

Markov Chain Monte Carlo solutions for radiative transfer problems

W. von Waldenfels^{1,2}, R. Wehrse^{3,2,*}, and B. Baschek³

¹ Institut f. Angewandte Mathematik, INF 294, 69120 Heidelberg, Germany

² Interdisziplinäres Zentrum für Wissenschaftliches Rechnen, INF 368, 19120 Heidelberg, Germany

³ ZAH, Institut f. Theoret. Astrophysik, Albert-Ueberle-Str. 2, 69120 Heidelberg, Germany
e-mail: baschek@ita.uni-heidelberg.de

Received 15 January 2010 / Accepted 19 September 2010

ABSTRACT

Algorithms for the solution of the radiative transfer equation that are simultaneously accurate, fast, and easy to implement are still highly desirable, in particular for multidimensional media. We present a stochastic algorithm that solves the transfer equation by assuming that the transfer of radiation can be modelled as a Markov process. It is a generalization of the classical Monte Carlo method and can be applied to the solution of the Milne equation.

Key words. radiative transfer – methods: numerical – methods: statistical – scattering – line: formation

1. Introduction

Monte Carlo methods have been used successfully for radiative transfer problems for many years (see e.g., Juvella 2005; Lucy 1999b). When these algorithms are employed usually the elementary processes, i.e., the flights of a (large) number of photons, are simulated or packages of photons cf. Lucy (1999a,b), of a given energy are tracked. These methods, which we refer to as *classical Monte Carlo methods* are popular since (i) the corresponding code is quite easy to implement; (ii) complicated geometries and various physical effects can be included in a straightforward way; and (iii) the computing time to solve a given problem with a reasonable accuracy is reasonable as long as the optical depths are not too high.

On the other hand, it is well-known that Monte Carlo methods can be used quite generally for the solution of linear ordinary and partial differential equations (see e.g. Zwillinger 1989). To our knowledge, Markov Chains have been proposed in astrophysics only once for radiative transfer problems (Esposito & House 1978) but only in the context of specific intensities. However, in the fields of computer visualisation, Bayesian statistics or global optimization similar approaches (called *Markov Chain Monte Carlo* or abbreviated *MCMC*) are frequently used. In this paper we follow this line of thought and apply it to the solution of the radiative transfer equation. Pseudo particles (or for short) particles are mostly not physical entities but symbolize stochastic variables, which are more intuitive.

For a general introduction to stochastic processes, we refer to e.g. Cox & Miller (1965), Papoulis (1991), or Williams (2001). An excellent introduction to these ideas can be found in Diaconis' article (2009).

To establish a sound mathematical basis, we present in the next two sections a short introduction to transient Markov processes including the corresponding potential equation (Sect. 2) and discuss the probabilistic background of the Monte Carlo method (Sect. 3). The potential of a Markov process is a well

known subject in the theory of Markov processes (see Meyer 1966). The classical potential given by the potential equation $\Delta\Phi = -4\pi\rho$ (cf. Eq. (6)) is the potential of the Brownian motion. In Sect. 4, we show that Beer's law and the radiative transfer equation can be derived by means of simple probabilistic arguments and the corresponding difference equation may be solved very simply by a Markov Chain Monte Carlo method. We also demonstrate how various forms of the transfer equation (differential, difference, or stochastic equation) and the corresponding solutions transform into each other. In Sect. 5, the probabilistic interpretation of the discretized Feautrier equation is derived and discussed. To highlight the wide range of application of the Markov Chain Monte Carlo algorithm, the Milne equation is shown in Sect. 6 to be the potential equation of a Markov process. We note, that Milne's equation is defined in three dimensional space, so it is much easier to treat than the usual radiative transfer equation in five dimensional space (for fixed time and frequency variable). A general scheme for the numerical application is given. In Sect. 7, the numerical properties of MCMC are exemplified by means of several 1D and 3D configurations. We end with a discussion of the advantages and disadvantages of MCMC compared to the classical Monte Carlo approach and grid methods.

2. The transient Markov process and the corresponding potential equation

In accordance with Wang & Uhlenbeck (1945), we consider a pseudo particle that can be in one of $i = 1, \dots, N + 1$ states at discrete times $t = 0, 1, 2, \dots$. A random process performed by this pseudo particle is described completely by the set of probability distributions $w_n((i_1, t_1), (i_2, t_2), \dots, (i_n, t_n))$ where

- $w_1(i, t)$ indicates the probability of finding the pseudo particle at place i at time t ;
- $w_2((i_1, t_1), (i_2, t_2))$ indicates the probability of finding the particle at time t_2 in the place i_2 and at time t_1 at place i_1 ;

* Deceased 08 December 2009.

- $w_3((i_1, t_1), (i_2, t_2), (i_3, t_3))$ gives the probability of finding the particle at time t_3 in the place i_3 , at time t_2 in the place i_2 and at time t_1 at place i_1 ;
- and so on.

The set of distributions must fulfill the following conditions:

- $w_n((i_1, t_1), \dots, (i_n, t_n)) \geq 0$;
- $w_n((i_1, t_1), \dots, (i_n, t_n))$ is symmetric in its arguments;
- $w_k((i_1, t_1) \dots (i_k, t_k)) = \sum_{i_{k+1}, \dots, i_n=1}^{N+1} w_n((i_1, t_1), \dots, (i_n, t_n))$
as the knowledge of the distributions at times t_1, \dots, t_n implies knowledge of the distributions at times t_1, \dots, t_k for $k < n$.

The set of distributions form therefore a kind of hierarchy. They describe successively the random process in more detail.

A *Markov process* is given by an initial probability π and by the matrix of transition probabilities P^1 . The initial probability is a vector $\pi = (\pi_1, \pi_2, \dots, \pi_{N+1})^T$ with $\pi_i \geq 0$ and $\sum \pi_i = 1$. The matrix elements of P are the transition probabilities $P_{i,j}$ that indicates the probability of a transition from state j to state i where $P_{i,j} > 0$ and $\sum P_{i,j} = 1$. A Markov process is now *defined* by

$$w_{n+1}((i_0, 0), \dots, (i_n, n)) = P_{i_n, i_{n-1}} \dots P_{i_2, i_1} P_{i_1, i_0} \pi_{i_0}. \quad (1)$$

Therefore, the probability of moving in one step from state j to state i is $P_{i,j}$ and the corresponding probability of this occurring in n steps is $(P^n)_{i,j}$. This Markov process with discrete states (here $1, \dots, N+1$) and discrete times (here $0, 1, 2, \dots$) is often called a *Markov chain*.

We now assume that the Markov process is *transient*, which means in the simplest case (used in this paper) that there exists a unique state Φ , called an absorbing state or “graveyard”, in which all pseudo particles which have left the system are gathered. Since the pseudo particles do not return, Φ indicates that the process remains in the graveyard after the particle has entered it. Transient means that all particles finally end in Φ with probability 1 regardless of the point in $\{1, \dots, N+1\}$ where they started, i.e. that each particle is finally either absorbed or leaves the medium.

We assume that the graveyard is the state $N+1$. Then P is of the form

$$P = \begin{pmatrix} Q & 0 \\ q & 1 \end{pmatrix}, \quad (2)$$

where Q is an $N \times N$ -matrix representing the probabilities for jumps outside the graveyard and q is the probability of a jump into the graveyard.

Since the process ends in state $(N+1)$ with probability 1, we have $Q^n \rightarrow 0$ for $n \rightarrow \infty$. Hence, $(\mathbf{1} - Q)(\mathbf{1} + Q + Q^2 \dots Q^n) = \mathbf{1} - Q^{n+1} \rightarrow \mathbf{1}$ for $n \rightarrow \infty$ and therefore $\mathbf{1} - Q$ is invertible, where $\mathbf{1}$ denotes the N -dimensional unit matrix.

We define the *potential* of the transient Markov process with the matrix

$$\Pi = \mathbf{1} + Q + Q^2 + \dots = (\mathbf{1} - Q)^{-1}. \quad (3)$$

The element $\Pi_{i,j}$ indicates the mean number of times that a pseudo particle enters state i during its “life time”, when it has started at position j .

The *potential equation* of the Markov process with S as a source is given by

$$\mathcal{I} = Q\mathcal{I} + S. \quad (4)$$

¹ Following the use in mathematics, scalars, vectors, and matrices are generally not distinguished typographically in this paper.

The solution of Eq. (4) is

$$\mathcal{I} = \Pi S. \quad (5)$$

We note that this potential equation is the analogue of the classical potential equation

$$\Delta\Phi_g = -4\pi\rho \quad (6)$$

in which the underlying Markov process is the Brownian motion. The matrix Q corresponds to the Laplace operator and S to $-4\pi\rho$, and Π is the Poisson kernel as can be seen from the discretisation of the Laplace operator. In the Dirichlet problem, i.e., for zero boundary conditions, the particle goes to the graveyard, when it crosses the boundary.

3. Probabilistic background of the Monte Carlo method

We choose an enormous, but *finite* number K and consider the space

$$\Omega_K = \{(i_0, i_1, \dots, i_K) : i_j \in \{1, \dots, N+1\}\}$$

of all sequences of numbers $1, \dots, N+1$ of length $K+1$. On the huge, but finite space, we define a probability \mathbb{P} by

$$\mathbb{P}(i_0, i_1, \dots, i_K) = w_{K+1}((i_0, 0), (i_1, 1), \dots, (i_K, K)).$$

We avoid treating the case $K = \infty$, for that would require some complicated mathematical constructions. A subset $A \subset \Omega_K$ has the probability

$$\mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\omega).$$

A *random variable* is a mapping from Ω_K into the set of real numbers. If Y is a random variable, we define the *expectation* or *mean* or *average value* by

$$\mathbb{E}Y = \sum_{\omega \in \Omega_K} \mathbb{P}(\omega)Y(\omega).$$

We consider the random variable

$$\mathbf{N}_j = \#\{l : i_l = j \text{ for } l = 0, \dots, K\},$$

where \mathbf{N}_j is the number of times the particle is in state j for the times $t = 0, \dots, K$, i.e. practically during its lifetime. It is well known that for a transient Markov process

$$\mathbb{E}\mathbf{N}_j = ((\mathbf{1} - Q)^{-1}\pi)_j = (\Pi\pi)_j.$$

When we return to the potential equation (Eq. (4)) and assume that $S_i \geq 0$, we have to assume in particular that $S_{N+1} = 0$. We then define the initial probability by $\pi_i = S_i / (\sum_l S_l)$, so that

$$I_j = \left(\sum_l S_l \right) \mathbb{E}\mathbf{N}_j.$$

These considerations suggest – as a recipe for the use of the Monte Carlo method – that we perform the following experiment: *We choose the starting point of the pseudo particle according to the initial probability π . We then let the particle jump until it enters the graveyard. We subsequently, count the number of times \hat{N}_j it has been in the state j . The experiment has to be repeated many times and the average over the numbers \hat{N}_j is to be computed. If the number of repetitions goes to infinity, then the averages approach $\mathbb{E}\mathbf{N}_j$, which equals I_j except for a factor.*

4. The radiative transfer equation for an absorbing medium

4.1. Derivation of Beer's law by means of stochastics

The derivation follows Reif (1976) and is analogous to the determination of the nearest neighbor distribution (cf. Hertz 1909; Traving 1960).

We assume that an absorption process can be considered as a collision of a photon with a disk of cross-section σ , the absorption cross-section of an atom. Furthermore, we assume that there are N of these disks in a unit volume and that the disks are randomly distributed (equal probability distribution) in the medium without any correlations.

On the basis of these assumptions, the probability of finding the nearest atom with cross-section σ along a beam at a distance s is given by the probability P_- that there is no atom between 0 and s and the probability P_+ that there is one atom between s and $s + ds$

$$P(s)ds = P_-(s)P_+(ds). \quad (7)$$

The probability that an interaction takes place in the range $s \dots s + ds$ is simply given by

$$P_+(ds) = N\sigma ds \quad (8)$$

and the probability that there is *no* interaction in the interval $s \dots s + ds$ is consequently

$$P_-(ds) = 1 - N\sigma ds. \quad (9)$$

The probability of having no interaction in the range is given – because of the assumed independence of the particle positions – by the product of the probabilities of the partial ranges of which there are s/ds

$$P_-(s) = (1 - N\sigma ds)^{s/ds}, \quad (10)$$

which for $ds \rightarrow 0$ gives

$$P_-(s) = \exp(-N\sigma s). \quad (11)$$

Therefore,

$$P(s)ds = N\sigma ds \exp(-N\sigma s) = \chi \exp(-\chi s)ds, \quad (12)$$

where $N\sigma = \chi$ is the extinction coefficient.

Note that $\int_0^\infty P(s)ds = 1$, i.e. that the probability density is normalized. The mean free path of photons is now given by

$$\langle s \rangle = \int_0^\infty sP(s)ds = \int_0^\infty s\chi \exp(-\chi s)ds = \frac{1}{\chi}. \quad (13)$$

Since $P(s)$ is given by the ratio of the specific intensity at s to that at 0, $I(s)/I(0)$, Beer's law

$$I(s) = I(0) \exp(-\chi s) \quad (14)$$

follows from Eq. (12) and hence also the radiative transfer equation

$$\frac{dI}{ds} = -\chi I. \quad (15)$$

4.2. Markov Chain Monte Carlo solution

As a most simple non-trivial example, we now assume that Eq. (15) is given and we solve it by means of the Markov Chain Monte Carlo method following Zwillingner (1989).

The finite total z -range $0 \dots z_0$ is divided into N subintervals of length $\Delta z = z_0/N$, which are numbered by the index i . Implicit (“up-stream”) discretisation of Eq. (15) results in

$$\frac{I_{i+1} - I_i}{\Delta z} = -\chi I_{i+1}$$

or

$$I_{i+1} = \frac{1}{1 + \chi \Delta z} I_i,$$

where $1/(1 + \chi \Delta z)$ is now interpreted as the probability of a jump from position i to position $i + 1$.

From a theoretical point of view, we now have a state space $\{1, \dots, N, N + 1\}$, where state $N + 1$ is the graveyard or absorbing state. The transition probability matrix is very simple in this case since it is only possible to jump from i either to $i + 1$ or the graveyard. The corresponding probabilities are $\alpha = 1/(1 + \chi \Delta z)$ and $1 - \alpha$, so that e.g. in the case of $N = 5$ the transition probability matrix is given by

$$P = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ \alpha & 0 & 0 & 0 & 0 & 0 \\ 0 & \alpha & 0 & 0 & 0 & 0 \\ 0 & 0 & \alpha & 0 & 0 & 0 \\ 0 & 0 & 0 & \alpha & 0 & 0 \\ 1 - \alpha & 1 - \alpha & 1 - \alpha & 1 - \alpha & 1 & 1 \end{pmatrix}. \quad (16)$$

At the right boundary, we note that the pseudo particle can jump only to the graveyard.

We now work the other way round and show that the transfer equation follows from the stochastic process, i.e. work from the probabilistic interpretation (cf. Zauderer 1983). For small Δz , we write

$$I(z + \Delta z) \approx I(z) + \frac{dI(z)}{dz} \Delta z = \frac{1}{1 + \chi \Delta z} I(z) \approx (1 - \chi \Delta z) I(z)$$

or

$$\frac{dI(z)}{dz} \Delta z = -\chi \Delta z I(z),$$

so that

$$\frac{dI(z)}{dz} = -\chi I(z),$$

i.e. we regain the transport equation when we assume the probabilistic interpretation to be valid.

5. Direct calculation of the mean intensity for a scattering and absorbing medium by means of the Feautrier equation

We now show that by means of a stochastic process one can calculate directly the mean intensity J from the well-known Feautrier equation (see e.g. Mihalas 1978),

$$\frac{d^2 J}{d\tau^2} = J - (1 - \varepsilon)J = \varepsilon J, \quad (17)$$

where τ is the optical depth defined by $d\tau = \chi ds$ and ε the ratio of absorption to extinction. We note that this equation is a special case of the *telegrapher's equation*, which is usually derived

from the Maxwell equations but that can also be derived from a stochastic process as shown by Zauderer (1983).

For simplicity, we restrict ourselves now to the two-stream-approximation and neglect internal sources and spatial variations in ε . Since we can discretize the second derivative by

$$\frac{d^2J}{d\tau^2} \approx \frac{1}{\Delta\tau^2}(J_i + J_{i+2} - 2J_{i+1}), \quad (18)$$

we find that

$$\varepsilon J_{i+1} = \frac{1}{\Delta\tau^2}(J_i + J_{i+2} - 2J_{i+1}) \quad (19)$$

or

$$J_{i+1} = \frac{1}{2 + \varepsilon\Delta\tau^2}(J_i + J_{i+2}) \quad (20)$$

or after the replacement $i + 1 \rightarrow i$ to have a symmetric representation

$$J_i = \frac{1}{2 + \varepsilon\Delta\tau^2}(J_{i-1} + J_{i+1}). \quad (21)$$

We interpret this equation (see Zwillinger 1989) as representing a pseudo particle that after the n th jump at position i will move at jump $n + 1$ either with probability $(2 + \varepsilon\Delta\tau^2)^{-1}$ to position $i + 1$ or with the same probability to position $i - 1$, or go to the absorbing state. The last case has a probability of $1 - (2 + \varepsilon\Delta\tau^2)^{-1}$.

We again have a state space of $N + 1$ states. The transition probability matrix has now in general two non-zero entries in most rows and is therefore slightly more complicated than the matrix above. It is given e.g. in the case $N = 5$ by

$$P = \begin{pmatrix} 0 & \alpha & 0 & 0 & 0 & 0 \\ \alpha & 0 & \alpha & 0 & 0 & 0 \\ 0 & \alpha & 0 & \alpha & 0 & 0 \\ 0 & 0 & \alpha & 0 & \alpha & 0 \\ 0 & 0 & 0 & \alpha & 0 & 0 \\ 1 - \alpha & 1 - 2\alpha & 1 - 2\alpha & 1 - 2\alpha & 1 - \alpha & 1 \end{pmatrix}$$

where

$$\alpha = \frac{1}{2 + \varepsilon\Delta\tau^2}.$$

In contrast, for small $\Delta\tau$ we can approximate

$$(2 + \varepsilon\Delta\tau^2)^{-1} = \frac{1}{2} \left(\frac{1}{1 + \varepsilon\Delta\tau^2/2} \right) \approx \frac{1}{2} (1 - \varepsilon\Delta\tau^2/2) \quad (22)$$

and

$$J_{i+1} + J_{i-1} = J(\tau + \Delta\tau) + J(\tau - \Delta\tau) \approx 2J(\tau) + \frac{d^2J}{d\tau^2}\Delta\tau^2 \quad (23)$$

since

$$J(\tau \pm \Delta\tau) \approx J(\tau) \pm \frac{dJ}{d\tau}\Delta\tau + \frac{1}{2} \frac{d^2J}{d\tau^2}\Delta\tau^2. \quad (24)$$

Therefore, Eq. (21) is now given by

$$J(\tau) = \frac{1}{2} (1 - \varepsilon\Delta\tau^2/2) \left(2J(\tau) + \frac{d^2J}{d\tau^2}\Delta\tau^2 \right) \quad (25)$$

or

$$0 = \frac{d^2J}{d\tau^2} - \varepsilon J(\tau) - \frac{1}{2}\Delta\tau^2 \frac{d^2J}{d\tau^2}, \quad (26)$$

such that in the limit $\Delta\tau \rightarrow 0$ we regain Eq. (17). In other words, based on the assumption of a Markov process with the jumping probabilities of Eq. (21) the transport equation in Eq. (17) follows in the limit of small $\Delta\tau$.

We note also that it can easily be shown that the analytical solution of the difference equation in Eq. (21) in the limit $\Delta\tau \rightarrow 0$ reduces to that of the differential equation of Eq. (17) as it should be.

6. The Milne equation as the potential equation of a Markov process in discrete time

In 1D radiative transfer, the Schwarzschild-Milne equations express the angular moments of the specific intensity by a linear integral operator acting on the source function. Of particular interest is the relation for the zero-order moment, i.e. the mean intensity J , if the source function is proportional to J (i.e. isotropic scattering and no absorption). The resulting linear integral equation

$$J(\tau) = \frac{1}{2} \int_0^\infty J(\tau') E_1(|\tau' - \tau|) d\tau' \quad \text{with } E_1(y) = \int_1^\infty t^{-1} e^{-yt} dt,$$

where the optical depth is defined by $d\tau = \chi dz$ with χ being the extinction coefficient, is called *Milne's equation*, cf. e.g. Mihalas (1978).

While the one-dimensional Milne equation has been extensively discussed in the theory of radiative transfer (cf. e.g. Mihalas 1978; Mihalas & Weibel Mihalas 1984) the Milne equation for *three* spatial dimensions is not well documented so that we start by deriving it here – following Führer (1993) – from the radiative transfer equation.

The radiative transfer equation for a static 3D medium with isotropic coherent scattering can be written as the directional derivative

$$n \cdot \nabla I(x + sn, n) = \frac{dI(x + sn, n)}{ds} = -\chi I(x + sn, n) + (1 - \varepsilon)\chi J(x + sn) + \varepsilon\chi B(x + sn), \quad (27)$$

where x is the position vector with components (x_1, x_2, x_3) . We assume that $x_3 = z$ and $I(x + sn, n)$ is the specific intensity (at a fixed wavelength) at spatial position $x + sn$ in the direction n . The mean intensity is denoted by $J(x + sn)$, and the (isotropic) Planck function by $B(x + sn)$. The ratio ε of absorption to extinction at the wavelength considered is assumed to be constant with respect to x .

The geometry of the medium is mathematically represented by a finite *convex* domain G . It follows that there exists a unique real number s_0 for any $x \in G$ such that $(x - s_0n)$ is on the boundary. For the boundary conditions, we have to assume that for all points on the boundary of G the intensities of the rays entering G are known since otherwise no sensible solution could be found.

If, however, G is not convex, some regions of the domain are not directly connected to each point of G . In particular, the mutual illumination by the domain's surface elements is not taken into account in this case. One way to deal with this problem is to imbed G in a larger domain in which then, however, the value of χ has to be set to 0 in the region outside G .

Integration of the differential equation in Eq. (27) yields

$$I(x + sn, n) = e^{-\chi(s+s_0)} I(x - s_0n, n) + \int_{-s_0}^s e^{-\chi(s-s')} \left((1 - \varepsilon)\chi J(x + s'n) + \varepsilon\chi B(x + s'n) \right) ds'. \quad (28)$$

To keep the expressions simple, we have assumed here that the extinction coefficient χ is independent of x . Otherwise we have to substitute here and in the following expressions $\chi \cdot s$ with the integral $\int \chi ds$.

We now set $s = 0$ and $s_0 = s_0(x, n)$ and obtain

$$I(x, n) = e^{-\chi \cdot s_0(x, n)} I(x - s_0(x, n), n) + \int_{-s_0(x, n)}^0 e^{+\chi \cdot s'} \left((1 - \varepsilon) \chi J(x + s'n) + \varepsilon \chi B(x + s'n) \right) ds'. \quad (29)$$

The mean intensity is now obtained by integrating the above equation over the unit sphere \mathbb{S}^2

$$J(x) = \int_{\mathbb{S}^2} e^{-\chi \cdot s_0(x, n)} I(x - s_0(x, n), n) \frac{d\omega(n)}{4\pi} + \int_{\mathbb{S}^2} \int_{-s_0(x, n)}^0 e^{+\chi \cdot s'} \left((1 - \varepsilon) \chi J(x + s'n) + \varepsilon \chi B(x + s'n) \right) ds' \frac{d\omega(n)}{4\pi}. \quad (30)$$

The source term equals

$$S(x) = \int_{\mathbb{S}^2} \left(e^{-\chi \cdot s_0} I(x - s_0(x, n), n) + \int_{-s_0(x, n)}^0 \varepsilon \chi e^{+\chi \cdot s'} B(x + s'n) ds' \right) \frac{d\omega(n)}{4\pi}. \quad (31)$$

The first term includes the boundary conditions, and the second contains the sources inside G . Then

$$J(x) = \int_G p(x, x') J(x') dx' + S(x), \quad (32)$$

where

$$\begin{aligned} \int_G p(x, x') J(x') dx' &= \int_{\mathbb{S}^2} \frac{d\omega(n)}{4\pi} \int_{-s_0(x, n)}^0 ds' e^{+\chi \cdot s'} (1 - \varepsilon) \chi J(x + s'n) \\ &= \int_{\mathbb{S}^2} \frac{d\omega(n)}{4\pi} \int_0^{s_0(x, -n)} ds' e^{-\chi \cdot s'} (1 - \varepsilon) \chi J(x + s'n). \end{aligned} \quad (33)$$

The third term arises from the second one by introducing $-n$ instead of n and $-s'$ instead of s' as new variables. We introduce the variable $x' = x + sn$, where $s = |x - x'|$ and the volume element is

$$dx' = 4\pi s^2 ds d\omega(n). \quad (34)$$

We obtain

$$\int_G p(x, x') J(x') dx' = \int_G \frac{e^{-\chi|x-x'|}}{4\pi|x-x'|^2} (1 - \varepsilon) \chi J(x') dx'$$

or

$$p(x, x') = (1 - \varepsilon) \chi \frac{e^{-\chi|x-x'|}}{4\pi|x-x'|^2}.$$

Finally

$$J(x) = \int_G \frac{e^{-\chi|x-x'|}}{4\pi|x-x'|^2} (1 - \varepsilon) \chi J(x') dx' + S(x), \quad (35)$$

which is the *3D Milne equation*.

We now define a *Markov process* and show that $J(x)$ is the solution of the corresponding potential equation.

The state space is continuous and given by $G \cup \Phi$, the union of G and the graveyard Φ . The initial probability has the density

$$\pi(x) = \frac{S(x)}{\int_G S(x') dx'}. \quad (36)$$

The graveyard is empty at the beginning. In the continuous case, the probability matrix P (Eq. (2)) becomes the transition probability $P(x, x_0)$ of moving from x_0 to x , which is given in the following way:

(i) if $x_0 = \Phi$, then $x = \Phi$;

(ii) if $x_0 \in G$, then

(iia) choose a random number

$$\underline{a} = \begin{cases} 1 & \text{with probability } (1 - \varepsilon) \\ 0 & \text{probability } \varepsilon. \end{cases}$$

If $\underline{a} = 1$, the particle is scattered; if $\underline{a} = 0$ it is absorbed and goes to the graveyard;

(iib) choose a random direction \underline{n} uniformly distributed across the unit sphere (n is the direction in which the particle is scattered);

(iic) choose a random number \underline{s} that is exponentially distributed. The probability that \underline{s} is in the interval $(s \dots s + ds)$ is $\chi e^{-\chi \cdot s} ds$. If $x = x_0 + \underline{n} \underline{s} \in G$, continue, if $x = x_0 + \underline{n} \underline{s} \notin G$, the particle goes to the graveyard.

We calculate the transition probability function of moving for coming from $x_0 \in G$ to $x \in G$, by using a test function f on G . The expectation value of f is

$$\begin{aligned} \mathbb{E} \underline{a} f(x_0 + \underline{n} \underline{s}) \hat{\mathbf{1}}_{\{\underline{s} \in [0, s(x_0, -n)]\}} &= \\ (1 - \varepsilon) \int \frac{d\omega(n)}{4\pi} \int_0^{s(x_0, -n)} \chi e^{-\chi \cdot s} f(x_0 + ns) ds &= \\ = \int f(x) p(x, x_0) dx, \end{aligned} \quad (37)$$

where $\hat{\mathbf{1}}$ is the characteristic function of the set. Comparing with Eqs. (33) and (35), we see that the Milne's equation is the potential equation for this Markov process. We obtain the transition probability from x_0 to the graveyard Φ

$$p(\Phi, x_0) = \varepsilon + (1 - \varepsilon) \int_{G^c} \frac{\chi}{4\pi|x-x_0|^2} e^{-\chi|x-x_0|} dx. \quad (38)$$

To perform the algorithm, we divide the set G into N cells of equal volume and numerate the cells by $1, \dots, N$. We calculate the source term $S(x)$ given by Eq. (32) and calculate the initial probability π with the help of Eq. (36). Then following the prescription above, choose random variables $(\underline{a}_1, \underline{n}_1, \underline{s}_1), (\underline{a}_2, \underline{n}_2, \underline{s}_2) \dots$, independent from each other until the particle ends up in the graveyard. We note the number of times \hat{N}_j the particle visits the cell j and calculate J , as it has been explained in Sect. 3.

We note that if we generalize to $G = \mathbb{R}^3$ and assume that there is no absorption, we have a random walk that ends in the graveyard at infinity. The probability density of moving from x_0 to x is

$$p(x, x_0) = (1 - \varepsilon) \frac{\chi}{4\pi|x-x_0|^2} e^{-\chi|x-x_0|}. \quad (39)$$

7. Selected astronomical applications of MCMC to the Milne equation

We discuss a few applications that demonstrate the versatility and ease of use of MCMC. We restrict ourselves to configurations that have the same values for ε , χ , and B everywhere but

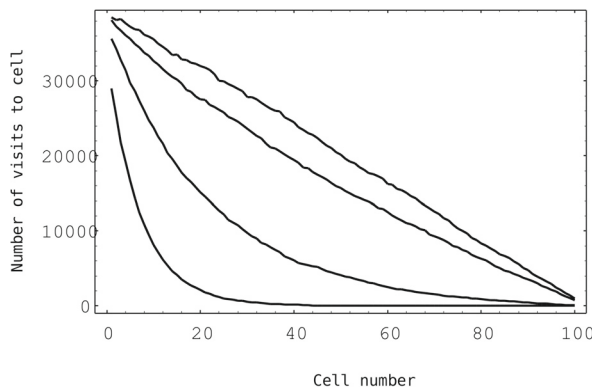


Fig. 1. Run of the number of visits of a pseudo particle in a cell (which is proportional to the mean intensity) with cell number (which is proportional to the optical depth starting at the left) in a slab that is illuminated from the left but does not have internal sources. The curves represent ratios of absorption to extinction $\varepsilon = 0, 10^{-3}, 10^{-2}, 10^{-1}$ (from top to bottom). In all cases, $N = 10^5$ pseudo particles are used.

emphasize that this is not an inherent limitation of this method. We note that the calculation of the mean intensity for $\varepsilon = 0$ is equivalent to the determination of the temperature distribution for a grey configuration. This implies that it is possible to determine this temperature distributions using MCMC without difficulties.

We begin with some easy 1D problems, the *radiation fields of slabs*, that can also efficiently be solved by standard numerical algorithms. The first test case is the mean intensity of a slab calculated for the two-stream approximation. Here the kernel of the Milne equation is the exponential function $\exp(-\chi s)$, so that the corresponding random numbers can easily be generated by $-1/\chi \cdot \log(\chi \times \text{Random})$ where *Random* refers to random numbers that are equally distributed between 0 and 1. We devised a corresponding Mathematica version 2.1 code for this purpose that is only 15 lines long. If the slab has an extinction coefficient of unity and a total optical depth of 50 and is illuminated from one side but does not have internal sources, the CPU times for 10^5 pseudo particles range between 12 s (ε close to 1) and 91 s (ε close to 0). The distributions for the mean intensities (Fig. 1) agree with those from the analytic solution to within the width of the line. If the illuminating source is switched off and a homogeneous distribution of internal sources is employed instead, the CPU times are similar. Figure 2 shows an example in which 10^6 pseudo particles are used.

We have performed calculations in which the full angle space is taken into account. These take much longer since the transformation law for the random numbers has to be evaluated numerically. On the other hand, if the Mathematica code is replaced by a corresponding FORTRAN code the execution times are reduced by about a factor of one hundred.

To convince ourselves of the versatility of the MCMC approach, we calculate some *3D radiation fields* of mean intensity (i) in a homogeneous rotational ellipsoid having internal sources; (ii) of a homogeneous accretion disk with constant flaring angle heated by a constant viscosity; and (iii) of a simple accretion funnel heated by the accretion shock at the bottom (stellar surface).

The corresponding FORTRAN code comprises just about hundred lines and is therefore one to several orders of magnitude shorter than grid codes for these problems. Owing to the larger number of cells, however, many more pseudo particles have to be used (in our case 10^8 to 10^9) to achieve high quality statistics

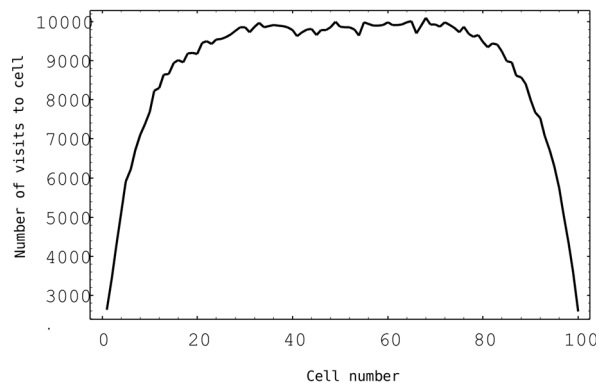


Fig. 2. Run of the number of visits of a pseudo particle in a cell (which is proportional to the mean intensity) with cell number (which is proportional to the optical depth starting at the left) in a slab with uniformly distributed sources but without illumination. The ratio of absorption to extinction is $\varepsilon = 10^{-1}$ and $N = 10^6$ pseudo particles are used.

but the CPU times still do not exceed about 10 min if the Fortran code is used on a 2 GHz laptop. Where the configuration is not fully convex as for the accretion disk or for the funnel, and when the illumination of parts of the medium by other parts cannot be neglected, one has to wrap the configuration with a box and assume that the intermediate domain is very optically thin. If the mutual illumination is small one may proceed as in the convex case.

8. Discussion

We propose that the most important advantages of MCMC as presented here are its sound mathematical basis and the clarification of the similarities between the treatments of radiative transfer problems using a Monte Carlo method, a difference equation method, or differential/integral equation methods. From the point of view of actual numerical calculations, when compared to grid methods (for a review, see Wehrse & Kalkofen 2006), the MCMC method presented here shares all the advantages of classical Monte Carlo methods: (i) it is easy to implement even for complex geometries; (ii) it allows the incorporation of many physical effects; (iii) it is quite fast for medium accuracy requirements; and (iv) the corresponding codes run well on classical parallel machines as well as on massively parallel devices such as modern graphics cards. As already observed by Lucy (1999a,b), Monte Carlo methods are very flexible. They allow the direct calculation of derived radiative quantities such as e.g., the mean intensity without resorting to the explicit calculation of specific intensities. They allow the computation of residual intensities occurring in the core saturation method, cf. Bastian et al. (1980). Furthermore, MCMC can be used to directly calculate a grey temperature distribution and we should be able to calculate directly mean intensities for differentially moving media. However, some disadvantages are obvious: (i) there is no precise and well-founded error estimator as it exists for finite-element approaches, cf. Meinköhn et al. (2007); (ii) additional quantities, such as e.g., the radiative flux, cannot be obtained by differentiation because of the error-amplifying nature of this process; (iii) the number of pseudo-particles required is sometimes very high (and therefore the CPU times are sometimes rather long) since the use of convergence accelerating methods is limited; and (iv) it can hardly be used when simultaneously non-linear equations (as e.g. hydrodynamic equations) have in addition to be solved. Here we have focused mainly on the geometrical structure of the

radiation. The distribution of frequency and the dependence of the temperature on the radiation that are included e.g., in Lucy's papers have not been considered.

However, it should be possible to generalize the algorithm described here (i) to calculate the radiation pressure and the radiative acceleration, since the calculation of these quantities can be done by integration, which can also be performed – as is well documented, see e.g. Liu (2001) – by MCMC methods; (ii) to evaluate integrals such as $\int_0^\infty \epsilon_\chi J d\lambda$ that are usually needed in the construction of model atmospheres; and (iii) to model differentially moving configurations. In this case, Eq. (17) becomes a full telegrapher's equation but one that can also be solved in a stochastic way, cf. Zauderer (1983).

References

- Bastian, U., Bertout, C., Stenholm, L. G., & Wehrse, R. 1980, *A&A*, 86, 105
- Cox, D. R., & Miller, H. D. 1965, *The Theory of Stochastic Processes* (London: Chapman and Hall)
- Diaconis, P. 2009, *Bull. AMS*, 46, 179
- Esposito, L. W., & House, L. L. 1978, *ApJ*, 219, 1058
- Führer, C. 1993, Diploma Thesis, Inst. f. Angewandte Mathematik, Heidelberg Univ.
- Hertz, P. 1909, *Math. Ann.*, 67, 387
- Juvela, M. 2005, *A&A*, 440, 531
- Liu, J. S. 2001, *Monte Carlo Strategies in Scientific Computing* (Heidelberg: Springer)
- Lucy, L. B. 1999a, *A&A*, 344, 282
- Lucy, L. B. 1999b, *A&A*, 345, 211
- Meinköhn, E., Kanschä, G., Rannacher, R., & Wehrse, R. 2007, *Numerical Methods for Multidimensional Radiative Transfer, in Reactive Flows, Diffusion and Transport*, ed. W. Jäger, R. Rannacher, & J. Warnatz (Heidelberg: Springer), 485
- Meyer, P.-A. 1966, *Probabilités et potentiel* (Paris: Herman)
- Mihalas, D. 1978, *Stellar Atmospheres*, 2nd edn. (San Francisco: Freeman)
- Mihalas, D., & Weibel Mihalas, B. 1984, *Foundations of Radiation Hydrodynamics* (New York: Oxford University Press)
- Papoulis, A. 1991, *Probability, Random Variables, and Stochastic Processes* (New York: McGraw-Hill)
- Reif, F. 1976, *Physikalische Statistik und Physik der Wärme* (Berlin: de Gruyter)
- Traving, G. 1960, *Über die Theorie der Druckverbreiterung von Spektrallinien* (Karlsruhe: G. Braun)
- Waldenfels, W. von 2009a, in *Stochastic Properties of the Radiative Transfer Equation*, in *Numerical Methods in Multidimensional Radiative Transfer*, ed. G. Kanschä, E. Meinköhn, R. Rannacher, & R. Wehrse (Heidelberg: Springer), 19
- Waldenfels, W. von 2009b, *Stochastics*, 81, 299
- Wang, M. C., & Uhlenbeck, G. E. 1945, *Rev. Mod. Phys.*, 17, 323
- Wehrse, R., & Kalkofen, W. 2006, *A&ARv*, 13, 3
- Williams, D. 2001, *Weighing the Odds* (Cambridge: Cambridge University Press)
- Zauderer, E. 1983, *Partial Differential Equations of Applied Mathematics* (Wiley & Sons)
- Zwillinger, D. 1989, *Handbook of Differential Equations* (Boston: Academic Press)