the convergence of the method to second order in Δt. (Every second term disappears in the Taylor series for the error, which is computed with respect to $t + \Delta t/2$.) Steps 2 and 3 do not need to be split because these operations commute to a reasonable approximation.

Actually, we adopted the advantages of the numerical solving pattern and some useful routines used by Ruffolo (1991, 1995). Hence, in steps 1 and 4, we update the distribution function with half the effect of the pitch angle scattering and adiabatic focusing term in Equation (2.12), so that the distribution function would evolve according to

$$\begin{aligned}\frac{\partial F}{\partial t} &= -\frac{\partial}{\partial \mu}\left[\frac{v}{2L}\left(1-\frac{\mu v u\cos\psi}{c^2}\right)+\frac{\mu u}{r}\left(1-\frac{3}{2}\sin^2\psi\right)\right]\cdot(1-\mu^2)F \\ &\quad +\frac{\partial}{\partial \mu}\frac{\varphi}{2}\frac{\partial}{\partial \mu}\left(1-\frac{\mu v u\cos\psi}{c^2}\right)F. \end{aligned} \tag{3.3}$$

This is repeated in step 4 to account for the second half. The scattering and focusing processes generate a μ-flux, S_μ, between each pair of neighboring grid points, μ and $\mu + \Delta\mu$. First, we have to specify boundary condtions at the edges

of the (t, μ, r, p) domain. As $\mu \to \pm 1$, we require the μ flux,

$$S_\mu(t,\mu,r,p) = \left[\frac{v}{2L}\left(1-\frac{\mu v u\cos\psi}{c^2}\right)+\frac{\mu u}{r}\left(1-\frac{3}{2}\sin^2\psi\right)\cdot(1-\mu^2)\right]F(t,\mu,r,p) - \frac{\varphi(\mu)}{2}\frac{\partial}{\partial\mu}\left(1-\frac{\mu v u\cos\psi}{c^2}\right)F(t,\mu,r,p), \tag{3.4}$$

to vanish so that no particles "flow" to the nonphysical region where $|\mu| > 1$. Using Equation (3.4) and the finite-difference approximation for derivatives, we obtain

$$S_{\bar\mu} \approx \left[\frac{v}{2L}\left(1-\frac{\bar\mu v u\cos\psi}{c^2}\right)+\frac{\bar\mu u}{r}\left(1-\frac{3}{2}\sin^2\psi\right)(1-\bar\mu^2)\right]\left[\frac{F(\mu+\Delta\mu)+F(\mu)}{2}\right] - \frac{\varphi_{\text{eff}}(\bar\mu)}{2}\left[\frac{F'(\mu+\Delta\mu)-F'(\mu)}{\Delta\mu}\right], \tag{3.5}$$

where $\bar\mu \equiv \mu + \Delta\mu/2$ and $F' = (E'/E)F = [1 - (\mu v u)/c^2]F$. The effective scattering coefficient, $\varphi_{\text{eff}}(\mu)$, used by Ng and Wong (1979) with a finite difference method to solve the transport equation numerically, can be written as (Ruffolo 1991, 1995)

$$\varphi_{\text{eff}}(\mu) = \frac{v}{2L(z)}(1-\mu^2)\frac{\Delta\mu}{\tanh\{v[I(\mu+\Delta\mu/2)-I(\mu-\Delta\mu/2)]/[2AL(z)]\}}, \tag{3.6}$$

where

$$I(\mu) \equiv \text{sgn}(\mu)\frac{|\mu|^{2-q}}{2-q}. \tag{3.7}$$

The μ-flux can be calculated using F at the start of the step (explicitly), at the end of the step (implicitly), or alternatively explicitly and implicitly for better stabillity and accuracy (Crank and Nicolson, 1947); this well-known method is called the "Crank-Nicolson" method. Using the last approach, the code first

solves the coupled explicit equations,

$$F(\mu) \leftarrow F(\mu) - \left(\frac{\Delta t}{4n}\right)\left[\frac{S_\mu(\mu+\Delta\mu/2) - S_\mu(\mu-\Delta\mu/2)}{\Delta\mu}\right], \tag{3.8}$$

for each grid point μ, then solves the implicit equations,

$$F(\mu) + \left(\frac{\Delta t}{4n}\right)\left[\frac{S_\mu(\mu+\Delta\mu/2) - S_\mu(\mu-\Delta\mu/2)}{\Delta\mu}\right] \leftarrow F(t,\mu), \tag{3.9}$$

and repeats the processes for a total of n explicit and n implicit substeps. Each substep represents a time increment of $\Delta t/(4n)$, so these $2n$ steps account for scattering and focusing over half of Δt. The number n is repeatedly doubled until the resulting flux changes by less than the preset absolute tolerance or the preset relative tolerance.

In step 2, we update the distribution function with the effect of adiabatic deceleration, which is the systematic decrease in p due to the divergence of the fluid flow. The splitting equation concerning the deceleration process alone can be written as

$$\frac{\partial F(t,\mu,r,p)}{\partial t} = -\frac{\partial}{\partial p} pu\left(\frac{1-3\mu^2}{2}\frac{\sin^2\psi}{r} - \frac{1-\mu^2}{r}\right) F(t,\mu,r,p). \tag{3.10}$$

Again, in this step, we adopted the solving methodology from Ruffolo (1995). By this way, the above equation can be written as

$$\frac{\partial}{\partial t}F(t,\mu,r,p) = \frac{1}{\tau_d}\frac{\partial}{\partial p}pF(t,\mu,r,p), \tag{3.11}$$

where the deceleration time, τ_d, is given by

$$\frac{1}{\tau_d} = u\left(\frac{1-3\mu^2}{2}\frac{\sin^2\psi}{r} - \frac{1-\mu^2}{r}\right). \tag{3.12}$$

This in turn gives us

$$\frac{\partial}{\partial t}pF(t,\mu,r,p) = \frac{1}{\tau_d}\frac{\partial}{\partial \ln(p/p_0)}pF(t,\mu,r,p), \tag{3.13}$$

for a fixed reference momentum p_0, which has the solution

$$pF(t+\Delta t,\mu,r,p) = p\, e^{\Delta t/\tau_d} F(t,\mu,z,p\, e^{\Delta t/\tau_d}). \tag{3.14}$$

Then the solution can be solved directly (for more details see Ruffolo (1995)).

In step 3 we implement TVD differencing (Sweby, 1984; explained in detail in the next section), tested and modified for a general Courant number (Nutaro, Riyavong and Ruffolo, 2000), and treat particles crossing the shock, allowing us to properly deal with a gradually varying $u\cos\psi$ or a discontinuous u. Other steps remain the same as in Ruffolo (1995). Away from a shock, step 3 for updating F for r-changing processes involves solving

$$\frac{\partial}{\partial t}F\left(t,\mu,r,p\right) = -\frac{\partial}{\partial r}\left(\mu v\cos\psi + u - \frac{\mu^2 v^2 u\cos^2\psi}{c^2}\right)F\left(t,\mu,r,p\right). \tag{3.15}$$

where $v_r = (\mu v\cos\psi + u - \mu^2 v^2 u\cos^2\psi/c^2)$ represents the spatial velocity of the particle guiding center as viewed in the fixed frame.

Note that in practice, the independent variable r will be replaced by z in our program. In the next section we will use z instead of r as a spatial coordinate to reduce any confusion about notation.

3.2 The Implementation of a Generalized TVD Algorithm

The term "total variation diminishing (TVD) method" was first presented by Harten (1983) for an oscillation-free scheme. The total variation (TV) of a function $U(x)$ is defined by

$$TV = \int \left| \frac{\partial U}{\partial x} \right| dx, \tag{3.16}$$

so the TV of the numerical solution is defined accordingly by

$$TV(U^n) = \sum \left| U^n_{i+1} - U^n_i \right|, \tag{3.17}$$

where i and n are spatial and time indices respectively, and the principal requirement of such a method is

$$TV(U^{n+1}) \leq TV(U^n). \tag{3.18}$$

Thus such a method is called total variation diminishing (TVD). The motivation for this requirement is to avoid creating new maxima or minima of the function in each time step. A unified description of several independently proposed, second-order accurate TVD schemes was provided by Sweby (1984). The TVD schemes employ flux limiters, and are related to the flux corrected transport (FCT) technique of Boris and Book (Boris and Book, 1973; Boris *et al.* 1973; Zelesak, 1979), although differing in the respect of being essentially one-step procedures as opposed to the two-step FCT. The purpose of flux limiting/correcting is to produce

a high resolution scheme without the spurious oscillations associated with the more classical second order schemes.

The TVD scheme that we employ is Roe's superbee limiter (Roe, 1983), which gives remarkably sharp profiles for the linear advection equation. Of the various schemes, this one provides the most aggressive antidiffusion that is consistent with the conditions of Sweby (1984). For some applications, this scheme can yield profiles that are too sharp (Jeng and Payne, 1995; Arora and Roe, 1997) but for our application the aggressive anti-diffusion is advantageous because the key goal is to avoid numerical diffusion. Of various the schemes reviewed by Sweby (1984), this scheme gave the most accurate solution for sharp pulses. In any case the choice of a TVD scheme, i.e., the choice of the flux limiter, is not critical, as evidenced by Hatzky's successful application (Hatzky 1996) of van Leer's method (van Leer, 1974).

Note that in early work with $u = 0$ and $v_z = \mu v$, i.e., neglecting processes such as convection and deceleration, Equation (3.15) could be solved exactly by simply moving the distribution function from one z-grid point to another (Ruffolo, 1991):

$$F(t + \Delta t, i\Delta\mu, z, p) = F(t, i\Delta\mu, z - i\Delta z, p), \tag{3.19}$$

where i labels the μ-grid points, $\mu = i\Delta\mu$, and $\Delta z = \Delta\mu\, v\Delta t$. In this case the Courant number $\gamma = v_z\, \Delta t/\Delta z$ was the integer i, and with no finite-difference approximation, there was no numerical diffusion. For a non-zero (but still small) u, this approach was modified (Ruffolo, 1995) by occasionally moving the distribu-

tion forward by an additional step, Δz. Thus numerical diffusion was still absent, and a sharp pulse or gradient in the distribution function could be maintained intact. While this method also gave the proper convection speed on average, it led to a somewhat irregular distribution function and still required the very small grid spacing $\Delta z = \Delta\mu v\Delta t$, leading to a slow run time.

The TVD method offers the possibility of a smoother and more accurate solution, and avoiding the assumption $v \gg u$ used in previous work. However, standard implementations of TVD imply a limitation on the Courant number, $\gamma = v_z\Delta t/\Delta z$, so that $0 \leq \gamma \leq 1$. For this application, we want to be able to set $\gamma > 1$ in order to use a smaller Δz, either for greater flexibility, for a comparison with previous results for $\Delta z = \Delta\mu v\Delta t$, or for improved numerical accuracy. In fact, our results will show that while Δz can be increased to values much higher than those used in Ref. 18, leading to a faster running speed, actually setting $\Delta z > v_z\Delta t\,(\gamma < 1)$ yields an unacceptable error for some cases of interest. Therefore, we have developed a generalization of the TVD routine to allow a general value of γ, which can also vary with position.

Let us first consider the case where v_z is independent of z. In our generalized TVD technique, we first move F by an integral number of steps, g, which is obtained by rounding γ downward. For example, if $\gamma = 3.4$, F is moved forward by 3 z-grid points. Then the remainder $\gamma' = \gamma - g$ (0.4 in this case) is between 0

and 1 and is accounted for by the usual TVD differencing:

$$F_\ell \leftarrow F_{\ell-g} - \frac{\Delta t}{\Delta z} S'_{\ell+1/2} + \frac{\Delta t}{\Delta z} S'_{\ell-1/2} \tag{3.20}$$

where $S'_{\ell+1/2}$ denotes the flux from z-cell ℓ to $\ell+1$ due to γ' and is calculated from

$$S'_{\ell+1/2} = v'_{\ell+1/2} F_{\ell-g} + \frac{1}{2} v'_{\ell+1/2} (1 - \gamma'_{l+1/2}) \left(F_{\ell-g+1} - F_{\ell-g}\right) \varphi_{\ell-g}, \tag{3.21}$$

where $\gamma'_{\ell+1/2}$ and $v'_{\ell+1/2}$ are γ' and $v'_z \equiv \gamma' \, \Delta z / \Delta t$, respectively, at $z = z_\ell + \Delta z/2$, and φ_ℓ is Roe's superbee limiter (Roe, 1983) given by

$$\varphi_\ell = \begin{cases} 0 & r_\ell \leq 0 \\ 2r_\ell & 0 < r_\ell \leq 0.5 \\ 1 & 0.5 < r_\ell \leq 1 \\ r_\ell & 1 < r_\ell \leq 2 \\ 2 & r_\ell > 2 \end{cases} \tag{3.22}$$

for $r_\ell \equiv (F_\ell - F_{\ell-1})/(F_{\ell+1} - F_\ell)$. Note that $\varphi_\ell \equiv 1$ corresponds to the Lax-Wendroff technique (Lax and Wendroff, 1960), which converges to second order in Δz; in choosing φ_ℓ that is sometimes different from unity, the TVD methods sacrifice convergence speed (then converging to first order in Δz) for the guarantee (Equation (3.18)) that no new minima or maxima are introduced by the second (anti-diffusive) term on the right hand side of Equation (3.20).

Next, we should consider variations in v_z and hence γ as a function of z. In the generalized method described above, variations within the range $0 \leq \gamma' < 1$

are acceptable as in the standard implementation of TVD, but what if γ crosses an integral value, i.e., g (the greatest integer $< \gamma$) changes from one point to the next? To take this possibility into account, F over a time step Δt is updated by

$$F_\ell \leftarrow \left(\sum_{m=\ell-g_-}^{\ell-g_+} F_m \right) - \frac{\Delta t}{\Delta z} \left(S'_{\ell+1/2} - S'_{\ell-1/2} \right), \tag{3.23}$$

where g_+ (g_-) is the rounded-down integer corresponding to $\gamma_{\ell+1/2}$ ($\gamma_{\ell-1/2}$). This formula is subject to the constraint that $g_+ \leq g_- + 1$; if $g_+ = g_- + 1$ we interpret the sum to be zero. In practice that constraint can often be avoided by reducing Δt, and in physical situations where $v_z(z)$ is discontinuous, one should use a special treatment such as that described shortly.

It is straightforward to demonstrate the consistency of this generalized TVD method, i.e., that in the limits $\Delta t \to 0$ and $\Delta z \to 0$ one recovers Equation (3.15). A detailed discussion of the consistency of standard TVD methods is given by Sweby (1984). Consistency of the present method can also be demonstrated by means of Taylor series expansions, assuming that $f'' \equiv \partial^2 f / \partial z^2$ and dv_z/dz exist in the region of interest. Note that $r_\ell = 1 - [f''(z_\ell)/f'(z_\ell)]\ \Delta z + O(\Delta z^2)$, so in the limit of a small step size the flux limiter φ_ℓ is given by 1 or r_ℓ. This limit also implies that $g_+ = g_-$ or $g_- \pm 1$. An important step in the Taylor series analysis is to recognize that $g\Delta z/\Delta t = v_z - v'_z$.

In all cases, Equation (3.15) is reproduced for the point of interest, (t_n, z_ℓ), with error $O(\Delta t) + O(\Delta z)$ or $O(\Delta t) + O(\Delta z^2)$. As mentioned earlier, for the case where $g_+ = g_- = g$, the convergence is only to second order in Δz when $\varphi \equiv 1$,

i.e., for both $\varphi_{\ell-g}$ (used in $S'_{\ell+1/2}$, according to Equation (3.21)) and $\varphi_{\ell-g-1}$ (used in $S'_{\ell-1/2}$). Otherwise the convergence is to first order in Δz. When $g_+ = g_- + 1$, the convergence is also to first order. However, when $g_+ = g_- - 1$, the order of convergence again depends on whether φ is 1 or r. If $\varphi_{\ell-g_-+1}$ (used in $S'_{\ell+1/2}$) and $\varphi_{\ell-g_--1}$ (used in $S'_{\ell-1/2}$) are both 1 or both the corresponding value of r, then the convergence is to first order; otherwise the convergence is to second order because of the symmetry of the resulting formula for F_ℓ.

Our tests have shown that there is no noticeable change in performance for γ' values near 0 or 1, so the generalized TVD method does give us freedom to use any continuous function $v_z(z)$ and any Δz. Next, we consider the case of a discontinuity in $v_z(z)$, which we treat by means of special matching conditions at the discontinuity. Physically, this situation often corresponds to a shock discontinuity in a fluid flow. Here, we discuss our implementation of this TVD method for the problem of energetic charged particle transport across an oblique magneto-hydrodynamic shock. As described in section 1, this situation is important for modeling the acceleration of charged particles (cosmic rays) throughout the universe, which in most cases is believed to occur at a shock, as well as the effect of the shock on existing cosmic rays.

Considering the solution of the transport Equation (2.12), in the framework of operator splitting, step 3 corresponds to spatial motions (streaming + convection), so the treatment of particles crossing the shock is naturally included in this step (see next section). We treat the transport of particles from a given cell

by considering whether the particles encounter the shock during a time increment Δt. If not, then we use the generalized TVD scheme as above.

3.3 Spherical Shock Treatment

The transport problem concerning a shock is special because of the drastic change in the fluid speed, magnetic field, *etc.*, in a very thin layer between the upstream and downstream sides, which we treat as a discontinuity in our model. The physical parameters on either side are not the same. The TVD scheme is no longer appropriate for the particles encountering the shock. We need some special assumptions or mechanisms to describe particles crossing the shock. Here, we exploited the basic idea following Ruffolo (1999).

The older literature has discussed two mechanisms, the first-order Fermi acceleration (Fermi 1954; Parker 1958) (see more details in Appendix A) and the shock drift mechanism (Schatzman, 1963), in which particles drift along an oblique shock front due to the sharp gradient in the magnetic field, and this drift is along the direction of the electric field so that particles can gain a substantial amount of energy in one encounter with the shock.

More recently, it has been shown that the distinction between these two mechanisms vanishes under a transformation to an appropriate frame of reference which is called the de Hoffmann-Teller (shock) frame (de Hoffmann and Teller, 1950). The electric field is zero in this frame. The entire energy change due to

both mechanisms is accounted for by transforming the particle momentum from the local fluid frame into the shock frame, considering the energy-conserving shock encounter in the shock frame, and then transforming the momentum into the new local fluid frame (Decker, 1983; Riyavong, 1996; Ruffolo, 1999).

In practice, the spatial simulation length is divided into two sides downstream ($r < r_{sh}$, where r_{sh} is the shock radius) and upstream ($r > r_{sh}$), where the shock is stationary and forms a radial boundary between these two regions. Actually, we will suppose that the shock boundary is thin; it should be much less than the gyroradius of a charged particle. If particles encouter the shock, we first perform a Lorentz transformation of p and μ into the shock frame. In general, for a static magnetic field, $\vec{F} = q\vec{v} \times \vec{B}$ is perpendicular to $\vec{v}$, so the rate of doing work on the particle, $\vec{F} \cdot \vec{v}$, is zero; thus the momentum in the shock frame is conserved throughout the encounter. Ruffolo (1999) used the common approximation that the magnetic moment $p^2(1 - \mu^2)/(2meB)$ is conserved as particles cross or are reflected by shock (Decker, 1983). Here we instead exploit a hybrid orbit-finite difference treatment of oblique shock acclereation (Sanguansak and Ruffolo, 1999), which numerically calculates energetic particle orbits near an oblique shock for a grid of momentum-space coordinates, without using the assumption of magnetic moment conservation. In this way, the transport equation on either side of the shock, which incorporates streaming, convection and pitch angle scattering and also includes adiabatic focusing and deceleration, is solved using a well-tested finite difference code. Finally, we perform a Lorentz

tranformation of p and μ back into the local wind frame.

Note that the TVD algorithm effectively splits a cell into fractions of particles destined to move to two different spatial locations. Here apply a similar method, since when particles cross the shock, some particles might be transported to quite different μ values as well. Note that for the nonrelativistic fluid speeds and energetic cosmic ray particle speed considered here, the fractional change in momentum for an individual shock encounter is not large. In other work, this numerical method has been applied for a grid of p points. Here we have treated one p value and simply assumed $F \propto p^{-\gamma}$, through there are problems with this assumption when u/v is not small (Ruffolo, 1999).