Monte Carlo simulation of the general elliptic operator



1987 J. Phys. A: Math. Gen. 20 2693

(http://iopscience.iop.org/0305-4470/20/10/013)

J. Phys. A: Math. Gen. 20 (1987) 2693-2697. Printed in the UK

# Monte Carlo simulation of the general elliptic operator

Jan Vrbik

Department of Mathematics, Brock University, St Catharines, Ontario L2S 3A1, Canada

Received 1 August 1986

**Abstract.** Several versions of an algorithm for Monte Carlo simulation of solutions of partial differential equations involving a general elliptic operator are developed, with the emphasis on quadratic accuracy and computational efficiency.

## 1. Introduction

Vrbik (1985) describes a quadratically accurate Monte Carlo technique for simulating the action of the operator

$$\exp[-t(-\tfrac{1}{2}\nabla^2+\nabla \boldsymbol{F}(\boldsymbol{r})+E(\boldsymbol{r}))] \tag{1}$$

when applied to an arbitrary non-negative function, represented by a random sample of 'configurations'.

Such a simulation can be used to solve

$$-\tfrac{1}{2}\nabla^2 f(\boldsymbol{r},t)+\nabla \boldsymbol{F}(\boldsymbol{r})f(\boldsymbol{r},t)+E(\boldsymbol{r})f(\boldsymbol{r},t)=-(\partial/\partial t)f(\boldsymbol{r},t) \tag{2}$$

or its time-independent counterpart posed, usually, as an eigenvalue problem, i.e.

$$-\tfrac{1}{2}\nabla^2 f(\boldsymbol{r})+\nabla \boldsymbol{F}(\boldsymbol{r})f(\boldsymbol{r})+(E(\boldsymbol{r})-E_0)f(\boldsymbol{r})=0. \tag{3}$$

We would like to extend these results to the more general

$$\exp[-t(-\tfrac{1}{2}\nabla_j\nabla_k D_{jk}(\boldsymbol{r})+\nabla \boldsymbol{F}(\boldsymbol{r})+E(\boldsymbol{r}))] \tag{4}$$

(with implied summation over each pair of identical indices), where $\boldsymbol{D}(\boldsymbol{r})$ is a symmetric positive-definite matrix for all values of $\boldsymbol{r}$ ($N$-dimensional vector). Also note that $D_{jk}(\boldsymbol{r})$ is inside the operational range of both $\nabla_j$ and $\nabla_k$.

This extension will be useful for solving all physical problems described by the Fokker-Planck equation and having a non-negative solution. The most eminent examples are the phenomenon of diffusion (possibly accompanied by creation/absorption of particles), heat conduction and flow of fluids through porous media, allowing a non-homogeneous environment undergoing its own motion.

Another potential application relates to solving the Schrödinger equation to obtain the ground state of a complex system of particles (atoms or molecules). One may seek a solution in the form of $f(\boldsymbol{R})g(\boldsymbol{R})$, where $f(\boldsymbol{R})$ is a known function of the particles' locations, constructed to approximate the cusps of the exact solution. This should smooth out the 'rough edges' of the remaining $g(\boldsymbol{R})$ function to be simulated (more accurately) by the proposed technique. Thus $f(\boldsymbol{R})$ will effectively become the new

0305-4470/87/102693+05$02.50  

(varying) diffusion coefficient and will similarly modify the other terms of the original equation.

## 2. Linearly accurate solution

In this section we aim to achieve first-order accuracy (in terms of expanding expression (4) in $t$), which means that the error of the corresponding procedure for solving (3) will be proportional to $t$ (the discrete time step of the simulation).

We already know how to simulate operators corresponding to the second and third terms of (4) (see Reynolds *et al* (1982) or Vrbik and Rothstein (1985)) via 'drift' and 'branching', respectively, so the only missing component remains the simulation of the 'diffusion' operator

$$\exp(\tfrac{1}{2}t\nabla_j\nabla_k D_{jk}(\boldsymbol{r})). \tag{5}$$

We claim that this can be achieved by the following.

(i) Finding a matrix $\boldsymbol{D}^{1/2}(\boldsymbol{r})$ such that

$$\boldsymbol{D}^{1/2}\cdot(\boldsymbol{D}^{1/2})^{\mathrm{T}}=\boldsymbol{D}(\boldsymbol{r}). \tag{6}$$

Note that this can be done easily by choosing $\boldsymbol{D}^{1/2}$ to have the lower-triangular form (and *not* the standard symmetric solution obtained by diagonalising $\boldsymbol{D}$).

(ii) Advancing each configuration $\boldsymbol{r}$ to a new location $\boldsymbol{r}'$ (see Vrbik 1985) by

$$\boldsymbol{r}'=\boldsymbol{r}+\boldsymbol{D}^{1/2}(\boldsymbol{r})\cdot z \tag{7}$$

where the $z_i$ are $N$ independent random values drawn from a symmetric (such as normal or uniform) distribution with zero mean and variance equal to $t$.

To prove (7), first consider the following function of $\boldsymbol{r}'$, $\boldsymbol{r}$ and $t$:

$$G(\boldsymbol{r}'\leftarrow\boldsymbol{r},t)=(2\pi t)^{-N/2}\det(\boldsymbol{D}^{-1/2}(\boldsymbol{r}))\exp[-(\boldsymbol{r}'-\boldsymbol{r})^{T}\cdot\boldsymbol{D}^{-1}(\boldsymbol{r})\cdot(\boldsymbol{r}'-\boldsymbol{r})/2t] \tag{8}$$

where $\boldsymbol{D}^{-1/2}=(\boldsymbol{D}^{1/2})^{-1}$. Note that

$$\int G(\boldsymbol{r}'\leftarrow\boldsymbol{r},t)f(\boldsymbol{r})\,\mathrm{d}\boldsymbol{r} \tag{9}$$

is, in statistical terms, equivalent to (7), with $f(\boldsymbol{r})$ being the probability density function of $\boldsymbol{r}$, and $z$ having the normal distribution mentioned above.

If we now define

$$B_{ij}=D_{ij}^{-1/2}(\boldsymbol{r})+[\nabla_j D_{ik}^{-1/2}](\boldsymbol{r}-\boldsymbol{r}')_k \tag{10}$$

(the operational range of $\nabla$ being restricted to the square brackets), the substitution

$$z=\boldsymbol{D}^{-1/2}\cdot(\boldsymbol{r}-\boldsymbol{r}') \tag{11}$$

(thus $\partial z_i/\partial r_j=\boldsymbol{B}_{ij}$) enables us to write

$$\lim_{t\to 0}\int G(\boldsymbol{r}'\leftarrow\boldsymbol{r},t)\cdot f(\boldsymbol{r})\,\mathrm{d}\boldsymbol{r}=\lim_{t\to 0}\int(2\pi t)^{-N/2}$$
$$\times\exp(-z^2/2t)\det(\boldsymbol{D}^{-1/2})/\det(\boldsymbol{B})\cdot f(\boldsymbol{r}(z))\,\mathrm{d}z=f(\boldsymbol{r}') \tag{12}$$

since

$$\lim_{t\to 0}(2\pi t)^{-N/2}\exp(-z^2/2t)=\delta(z)$$
$$\delta(z)\,\mathrm{d}z=\delta(\boldsymbol{r}-\boldsymbol{r}')\,\mathrm{d}\boldsymbol{r} \qquad \text{and} \qquad \det(\boldsymbol{D}^{-1/2})/\det(\boldsymbol{B})|_{r=r'}=1. \tag{13}$$

Thus the zeroth-order term of (9) (seen as an operator applied to $f$) is the identity. To derive its first-order terms, we need

$$-\lim_{t\to 0}\frac{\partial}{\partial t}\int G(\boldsymbol{r}'\leftarrow\boldsymbol{r},t)f(\boldsymbol{r})\,\mathrm{d}\boldsymbol{r}$$
$$=-\lim_{t\to 0}\tfrac{1}{2}\int(2\pi t)^{-N/2}[\nabla_z^2\exp(-z^2/2t)]\cdot\det(\boldsymbol{D}^{-1/2})/\det(\boldsymbol{B})f(\boldsymbol{r}(z))\,\mathrm{d}z$$
$$=-\lim_{t\to 0}\tfrac{1}{2}\int(2\pi t)^{-N/2}\exp(-z^2/2t)\cdot\nabla_z^2\det(\boldsymbol{D}^{-1/2})/\det(\boldsymbol{B})f(\boldsymbol{r}(z))\,\mathrm{d}z$$
$$=-\tfrac{1}{2}\int\delta(\boldsymbol{r}-\boldsymbol{r}')B_{ji}^{-1}\nabla_j B_{ki}^{-1}\nabla_k\det(\boldsymbol{D}^{-1/2})/\det(\boldsymbol{B})f(\boldsymbol{r})\,\mathrm{d}\boldsymbol{r} \tag{14}$$

as $\nabla_{zi}=B_{ji}^{-1}\nabla_j$.

Explicit evaluation of the last integral is rather tedious and rests on the following formulae:

$$\nabla_j B_{ki}^{-1}|_{r=r'}=(A_{jil}+A_{lij})D_{lk}^{1/2}(\boldsymbol{r}')$$
$$\nabla_k\det(\boldsymbol{D}^{-1/2})/\det(\boldsymbol{B})|_{r=r'}=C_k(\boldsymbol{r}') \tag{15}$$

and

$$\nabla_j\nabla_k\det(\boldsymbol{D}^{-1/2})/\det(\boldsymbol{B})=C_kC_j+\nabla_kC_j+\nabla_jC_k+A_{muk}A_{umj} \tag{16}$$

where $C_k=[\nabla_m D_{mu}^{1/2}]D_{ui}^{-1/2}$ and $A_{jil}=[\nabla_j D_{iu}^{1/2}]D_{ul}^{-1/2}$.

When confirming these, remember that, in general,

$$\nabla H_{ki}^{-1}=-H_{kl}^{-1}[\nabla H_{lm}]H_{mi}^{-1} \qquad \nabla\det\{\boldsymbol{H}\}=[\nabla H_{jk}]H_{kj}^{-1}\det\{\boldsymbol{H}\}.$$

With the help of these, one obtains as the final result of integration (14)

$$-\tfrac{1}{2}\nabla_j\nabla_k D_{jk}f(\boldsymbol{r}') \tag{17}$$

as desired.

## 3. Quadratically accurate solution

To derive a $t^2$ accurate analogue of (8), it is necessary to use the more powerful technique of Risken (1984) which utilises the Fourier transform of the $\delta$ function. This yields

$$G(\boldsymbol{r}'\leftarrow\boldsymbol{r},t)=\exp[-t(-\tfrac{1}{2}\nabla_i\nabla_jD_{ij}(\boldsymbol{r}))^{\mathrm{A}}]\delta(\boldsymbol{r}-\boldsymbol{r}')$$
$$=\exp[(t/2)D_{ij}\nabla_i\nabla_j](2\pi)^{-N}\int\exp\{\mathrm{i}\boldsymbol{u}(\boldsymbol{r}-\boldsymbol{r}')\}\,\mathrm{d}\boldsymbol{u}$$

$$= (2\pi)^{-N} \int [1 - (t/2) D_{ij} u_i u_j + (t^2/8) D_{kl} u_k u_l D_{ij} u_i u_j$$
$$- (t^2/8) D_{kl} [\nabla_k \nabla_l D_{ij}] u_i u_j$$
$$- \mathrm{i}(t^2/4) D_{kl} [\nabla_l D_{ij}] u_i u_j u_k + \ldots] \exp[\mathrm{i}u(r - r')]\,\mathrm{d}u \tag{18}$$

where the dots imply terms of third and higher order in $t$, $\mathrm{i}^2 = -1$, and the superscript A turns an operator into its adjoint ($-\frac{1}{2}D_{ij}\nabla_i\nabla_j$ in this case). $\boldsymbol{D}$ and its derivatives are to be evaluated at $\boldsymbol{r}$ ($\boldsymbol{r}'$ is considered a fixed parameter here).

The last integral can be easily carried out (to within the $t^2$ accuracy), resulting in the probability density function of the random variable $\boldsymbol{r}' - \boldsymbol{r}$ (now, $\boldsymbol{r}'$ is considered varying and $\boldsymbol{r}$ fixed!). Unfortunately, this itself would not provide an explicit prescription for simulating values from such a distribution; thus we have to employ the following alternative approach.

From the definition of a characteristic function of a distribution and the related theory (see, for example, Cramer 1971), it is immediately obvious that the following expression from the last integral

$$1 - (t/2) D_{ij} u_i u_j + (t^2/8) D_{kl} u_k u_l D_{ij} u_i u_j - (t^2/8) D_{kl} [\nabla_k \nabla_l D_{ij}] u_i u_j$$
$$- \mathrm{i}(t^2/4) D_{kl} [\nabla_l D_{ij}] u_i u_j u_k + \ldots \tag{19}$$

represents the characteristic function of the required distribution. From this, we can determine all the distribution's moments. These are

$$\begin{aligned} M_i^{(1)} &= 0 + \ldots \\ M_{ij}^{(2)} &= tD_{ij} + (t^2/4) D_{kl} [\nabla_k \nabla_l D_{ij}] + \ldots \\ M_{ijk}^{(3)} &= (t^2/2)(D_{kl}[\nabla_l D_{ij}] + D_{il}[\nabla_l D_{kj}] + D_{jk}[\nabla_l D_{ik}]) + \ldots \\ M_{ijkl}^{(4)} &= t^2(D_{ij} D_{kl} + D_{ik} D_{jl} + D_{il} D_{jk}) + \ldots \end{aligned} \tag{20}$$

for $i, j, k, l = 1, 2, \ldots, N$, all the higher moments being zero (in the $+\ldots$ sense). Note that the moments have been properly symmetrised.

All we have to do now is to construct a random vector with these moments and use it for the actual simulation of $\boldsymbol{r}' - \boldsymbol{r}$. It is not difficult to check that a possible solution is

$$D_{ij}^{1/2} z_j + (t/8) D_{jm}^{-1/2} D_{kl} [\nabla_k \nabla_l D_{mi}] z_j + R_{ijk} z_j z_k - tR_{ijj} + tS_{ij} z_j \tag{21}$$

where

$$R_{ijk} = \tfrac{1}{4} D_{jm}^{-1/2} [\nabla_l D_{mi}] D_{lk}^{1/2}$$
$$S_{ij} = -\tfrac{1}{2} D_{jm}^{-1/2} (R_{mlk} R_{ilk} + R_{mlk} R_{ikl})$$

and $z$ is a random vector with independent components generated from the normal distribution with zero mean and variance equal to $t$ (or an equivalent—any symmetric distribution with the same first *four* moments will do; let us call such a distribution $\mathcal{N}\{0, t\}$).

The essential formulae to help verify (21) are

$$\begin{aligned} \mathbb{E}(z_i) &= 0 \\ \mathbb{E}(z_i z_j) &= \delta_{ij} t \\ \mathbb{E}(z_i z_j z_k) &= 0 \end{aligned} \tag{22}$$

and

$$\mathbb{E}(z_i z_j z_k z_l) = (\delta_{ij}\delta_{kl} + \delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}) t^2$$

where $\mathbb{E}$ denotes the expected value of a random variable.

Note that in the isotropic case of $D_{ij}(\boldsymbol{r}) = \delta_{ij} D(\boldsymbol{r})$, expression (21) reduces to

$$D^{1/2} z_i + (t/8) D^{1/2} [\nabla^2 D] z_i + (1/4)[\nabla_j D] z_i z_j - (t/4)[\nabla_i D]$$
$$- (t/32) D^{-1/2} ([\nabla_j D][\nabla_j D] z_i + [\nabla_i D][\nabla_j D] z_j). \qquad (23)$$

The actual simulation (of both the general and isotropic case) can be simplified even further (with the objective of avoiding derivatives of $D_{ij}$) if we replace, in expression (21),

$$tD_{kl}[\nabla_k \nabla_l D_{mi}] \quad \text{by} \quad (D^{(+)} + D^{(-)} - 2D(\boldsymbol{r}))_{mi}$$

and (24)

$$R_{ijk} \quad \text{by} \quad (1/8t) D_{jm}^{-1/2} (D^{(+)} - D^{(-)})_{mi} x_k$$

where $\boldsymbol{x}$ is a random vector generated, *independently of* $\boldsymbol{z}$, from $\boldsymbol{N}\{0, t\}$ and

$$\boldsymbol{D}^{(+)} = \boldsymbol{D}(\boldsymbol{r} + \boldsymbol{D}^{1/2}(\boldsymbol{r}) \cdot \boldsymbol{x})$$
$$\boldsymbol{D}^{(-)} = \boldsymbol{D}(\boldsymbol{r} - \boldsymbol{D}^{1/2}(\boldsymbol{r}) \cdot \boldsymbol{x}). \qquad (25)$$

It is a simple exercise to check that such a replacement will not change the moments of (21); thus the final version of simulating 'diffusion' of a configuration with an initial location at $\boldsymbol{r}$ is to advance it by adding, to $\boldsymbol{r}$, the following random vector:

$$D_{ij}^{1/2} z_j + (1/8)(D^{(+)} + D^{(-)} - 2D)_{ij} Z_j$$
$$+ (1/8t) z_k x_k (D^{(+)} - D^{(-)})_{ij} Z_j - (1/8)(D^{(+)} - D^{(-)})_{ij} X_j$$
$$- (1/128t) x_k x_k (D^{(+)} - D^{(-)})_{ij} D_{jl}^{-1} (D^{(+)} - D^{(-)})_{lm} Z_m$$
$$- (1/128t)(D^{(+)} - D^{(-)})_{ij} X_j \cdot X_l (D^{(+)} - D^{(-)})_{lm} Z_m \qquad (26)$$

where $z$ and $\boldsymbol{x}$ are generated, independently, from $\boldsymbol{N}\{0, t\}$ and

$$Z_j = z_l D_{lj}^{-1/2} \quad \text{and} \quad X_j = x_l D_{lj}^{-1/2}.$$

Computationally, this will involve three evaluations of $\boldsymbol{D}$, one matrix inversion and some further simple matrix manipulation (note that the full matrix multiplication is *not* required). This seems a relatively modest cost of a potentially significant increase in accuracy.

## References


Cramer H 1971 *Mathematical Methods of Statistics* 12th part (Princeton: Princeton University Press) p 101

Reynolds P J, Ceperley D M, Alder B J and Lester W A Jr 1982 *J. Chem. Phys.* **77** 5593

Risken H 1984 *The Fokker–Planck Equation* (Berlin: Springer) p 69

Vrbik J 1985 *J. Phys. A: Math. Gen.* **18** 1327

Vrbik J and Rothstein S M 1985 *J. Comput. Phys.* **63** 130