Chapter 3

The formal transfer equation

The words "radiative transfer" make it sound as if we are mainly interested in studying the movement of photons. In reality, the interaction of the radiation with the medium is actually the main issue. As we discussed in Section 2.4, in the absense of any interaction with matter, the transport of radiation is fairly trivial: the intensity in any direction then remains constant along a ray in that direction. However, interaction with the medium can remove radiation from the ray or add to it. In most cases we can assume that light propagates so fast that we can ignore the light travel time effects. In other words: in most cases we can assume that all photons travel through the medium on a time scale much shorter than any changes that happen to the medium. We can thus regard the radiation as a steady-state flow of photons. We will, however, discuss the limits of validity of this approach in Section 3.6.

3.1 Extinction coefficient, opacity

The interaction of radiation with matter can be typically understood in terms of two processes: radiation being injected into the ray and radiation being removed from the ray. Let us start discussing the latter.

Suppose that there are particles in the medium that can absorb a photon. This absorption process is random: a photon can be lucky, and travel quite far, but it can also be unlucky and be absorbed quickly. The efficiency of the medium to absorb a photon is given by the *photon mean free path* l_{free}, which has the CGS dimension of cm. Typically this is a function of frequency ν and position in space $\mathbf{x}$, i.e. $l_{\nu,\mathrm{free}}(\mathbf{x})$. The denser the medium is, the smaller is the mean free path. Suppose that we have a cloud in which there is no emission, only extinction. If we have a light source on one side of the cloud, and we observe this light source from the other side of the cloud, then the cloud may extinct part of the radiation from the source that passes through the cloud before it reaches us. If the path length of the light travelling through the cloud is equal to one mean free path, then on average only 36.8% of the photons manage to get through the cloud (since $e^{-1} = 0.368$). If the path length is twice the mean free path, only 13.5% survives the journey through the cloud. This is what is called *extinction*. The number of mean free path lengths the photons travel through the cloud is called *optical depth* and is typically written with the symbol τ. A medium that has $\tau \gg 1$ is called *optically thick* while a medium that has $\tau \ll 1$ is called *optically thin*.

Rather than using mean free paths, in radiative transfer theory we more commonly use its reciprocal: the *extinction coefficient* $\alpha_\nu = 1/l_{\nu,\mathrm{free}}$, with CGS unit of cm^{-1}. This quantity is also often called the *opacity*. With this coefficient the optical depth between two points along a ray can be expressed as the following integral:

$$\tau_\nu(s_0, s_1) = \int_{s_0}^{s_1} \alpha_\nu(s)\,ds \tag{3.1}$$

where $\alpha_\nu(s)$ is the extinction coefficient at point s along the ray (using $\mathbf{x} = \mathbf{x}_0 + s\mathbf{n}$, see Eq. 2.24).

Often the density-dependence of the opacity is explicitly written as:

$$\alpha_\nu = \rho\kappa_\nu \tag{3.2}$$

where ρ is the density, with CGS units of gram cm^{-3}, and κ_ν is the mass-weighted opacity, with CGS units of cm^2 gram^{-1}. Often κ_ν is also simply called the opacity. When one talks about "opacity" one should therefore be careful what is actually meant. Note that sometimes people use the word "opacity" when they actually mean "optical depth". We shall stick to the strict separation of these two terms.

If we regard the medium as a collection of particles (for instance: dust particles), then we can introduce yet another way to write the opacity: the cross section per particle σ_ν, with CGS units cm^2. If the particles are very large compared to the wavelength, this cross section is typically equal to the geometric cross section, which for spherical particles of radius a is equal to $\sigma = \pi a^2$. This is, however, only valid for particles with $a \gg \lambda$. For small particles and large wavelength ($a \ll \lambda$) one typically finds that $\sigma \ll \pi a^2$. We will discuss this at length in Chapter 6. For now we limit ourselved by stating a relation between κ_ν and σ_ν:

$$\kappa_\nu = \frac{\sigma_\nu}{m} \tag{3.3}$$

where m is the mass of the particles.

3.2 The formal radiative transfer equation

Let us now introduce the concept of extinction into the differential equation for the intensity along a ray, Eq. (2.25). Instead of a zero right-hand-side we now have

$$\frac{dI_\nu(\mathbf{n}, s)}{ds} = -\alpha_\nu(s) I_\nu(\mathbf{n}, s) \tag{3.4}$$

This is the *formal radiative transfer equation* for the case of a purely absorbing (and non-emitting) medium. Note that this equation is an equation along a given ray. It is valid along any ray passing through the medium. We can integrate Eq. (3.4) to obtain the integral form of this equation:

$$I_\nu(\mathbf{n}, s_1) = I_\nu(\mathbf{n}, s_0) e^{-\tau_\nu(s_0, s_1)} \tag{3.5}$$

where $\tau_\nu(s_0, s_1)$ with $s_1 > s_0$ is given by Eq. (3.1). This equation expresses what we already qualitatively argued in Section 3.1.

Now let us assume that the cloud also injects radiation into the ray. We then add an emission term to Eq. (3.4):

$$\frac{dI_\nu(\mathbf{n}, s)}{ds} = j_\nu(s) - \alpha_\nu(s) I_\nu(\mathbf{n}, s) \tag{3.6}$$

This is the complete *formal radiative transfer equation*. The source term j_ν is called the *emissivity* and has CGS dimensions of erg s^{-1} cm^{-3} Hz^{-1} ster^{-1}. Also this can be cast in integral form:

$$I_\nu(\mathbf{n}, s_1) = I_\nu(\mathbf{n}, s_0) e^{-\tau_\nu(s_0, s_1)} + \int_{s_0}^{s_1} j_\nu(s) e^{-\tau_\nu(s, s_1)} ds \tag{3.7}$$

The formal radiative transfer equation, Eq. (3.6), can also be written in a form similar to Eq. (2.23):

$$\mathbf{n}\cdot\nabla I_\nu(\mathbf{x},\mathbf{n}) = j_\nu(\mathbf{x}) - \alpha_\nu(\mathbf{x}) I_\nu(\mathbf{x},\mathbf{n}) \tag{3.8}$$

This form is mathematically equivalent to Eq. (3.6) and it can be useful, for instance, for deriving the radiative diffusion equation (see Section 4.5).

3.3 Kirchhoff's law

To satisfy the laws of thermodynamics, the formal radiative transfer equation, Eq. (3.6), must obey a certain condition. Suppose we have a thermal cavity filled with gas with some extinction coefficient α_ν. Suppose that this gas is in thermal equilibrium with the temperature T in the cavity. Then, irrespective of whether the gas is optically thick or thin, the intensity should everywhere be equal to $I_\nu = B_\nu(T)$. For the formal radiative transfer equation this means that $dI_\nu/ds = 0$ and thus

$$j_\nu(s) - \alpha_\nu(s) I_\nu(\mathbf{n},s) = j_\nu(s) - \alpha_\nu(s) B_\nu(T) = 0 \tag{3.9}$$

In other words:

$$\frac{j_\nu}{\alpha_\nu} = B_\nu(T) \tag{3.10}$$

This is *Kirchhoff's law*. It says that a medium in thermal equilibrium can have any emissivity j_ν and extinction α_ν, as long as their ratio is the Planck function.

This law does not only apply in a thermal cavity. It applies everywhere where the medium is in *local thermodynamic equilibrium* (LTE). While LTE is not always guaranteed (and we shall see plenty of examples where LTE breaks down in this lecture), in media where it is valid, Kirchhoff's law greatly simplifies the radiative transfer problem: In LTE we can use Kirchhoff's law to write the formal radiative transfer equation in the form

$$\frac{dI_\nu(\mathbf{n},s)}{ds} = \alpha_\nu(s)[B_\nu(T(s)) - I_\nu(\mathbf{n},s)] \tag{3.11}$$

Note that here the Planck function is allowed to vary along the ray. This form of the equation clearly demonstrates that the intensity I_ν is always trying to asymptotically approach $B_\nu(T(s))$. If the temperature is constant along the ray, then the intensity will indeed exponentially approach $B_\nu(T)$. If the temperature varies along the ray, the intensity will always lag behind by a few mean free paths, but it will always tend to approach the Planck function.

3.4 The "source function"

Inspired by Kirchhoff's law we can try to express the transfer equation in a way similar to Eq. (3.11) irrespective of whether we have LTE or not. Let us define the *source function* S_ν as

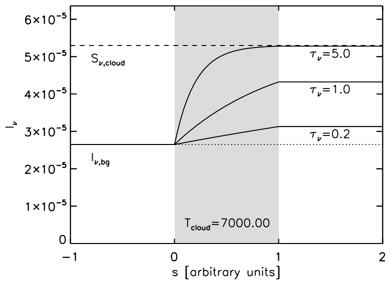
$$S_\nu \equiv \frac{j_\nu}{\alpha_\nu} \tag{3.12}$$

The formal radiative transfer equation then becomes

$$\frac{dI_\nu(\mathbf{n},s)}{ds} = \alpha_\nu(s)[S_\nu(s) - I_\nu(\mathbf{n},s)] \tag{3.13}$$

For the case of LTE the source function is equal to the Planck function: $S_\nu = B_\nu(T)$, and we retrieve Eq. (3.11). For a non-LTE case the source function can be unequal to the Planck function. In this lecture we will encounter radiative transfer problems in which we will in fact need to solve for the source function as part of the solution to the transfer equation.

In LTE or non-LTE alike, the source function is a useful quantity because it acts, like the Planck function for the LTE case, as an "attractor" for the intensity: At every point along the ray the intensity wants to approach S_ν as it proceeds its journey along the ray. If S_ν is constant along the ray, then within a few mean free path lengths the intensity will have exponentially approached $I_\nu \to S_\nu$. If S_ν varies with s, then I_ν will lag behind, but always tries to approach S_ν along the way.



How this works for a ray passing through a slab of given temperature in front of a radiation source of another temperature is illustrated in the margin figure. It shows the intensity I_ν at $\nu = c/\lambda$ with $\lambda = 0.5\,\mu$m starting at a background intensity corresponding to a temperature $T_{\mathrm{bg}} = 6000\,$K going through a slab or cloud of gas with a temperature $T_{\mathrm{cloud}} = 7000\,$K with three values of the total optical depth, as annotated in the figure. The slab is assumed to be in LTE, so that the source function equals the Planck function.

3.5 Spectroscopic absorption- and emission features

With the simple formal radiative transfer framework we have developed so far we can already study - and understand - how spectroscopic emission features and absorption features are formed. A "spectral feature" here means any change in the spectrum that is limited to a small wavelength domain and which can be associated to some physical property of the medium. Gas spectral lines are "features" in that sense (see Chapter 7 for an in-depth discussion on gas spectral lines). So when we talk about gas lines, we would talk about emission lines and absorption lines. However, also solids such as dust grains can have spectral features. They will, however, be much wider in wavelength than gas lines. Yet, also these features can be both in emission and in absorption. Therefore we will refer here to "features" as a more general class of spectral signatures than just "lines", and we will discuss here how such features are formed.

3.5.1 Optically thin case

In the optically thin case we can ignore the extinction part of the radiative transfer equation and the formal transfer equation becomes

$$\frac{dI_\nu(s)}{ds} = j_\nu(s) \tag{3.14}$$

where, for notational convenience, the **n** vector is omitted. Integrating this yields

$$I_\nu(s_1) = I_\nu(s_0) + \int_{s_0}^{s_1} j_\nu(s)ds \tag{3.15}$$

What we observe ($I_\nu(s_1)$) is equal to the background intensity ($I_\nu(s_0)$) plus the emission from the cloud between s_0 and s_1.

If the frequency-dependent function j_ν has a feature around some frequency ν_0, i.e. if the function j_ν has a particularly large value near ν_0 but is much smaller (or even zero) at frequencies far away from ν_0, then this is, via the integral Eq. (3.15), also visible as a feature in the function $I_\nu(s_1)$. The value of $I_\nu(s_1)$ will be particularly high for frequencies close to ν_0. If the background intensity is even much higher than that, then the feature could be "drowned" by the background intensity. In many cases of optically thin clouds in astrophysics, however, the background is dark, so the spectral feature will be strongly apparent and will have the same shape as the feature in the emissivity function j_ν. Since the feature points upward (i.e. it is the brightest around ν_0) we call this an *emission feature*. Or in other words: the feature *is in emission*.

3.5.2 Optically thick case

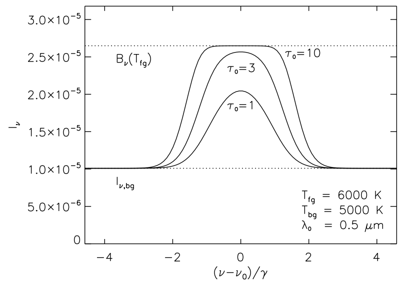
A medium that is optically thick can, in addition to emission features also create *absorption features*. Both are possible, and which of the two is created depends strongly on the temperature gradient.

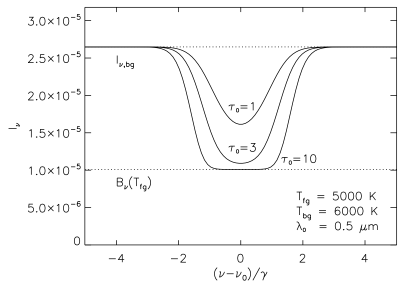
Let us assume that we have a medium that is in LTE, so that Kirchhoff's law is valid. Let us now assume that the medium consists of an optically thick background of temperature $T_{\rm bg}$ and a foreground layer of gas in front of it (as seen by the observer) of tempeature $T_{\rm fg}$. The "feature" is a bump in the opacity in the gas layer α_ν around frequency ν_0. Together with the thickness ΔX of the foreground layer this opacity bump yields an optical depth $\tau_\nu = \alpha_\nu \Delta X$ which has the following functional form:

$$\tau_\nu = \tau_0 \exp\left(-\frac{(\nu-\nu_0)^2}{\gamma^2}\right) \tag{3.16}$$

where γ denotes the width of the feature. Let us assume that the emission from the optically thick background is a perfect blackbody $I_{\nu,\rm bg} = B_\nu(T_{\rm bg})$. The question is: how will the opacity "feature" of Eq. (3.16) appear as a spectral feature in the observed intensity I_ν? We can find out by integrating the formal transfer equation in the form of Eq. (3.11). We obtain

$$I_{\nu,\rm observed} = I_{\nu,\rm bg} e^{-\tau_\nu} + (1 - e^{-\tau_\nu})\, B_\nu(T_{\rm fg}) \tag{3.17}$$





In the figures in the margin the results are shown for a feature at $\nu_0 = 6\times 10^{14}$ Hz (corresponding to a wavelength of $\lambda_0 = 0.5\,\mu$m), for $T_{\rm bg} = 5000\,$K and $T_{\rm fg} = 6000\,$K, which yields an emission feature, and for the opposite ($T_{\rm bg} = 6000\,$K and $T_{\rm fg} = 5000\,$K), which yields an absorption feature. The results are shown for three different values of τ_0.

From these figures we learn a number of things. The most important one is that if a hot layer is in front of a cool layer, we get emission features, and if a cool layer is in front of a hot layer, we get absorption features. The famous Heidelberger scientists Kirchhoff and Bunsen in fact discovered this (and published it in 1860), and were thus able to explain the absorption features of the solar spectrum.

Another thing we learn from the figures is that the emission feature has the same shape as the opacity feature as long as $\tau_0 \lesssim 1$. But when $\tau_0 \gg 1$, the feature becomes optically thick and *saturates*. This is exactly the "attractor effect" mentioned above: the intensity wants to approach the Planck function of the foreground layer. Once it has arrived at that Planck function, it will not change any further.

Another important thing we can learn is: If we have an optically thick cloud or atmosphere with a constant temperature (which here would translate to: the layer temperature being equal to the background temperature), then we would not observe any features in the spectrum - neither in absorption nor in emission.

It is important to understand that this very same principle of feature formation (for gas spectral lines: line formation) can be applied to cases of non-LTE. We should then just replace the Planck function $B_\nu(T_{\rm fg})$ with the source function $S_{\nu,\rm fg}$. The rest stays the same. For such a case we *can* in fact form a feature even if the temperature is constant, as long as $S_{\nu,\rm fg} \neq I_{\nu,\rm bg}$.

3.5.3 Eddington-Barbier estimation

In most real application we do not have a clean two-temperature situation as sketched above. We will have a smooth temperature gradient. The solution to the formal transfer equation can still be easily calculated, simply by performing a numerical integration (see Section 3.8).

However, there is a simple trick to get a reasonably good estimate of the observed intensity, called the *Eddington-Barbier estimation*. If the medium is optically thick but it has a temperature gradient, the intensity you observe is roughly equal to the source function at the location where the optical depth is $\tau_\nu = 2/3$:

$$I_\nu^{\rm observed} \simeq S_\nu(\tau_\nu = 2/3) \tag{3.18}$$

For media in LTE this means: you observe a blackbody intensity of temperature T at the location where the optical depth toward you is 2/3:

$$I_\nu^{\text{observed}} \simeq B_\nu(T(\tau_\nu = 2/3)) \tag{3.19}$$

With the Eddington-Barbier estimation we have another, and quite powerful, way to understand how spectral lines and features are formed. Consider the solar photosphere. Deep down into the photosphere the temperature is higher than at the top of the photosphere. In other words: there is a negative temperature gradient: $dT/dz < 0$. If we look at the atmosphere at a frequency ν that is right at the center of a spectral line, where the opacity α_ν of the photosphere is very high, then the location z where $\tau_\nu = 2/3$ is somewhere in the top of the photosphere, where temperatures are comparatively low. If, however, we shift ν far from the spectral line, the opacity α_ν of the photosphere drops, meaning that the location z where $\tau_\nu = 2/3$ is now much deeper, where the temperatures are higher. This predicts that the spectrum of the Sun should have its lines typically in absorption, which is indeed the case.

The Eddington-Barbier estimation is, however, not always valid. You can see this, again, by an example of the Sun's atmosphere. Above the photosphere there is the chromosphere, which is much hotter than the photosphere, but also much more tenuous. The optical depth of the chromosphere is small, yet in some strong spectral lines it may still dominate the photospheric emission. Eddington-Barbier would not predict this to happen. It shows that Eddington-Barbier can be used if the temperature gradient is moderate, but not in cases where there is an extremely hot tenuous layer in front of a much cooler optically thick medium.

3.6 A note on time-dependence

So far we have assumed that the speed of the propagation of radiation is so large, that photons pass through our object of interest in a time much shorter than that the object can change its properties. For the vast majority of radiative transfer problems in astrophysics this is a good approximation. There are, however, some occasions where the light travel time plays a role. Consider, for instance, a star surrounded by a large disk or envelope. If the star exhibits a sudden outburst of luminosity (for instance, an accretion event or an instability), and the onset of this outburst takes only minutes to hours, then, compared to that minute or hour time scale, the outer regions of the circumstellar disk or envelope will receive that light much later. Another example is light echos of a supernova: a supernova lasts only weeks to months, while the outgoing light may excite molecules or dust grains many parsecs away for tens to hundreds of years afterward. Clearly in these problems the steady-state formal transfer equation, Eq. (3.6) is not valid. We can, however, easily extend it to account for the light travel time effects:

$$\frac{1}{c}\frac{\partial I_\nu(\mathbf{n},s,t)}{\partial t} + \frac{\partial I_\nu(\mathbf{n},s,t)}{\partial s} = j_\nu(s) - \alpha_\nu(s) I_\nu(\mathbf{n},s,t) \tag{3.20}$$

Solving such problems numerically is, however, not entirely trivial, even if we know j_ν and α_ν perfectly in advance. Eq. (3.20) is a partial differential equation of hyperbolic type: an advection equation with source and sink terms. Problems of this mathematical kind are routinely encountered in physics, in particular in the area of hydrodynamics, but they are also known to be numerically tricky. In the present lecture, however, we will focus on the more common problems in which the time-derivative in Eq. (3.20) can be ignored.

There is, however, an entirely other kind of time-dependence in radiative transfer theory that is much more common: matter does not react instantly to changes in the radiation field. It takes time to radiatively heat up or cool down a parcel of gas. And as we shall see later, in very optically thick media the transport of radiative energy proceeds through many absorption and re-emission events, each involving a certain time-delay. The slowness of radiative energy transport is then not due to the finite speed of light, but due to the latency introduced by the slow re-emission process. We will discuss such effects at length in the chapter on radiation hydrodynamics (Chapter 12).

3.7 1-D Plane-parallel radiative transfer problems

Many radiative transfer problems in astrophysics are truly 3-dimensional in nature. As we have seen above, this essentially means that the mathematical problem is 6-dimensional, since we have to also account for 2 directions and 1 frequency. For notational convenience let us write that our problem is 3-D (with capital D) when we talk about the spatial dimensions, and 6-d (with small-letter d) when we talk about the full number of independent variables. Only in the first decade of this millennium have computers become powerful enough to tackle 6-d problems with sufficient resolution.

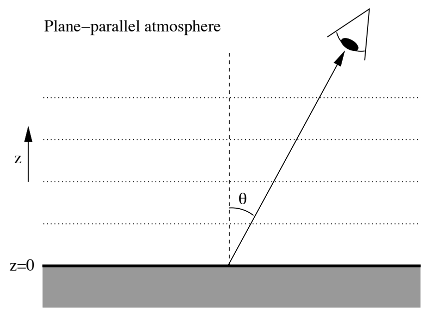
Fortunately, often one can identify symmetries in the problem that allow one to reduce the dimensionality from 6-d down to 5-d or even down to just 3-d.

Consider the simplest possible model of a planetary or stellar atmosphere: a *plane parallel* atmosphere. In this model the variables of the gas, such as gas density, gas temperature etc, depend only on z, but not on x and y. We thus have perfect translational symmetry in x and y. The coordinate z is the vertical coordinate. For the Earth's atmosphere this could mean that $z = 0$ is the surface while $z > 0$ is the atmosphere. We also assume rotational symmetry in the $x - y$-plane. We have thus reduced the problem from 3-D to 1-D. The total dimensionality has been reduced from 6-d to 3-d: In addition to z, we still have the angular coordinate $\mu = \cos\theta$ and the frequency ν. The angle ϕ drops out, because of the rotational symmetry in the $x - y$-plane.

It is important to understand that while a plane-parallel atmosphere is formally a 1-D problem, this does not mean that the photons can travel only either upward or downward! This is a very common misunderstanding. The problem remains, in some way, fully 3-D: photons can still move in all three directions, and there exists an atmosphere not only at $x = 0, y = 0$, but also at $x = 5$ and $y = -8$. The only thing that makes it 1-D is that we do not need to explicitly care about the dependency of variables on x and y. The differential operators $\partial/\partial x$ and $\partial/\partial y$ will yield 0 for such problems. And while photons can still move in any direction ϕ, we do not have to keep track of this. Only the dependence on μ matters. The differential operator $\partial/\partial\phi$ always yields 0. Solving the 1-D plane-parallel transfer problem therefore means that we actually solve the full 3-D problem; it is just that the problem has plane-parallel symmetry.

The formal transfer equation in the form of Eq. (2.23) can, for such a plane-parallel geometry, be written in the following form:

$$\mu\frac{dI_\nu(z,\mu)}{dz} = j_\nu(z) - \alpha_\nu(z)I_\nu(z,\mu) \tag{3.21}$$



(with $\mu = \cos\theta$) or equivalently with the source function S_ν:

$$\mu\frac{dI_\nu(z,\mu)}{dz} = \alpha_\nu(z)[S_\nu(z) - I_\nu(z,\mu)] \tag{3.22}$$

In comparison to the "along the ray" form of the transfer equation, Eq. (3.6), the d/ds was replaced by $\mu d/dz$. For fixed μ Eq. (3.21) can be integrated over z, which is equivalent to integrating Eq. (3.6) along a ray for fixed $\mathbf{n}$.

The first three moments of radiation are, in plane-parallel geometry:

$$J_\nu = \frac{1}{2}\int_{-1}^{+1} I_\nu(\mu)\, d\mu \tag{3.23}$$

$$H_\nu = \frac{1}{2}\int_{-1}^{+1} I_\nu(\mu)\,\mu\, d\mu \tag{3.24}$$

$$K_\nu = \frac{1}{2}\int_{-1}^{+1} I_\nu(\mu)\,\mu^2 d\mu \tag{3.25}$$

All these are scalars, because in 1-D we are only interested in the z components of the tensors.

Throughout this lecture we will regularly deal with plane-parallel transfer problems, and we will demonstrate many radiative transfer effects using Eq. (3.21). It is the simplest form of the transfer equation that is still general enough to demonstrate many aspects of radiative transfer theory. In Section 5.2.8 we will discuss another 1-D radiative transfer geometry: that of spherically symmetric radiative transfer problems. However, those problems involve a few tricky elements that we will try to avoid for most of the lecture - hence our focus on 1-D plane-parallel geometries.

3.8 A numerical algorithm for integrating the formal transfer equation

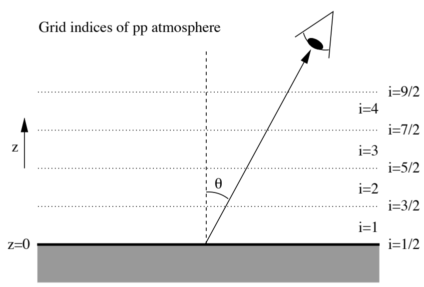
In most cases of practical interest, to get an accurate solution to the formal transfer equation requires numerical integration. Doing this in a reliable and stable way is not entirely trivial. A naive approach can lead to vastly wrong results. In this section we will discuss the first, second and third order versions of the method of Olson & Kunasz (1987, J. Quant. Spectros. Radiat. Transfer 38, 325), which we will henceforth call the *OK87 method*. This method, with some minor additions, has turned out to be extremely stable and reliable for all purpuses that I have encountered. There are also other reliable methods in the literature, such as the famous Feautrier method. But the OK87 method is more generally applicable, and we will therefore take the OK87 method as our workhorse method throughout this lecture.

Warning: *To understand what follows, a basic knowledge of numerical methods is required. I assume that you know how to numerically integrate simple differential equations on a grid using e.g. forward Euler or Runge-Kutta integration. If you are completely new to numerical methods, please read some of the relevant chapters of "Numerical Recipes" by Press, Teukolsky, Vetterling and Flannery*[1].

3.8.1 Putting the formal transfer problem on a grid

So the objective is, to find a reliable and accurate numerical algorithm to integrate the equation[2]

$$\frac{dI}{ds} = j - \alpha I \equiv \alpha(S - J) \tag{3.26}$$



We omitted all $(\mathbf{x})$, $(\mathbf{n})$ and ν for notational convenience.

Let us, for the moment, confine ourselves to a 1-D plane-parallel case, so that we have only one coordinate: the vertical dimension z. Let us divide z up into cells. The cells have indices $i = 1, 2, 3, \cdots, N_z$ where N_z is the number of grid cells. The cell walls, which separate the cells, also have indices, which we will give half-numbers: $i = \frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \cdots, N_z + \frac{1}{2}$. This is shown in the figure.

A computer cannot handle half-numbers as indices to array elements. So if we want to store variables that are located at the cell walls, we must use integers again. By

[1] http://www.nr.com/

[2] Numerical integration is also often called *quadrature*.

convention we will then use index 1 for cell wall $i = \frac{1}{2}$ and we will have an array of $N_z + 1$ elements, because we have $N_z + 1$ cell walls. Note that several programming languages start their array indices with 0 instead of 1. That would mean that the indices shift by one. We would then have cell indices $0, \cdots, N_z - 1$ and cell wall indices $0, \cdots, N_z$. How you index the cells and cell walls in your computer program is, in the end, a matter of taste and is left up to you to decide.

The grid is now defined by the z-locations of the cell walls: $z_{1/2}, z_{3/2}, \cdots, z_{N_z+1/2}$. In the figure the bottom cell wall is located at $z_{1/2} = 0$.

The cell walls are, in 1-D, grid *points*. We will see that there are *cell-based radiative transfer algorithms* and *grid-point-based radiative transfer algorithms*, and that the two classes of methods work a bit differently. To distinguish between cell-based and point-based variables we will use the integer and half-integer indexing for cell-based and point-based (or wall-based) variables.

If we have a ray passing through our grid, then the ray crosses the cell walls. These cell-wall-crossings divide the ray into ray segments. We will apply the same method of indexing these segments: the segments have integer indices i while the joining-points between the segments have half-integer indices $i + 1/2$. If s is our coordinate along the ray, then the joining-points have $s_{i+1/2}$.

In our 1-D setting, if we integrate upward ($\mu > 0$), we can match the indices along the ray with the indices of z. If we integrate downward, then increasing s means decreasing z. If we decide to still use a matched indexing of the ray and the grid[3], then we will be integrating from high to low indexing. In that case, in all quadrature formulae shown below we would have to swap: $i + 3/2 \leftrightarrow i - 1/2$ and $i \leftrightarrow i + 1$.

3.8.2 First order integration

The key of the OK87 quadrature formulae is to make an assumption for the functional form of $j(z)$ and $\alpha(z)$ between the cell boundaries, and then to analytically solve the formal transfer equation exactly. The first order version of this method assumes that the emissivity j and extinction coefficient α are constant within each cell, but can be different from one cell to the next.

For each cell i we can thus calculate an optical depth:

$$\Delta\tau_i = (s_{i+1/2} - s_{i-1/2})\alpha_i \tag{3.27}$$

For our 1-D example, assuming that we integrate upward (i.e. $\mu > 0$), we can equivalently write:

$$\Delta\tau_i = (z_{i+1/2} - z_{i-1/2})\frac{\alpha_i}{\mu} \tag{3.28}$$

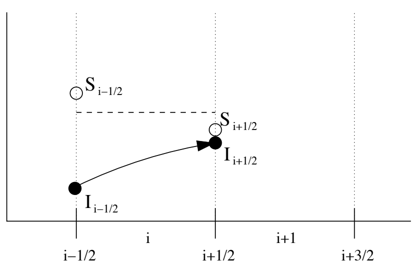
We can now calculate the source function on each ray segment, which in 1-D means the source function in each cell:

$$S_i = \frac{j_i}{\alpha_i} \tag{3.29}$$

which is constant throughout each cell. Now we can write the exact integral to the formal transfer equation from the bottom to the top of cell i as

$$I_{i+1/2} = e^{-\Delta\tau_i} I_{i-1/2} + \left(1 - e^{-\Delta\tau_i}\right) S_i \tag{3.30}$$

First order integration



Assuming that j and α are indeed constant within the cell, Eq. (3.30) is an exact result! It is therefore valid for any grid cell size, i.e. for any value of $\Delta\tau_i$.

We can now use Eq. (3.30) to integrate systematically from one grid wall to the next. For $\mu > 0$ we do this using Eq. (3.30) starting at $i = 1$ (i.e. starting from $I_{1/2}$) and

[3] In 1-D this can make sense. In 3-D, however, a ray should always have its own indexing, which is then increasing with increasing s.

working our way up to $i = N_z$ (i.e. arriving at $I_{N_z+1/2}$). Here $I_{1/2}$ is the boundary condition, which we discuss in Subsection 3.8.5 below. For $\mu < 0$ we start from the top and work our way down. The quadrature formula has to be accordingly adapted.

3.8.3 Second order integration

For many purposes the first order integration scheme of Subsection 3.8.2 is sufficiently accurate. "Sufficiently" here means that it does not produce results that are dramatically wrong. It may, however, not be very accurate either. To get a result that is within some tolerance margin, it is known that results obtained from first order integration schemes require a much finer gridding (and thus many more gridpoints) than when higher-order integration schemes are used. It is therefore worthwhile to consider higher order schemes.

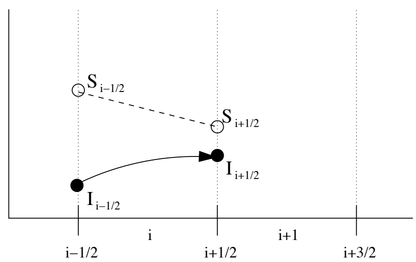
Moreover, as we will see in Chapter 4, higher order integration schemes may be crucial when we use them in iteration methods for solving multiple scattering or non-LTE radiative transfer problems.

The philosophy of the second order integration scheme presented by Olson and Kunasz (1987, J. Quant. Spectros. Radiat. Transfer 38, 325) is very similar to that of Subsection 3.8.2. Also in their scheme an exact analytical solution is computed for the formal transfer equation along a segment of the ray. The difference is now that we define the S and α on the cell walls ("grid points"), and assume that they vary linearly between the cell walls (instead of constant within the cell). For cell i this implies:

$$S(z) = \frac{z_{i+1/2} - z}{z_{i+1/2} - z_{i-1/2}} S_{i-1/2} + \frac{z - z_{i-1/2}}{z_{i+1/2} - z_{i-1/2}} S_{i+1/2} \tag{3.31}$$

$$\alpha(z) = \frac{z_{i+1/2} - z}{z_{i+1/2} - z_{i-1/2}} \alpha_{i-1/2} + \frac{z - z_{i-1/2}}{z_{i+1/2} - z_{i-1/2}} \alpha_{i+1/2} \tag{3.32}$$

Second order integration



Without proof the exact solution of the integral of the formal transfer equation across cell i is such that, for $\mu > 0$,

$$I_{i+1/2} = e^{-\Delta\tau_i} I_{i-1/2} + Q_i \tag{3.33}$$

with

$$Q_i = \left(\frac{1 - (1 + \Delta\tau_i)e^{-\Delta\tau_i}}{\Delta\tau_i}\right) S_{i-1/2} + \left(\frac{\Delta\tau_i - 1 + e^{-\Delta\tau_i}}{\Delta\tau_i}\right) S_{i+1/2} \tag{3.34}$$

There are two caveats with this quadrature formula. First, if $\Delta\tau_i \ll 10^{-6}$, the finite machine precision may cause problems. In that limit one can write

$$\lim_{\Delta\tau_i \to 0} Q_i = \frac{1}{2}\Delta\tau_i (S_{i-1/2} + S_{i+1/2}) \tag{3.35}$$

So if we use this formula, in case $\Delta\tau_i < 10^{-6}$, then this problem is solved. Secondly, under very pathological circumstances this second order quadrature recipe can sometimes yield overshoots. This can happen in cases in which $S_{i+1/2} < S_{i-1/2}$ but $\alpha_{i+1/2} > \alpha_{i-1/2}$ (or vice versa). Inside of cell i the functions $S(z)$ and $\alpha(z)$ are linear interpolations between these values. Their product $j(z) = \alpha(z)S(z)$ is therefore a parabola which, for the case described above, can have a local maximum somewhere inside the cell. If $\Delta\tau_i \lesssim 1$, then such a parabolic functional form of $j(z)$ will give an intensity that is larger than one would expect when one would have linearly interpolated $j(z)$ instead of $S(z)$. Therefore it is important to supplement the above second order integration recipe with the following "quadrature limiter":

$$Q_i = \min\left(Q_i^{\rm 2nd}, Q_i^{\rm max}\right) \tag{3.36}$$

with $Q_i^{\rm 2nd}$ given by Eq. (3.34) and

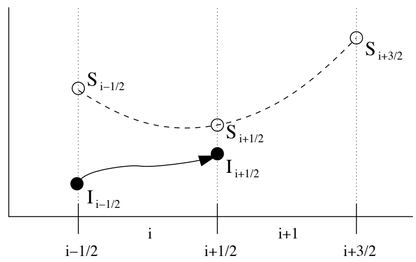
$$Q_i^{\rm max} = \frac{1}{2}(j_{i+1/2} + j_{i-1/2})\Delta s \tag{3.37}$$

This quadrature limiter will only intervene if the gradients of S and α have opposite signs. Otherwise the second order recipe stays in effect.

3.8.4 Third order integration

It may sound like overkill, but as we will discover in Chapter 4, even second order integration can under some circumstances be not accurate enough. If we just want to integrate the formal transfer equation to obtain a spectrum or image, then first and second order are absolutely fine (of course, with second order giving nicer results than first order). But when we use first or second order integration for iteration schemes to solve non-LTE and/or multiple scattering problems, we may be forced to use third order integration. This was recognized by Olson and Kunasz (1987, J. Quant. Spectros. Radiat. Transfer 38, 325), who, in addition to their second order integration scheme also presented a third order one. Here is how it goes[4].

Third order integration



In the third order integration scheme for obtaining $I_{i+1/2}$ for $\mu > 0$ we do not only use $S_{i-1/2}$ and $S_{i+1/2}$, but also $S_{i+3/2}$. The "subgrid model" for $S(z)$ is now a quadratic fit through these three values. Also for this $S(z)$ the formal transfer equation can be analytically solved. The result:

$$I_{i+1/2} = e^{-\Delta\tau_i} I_{i-1/2} + Q_i \tag{3.38}$$

(i.e. the same as Eq. 3.33). But now we define Q_i as

$$Q_i = u\,S_{i-1/2} + v\,S_{i+1/2} + w\,S_{i+3/2} \tag{3.39}$$

with

$$u = e_0 + \frac{e_2 - (2\Delta\tau_i + \Delta\tau_{i+1})e_1}{\Delta\tau_i(\Delta\tau_i + \Delta\tau_{i+1})} \tag{3.40}$$

$$v = \frac{(\Delta\tau_i + \Delta\tau_{i+1})e_1 - e_2}{\Delta\tau_i \Delta\tau_{i+1}} \tag{3.41}$$

$$w = \frac{e_2 - \Delta\tau_i e_1}{\Delta\tau_{i+1}(\Delta\tau_i + \Delta\tau_{i+1})} \tag{3.42}$$

where the symbols e_0, e_1 and e_2 are defined as

$$e_0 = 1 - e^{-\Delta\tau_i} \tag{3.43}$$

$$e_1 = \Delta\tau_i - 1 + e^{-\Delta\tau_i} \equiv \Delta\tau_i - e_0 \tag{3.44}$$

$$e_2 = \Delta\tau_i^2 - 2\Delta\tau_i + 2 - 2e^{-\Delta\tau_i} \equiv \Delta\tau_i^2 - 2e_1 \tag{3.45}$$

With this third order integration we must be even more careful than for second order integration: now not only an overshoot could happen, but also an undershoot: we might even obtain negative results, because the quadratic interpolation of the source function might go negative. We must thus, in addition to the upper limiter set by Eqs. (3.36, 3.37), also introduce a bottom limiter:

$$Q_i = \max\left(\min\left(Q_i^{\mathrm{3rd}}, Q_i^{\max}\right), 0\right) \tag{3.46}$$

with $Q_i^{\max}$ still given by Eq. (3.37).

3.8.5 Boundary conditions

When we integrate from $z = 0$ upward, for $\mu > 0$, we must start with some value of $I_{1/2}$. Which value should we take? The answer depends on the problem we wish to solve. If $z = 0$ represents the surface of the Earth, and if the wavelength we are

[4]Note that in the Kunasz & Olson paper the grid indices are from top to bottom, so for $\mu > 0$ the cell indices get smaller each step. We have our indices from bottom to top, i.e. $z_{i+1/2} > z_{i-1/2}$.

interested in is the mid-infrared, then it is reasonable to take it to be $I_{1/2} = B_\nu(T)$ with T the temperature of the ground. But if we consider the optical wavelength regime, then it depends entirely on the reflection of light impinging on the surface. That is: we do know know $I_{1/2}$ in this case until we calculate the downward radiative transfer. This gives us a glimpse of the true complexity of radiative transfer: To calculate the radiation field, we have to know it in advance. So let us put this issue to rest for the moment. We will discuss it at length in Chapter 4.

So what about the downward integration, for $\mu < 0$? In that case we must impose a boundary condition for $I_{N_z+1/2}$. In the case of our Earth's atmosphere, most of the sky above the atmosphere is pitch black. We can then set $I_{N_z+1/2} = 0$. This is also true if we model a stellar atmosphere.

For the Earth's atmosphere (and any planetary atmosphere) there is, however, one exception: The irradiation of the Earth's atmosphere by the Sun. Typically this occurs under some inclination angle: $\theta > 0$ for some ϕ. While this does not break the plane-parallel translational symmetry, it does break the rotational symmetry in the $x - y$-plane. The irradiation by the Sun thus would force us to go from a 3-d problem (z, μ, ν) to a 4-d problem (z, μ, ϕ, ν). We will discuss this at length in Chapter 9.

3.8.6 Choosing the right spatial resolution

A (wrong) rule of thumb for the sufficiently narrow spacing of the z-grid that is often quoted is to always make sure that $\Delta\tau \lesssim 1$ over each grid cell. As we have seen above, if you do the integration properly, this is not always necessary. In fact, it is often prohibitly numerically expensive to make the grid finely enough spaced that all cells are optically thin. This is even more true in 3-D radiative transfer problems. If we were to strictly adhere to that rule, most 3-D radiative transfer problems would presumably be not feasible. This rule of thumb is also simply wrong: there are regions, at sufficiently high optical depth, where it is not at all necessary to have optically thin grid spacing. Fortunately!

The "real" rules of thumb are:

1. Use a stable numerical integration scheme that also works properly when large steps in τ are taken.
2. For high optical depths use second order integration if possible.
3. Try to spatially resolve the photosphere of the object with sufficient number of grid points, because it is here where the observed spectrum is formed.
4. Regions that are at high optical depth at all wavelengths[5] can be mapped with optically thick grid spacing.
5. One can always do an a-posteriori check if the grid resolution was chosen sufficiently high: the intensity function $I_\nu(s)$ along the ray should not make large jumps from one grid point (or grid cell) to the next.

3.8.7 Numerical integration of rays in 2-D and 3-D

While the 1-D plane-parallel example is sufficient to explain the basic principles of the numerical integration methods, it is worthwhile to briefly discuss how this can be generalized to 2-D and 3-D grids. If we divide our 2-D or 3-D space up in cells, then a ray passing through that grid will intersect with the cell walls, thus dividing the ray up into ray segments, each segment belonging to a cell that is being traversed. Their lengths Δs_i can be quite irregular: a long segment can be followed by a tiny one,

[5] To be more precise: all wavelengths near the peak of the Planck function at the temperatures involved.

followed by an intermediate one etc. But apart from that the integration along the ray remains the same as we have seen so far: we simply use the quadrature formulae we have discussed in this section.

The main new aspect in 3-D compared to what we have done so far in 1-D is that the variables such as α and S are stored either in the cell or on the cell corners. When they are stored in the cell (cell-based radiative transfer), we must use the first order quadrature formula, because α and S are then assumed to be constant throughout the cell. If they are stored at the cell corners (grid-point-based radiative transfer), then we must employ interpolation from the cell corners to the point where the ray crosses the cell wall. The simplest would be bilinear interpolation, because a cell wall has four corner points. A better way would be bi-quadratic or bi-cubic.

Ray made up of ray segments