JOURNAL OF COMPUTATIONAL PHYSICS 111, 33–48 (1994)

A Synopsis of Monte Carlo Perturbation Algorithms

HERBERT RIEF

Commission of the European Communities, Ispra Establishment, I-21020 Ispra (Va), Italy
Guest lecturer at the Institut für Ionenphysik der Universität Innsbruck. A-6020 Innsbruck (Austria)

Received May 20, 1992; revised January 4, 1993

Fundamental aspects of correlated sampling and differential operator procedures applied to integrals and systems of linear equations modelling Markov processes are investigated. Algorithms providing sensitivities (gradients, Jacobians) and perturbation estimates obtained by a single simulation experiment are described in detail and explained by examples. Mathematical proofs are provided which show that under most conditions a finite relative variance can be obtained for arbitrarily small parameter variations.

1. INTRODUCTION

In science and technology the demand for the analysis of complex, multidimensional systems is continuously increasing. Compared to deterministic methods Monte Carlo allows a less demanding modelling of such problems, provided the underlying elementary processes are of a stochastic nature or can be formulated in such terms.

Until quite recently most Monte Carlo programs had to be run on expensive mainframe computers. The advent of inexpensive systems based on mass produced CPUs, for example, transputers or acceleration boards, is going to change this situation dramatically. Right from the beginning it was realized that the Monte Carlo method was one of the most promising candidates specifically suited for parallel computers. It is the general conviction that in the light of these facts the role of Monte Carlo as a powerful tool of numerics will have to be re-evaluated in the near future.

Despite this advantage the probabilistic nature of the method imposes certain limitations if it is applied in a straightforward manner. Parameter studies which constitute a great fraction of any complex system analysis prove, for example, impossible if the differences of the sampled responses are of the same size or smaller than their uncertainties. Actually, more often than not, the perturbation of a response as a consequence of small parameter changes, turns out to be of greater interest than the response itself. But in proportion to the rich literature on the theory of stochastic simulations [1, 3, 4, 6, 12, 13] and the countless papers describing its application to a specific class of problems, the aspect of Monte Carlo perturbation algorithms has been dealt with only sporadically in the context of specific problems [2, 5, 7–11].

To overcome this shortcoming, two methods allowing for the calculation of small parameter variations will be described here in a systematic manner. One is based on *correlation techniques* and the other deals with *differential operator sampling*. The author tries to prove that for most conditions the perturbation estimates determined by these sampling schemes have a finite relative variance.

While *correlation techniques* provide straightforward perturbation estimates, *differential operator sampling* allows for the determination of gradients and higher order derivatives. The latter method offers the possibility to perform perturbation calculations by the use of a multivariate Taylor expansion and to analyze the stochastic simulation process in terms of additional criteria, such as:

— the determination of the sensitivity coefficients for the parameter vector $\mathbf{p}$ describing the system being analyzed,

— the estimation of the uncertainty (variance) of the target quantities as a function of the scatter (uncertainty) of the parameters p_i,

— the enventual adjustment of the parameters p_i such that the discrepancy between measurements and calculations is minimized.

2. TERMINOLOGY AND DEFINITIONS

Stochastic simulation procedures, usually called Monte Carlo methods, always aim at the determination of a mean value which can be interpreted as being an integral or the solution of a (linear) system of equations. The functions to be integrated are probability density functions (pdfs) from which discrete variables are sampled randomly to perform the simulation process.

Random variables are a set of real numbers which correspond to a probability distribution function. In the case of continuous variables, we define

$$g(u)\,du = P[u \leqslant u' < u + du];$$

0021-9991/94 $6.00

$g(u)$ is the pdf of u and gives the probability of finding the random variable u' within the interval $[u, u+du]$.

As we shall see later on, it is sometimes more convenient to use the integrated distribution

$$G(u) = \int_{-\infty}^{u} g(x)\,dx, \qquad g(u) = dG/du.$$

Since the pdf $g(u)$ is always non-negative and normalized so that its integral over all u is one, it follows that $G(u)$ is a monotonously non-decreasing function taking values from zero to one.

The expectation E of a function $f(u)$ is defined as the average or mean value of the function

$$r = E(f) \equiv \int_{-\infty}^{\infty} f(u)\,g(u)\,du = \int_0^1 f(u)\,dG, \qquad (2.1)$$

where $G(u)$ is the cumulative distribution function of u'. In a Monte Carlo integration procedure it defines the pdf from which a random variable is sampled.

The special case in which u' is uniformly distributed between a and b results in $dG(x) = dx/(b-a)$, leading to the expectation:

$$E(f) = \frac{1}{b-a}\int_a^b f(x)\,dx.$$

The *variance* V of the function f is defined as the average of the squared deviation from its expectation:

$$V(f) \equiv E[f - E(f)]^2 = \int_0^1 [f - E(f)]^2\,dG. \qquad (2.2)$$

Any product $h = f \cdot g$ (of an arbitrary function f and a pdf g with the properties defined above) resulting in the same function h has the same expectation (2.1). This does not hold for the variance. It depends on f and g and is a measure for the convergence of a Monte Carlo integration procedure.

3. CORRELATED SAMPLING

3.1. *Basic Considerations*

The function f of the previous section may depend on the variables x_i and on a set of parameters p_j. Quite frequently it is of interest to provide Mont Carlo estimates of (2.1) which show the effect of a parameter variation Δp_j. If such a parameter change is small, straightforward Monte Carlo may fail, as will be shown and special techniques must be employed.

In the following for the sake of brevity the mathematics is carried out for one Δp_j only. For this reason the index j is dropped as well as the index i in the integration variables. The integration space is $B = [0, 1]$.

In terms of classical Monte Carlo estimates the two integrals to be compared are calculated as

$$E(f^*) = \int_B f(x; p+\Delta p)\,dG^* \qquad (3.1a)$$

and

$$E(f) = \int_B f(x; p)\,dG, \qquad (3.1b)$$

where $f^* \equiv f(x; p+\Delta p)$, $f \equiv f(x; p)$ and

$$dG \equiv g(x; p)\,dx \qquad (3.2a)$$

$$dG^* \equiv g(x; p+\Delta p)\,dx. \qquad (3.2b)$$

The expectation of the difference $\Delta f \equiv f^* - f$ is

$$E(\Delta f) \equiv E(f^*) - E(f), \qquad (3.3)$$

and the sample variance of Δf,

$$\begin{aligned} V(\Delta f) &\equiv E[(f^* - f) - E(f^* - f)]^2 \\ &= E\{[f^* - E(f^*)][f - E(f)]^2\} \\ &= [V(f^*) + V(f)] - 2E\{[f^* - E(f^*)] \\ &\quad \times [f - E(f)]\}, \end{aligned} \qquad (3.4)$$

where

$$V(f^*) = \int_B [f^* - E(f^*)]^2\,dG^*,$$

$$V(f) = \int_B [f - E(f)]^2\,dG.$$

In case the two estimates are obtained by two independent Monte Carlo runs, the correlation term $2E\{\ldots\}$ of (3.4) vanishes, so that the variance $V(\Delta f)$ is just the sum of the two variances $V(f^*)$ and $V(f)$,

$$V(\Delta f) = [V(f^*) + V(f)]. \qquad (3.5)$$

For small changes Δp the estimate is usually expressed by the *relative change* (differential quotient) $\Delta f/\Delta p$. As the sum in (3.5) is positive and finite, it is now easy to show that for

small parameter changes, Δp, the relative variance can increase without limit:

$$\lim_{\Delta p \to 0} V(\Delta f/\Delta p) = [V(f^*) + V(f)]/\Delta p^2 \to \infty. \quad (3.6)$$

If, on the other hand, it is possible to correlate the calculation of $E(f^*)$ and $E(f)$ then the relative variance may assume a finite value:

$$\lim_{\Delta p \to 0} V(\Delta f/\Delta p) = \frac{V(f^*) + V(f) - 2[E(f^*f) - E(f^*)\,E(f)]}{\Delta p^2} \to c < \infty. \quad (3.7)$$

Proof. One possibility of correlating the Monte Carlo integration of $f(x; p+\Delta p)\,dG^*$ with $f(x; p)\,dG$ (Eqs. (3.1)) consists in expressing dG^* in terms of dG by the use of Eq. (3.2). It leads to

$$dG^* = g(x; p+\Delta p)\frac{g(x; p)\,dx}{g(x; p)} = \frac{g(x; p+\Delta p)}{g(x; p)}\,dG, \quad (3.8)$$

a so-called "likelihood ratio." Formally this procedure is similar to *importance sampling*. Its restrictions are valid in this context too. With (3.8) inserted in (3.1) and then, together with (3.2) in 3.3, we obtain

$$E(\Delta f) = \int_B \left[f(x; p+\Delta p)\frac{g(x; p+\Delta p)}{g(x; p)} - f(x; p) \right] dG. \quad (3.9)$$

Note that the two functions $f(x; p+\Delta p)$ and $f(x; p)$ are now under the same integral and random variables are taken only from dG. This changes the Monte Carlo integration procedure to

$$\hat{r}^* = \frac{1}{N}\sum_n f(x_n; p+\Delta p)\frac{g(x_n; p+\Delta p)}{g(x_n; p)} \quad (3.10a)$$

and

$$\hat{r} = \frac{1}{N}\sum_n f(x_n; p), \quad (3.10b)$$

where the x_n's are sampled from $g(x; p)$ and inserted simultaneously in both Eqs. (3.10a) and (3.10b). The total number of samples is N, while $\hat{r}^*$ and $\hat{r}$ are sample means approaching $E(f^*)$ and $E(f)$ for large Ns.

To estimate the variance of the differential quotient $\Delta f/\Delta p$ for small values of Δp, we develop in (3.9) the product $h(x; p+\Delta p) = f(x; p+\Delta p)\,g(x; p+\Delta p)$ in a *first-order* Taylor series in p,

$$h(x; p+\Delta p) \approx h(x; p) + \Delta p \frac{\partial}{\partial p} h(x; p) \quad (3.11a)$$

$$= f(x; p)\,g(x; p) + \Delta p \times \left[\frac{\partial f(x; p)}{\partial p} g(x; p) + f(x; p)\frac{\partial g(x; p)}{\partial p} \right]. \quad (3.11b)$$

Using Eq. (3.11a) the integrand of Eq. (3.9) becomes

$$f(x; p+\Delta p)\frac{g(x; p+\Delta p)}{g(x; p)} - f(x; p) \approx \Delta p \cdot f(x; p)\left[\frac{1}{g(x; p)}\frac{\partial}{\partial p} g(x; p) + \frac{1}{f(x; p)}\frac{\partial}{\partial p} f(x; p) \right].$$

Inserting Eq. (3.11b) into Eq. (3.9) we obtain (after dividing Δf by Δp and reorganizing the formula)

$$E(\Delta f/\Delta p) = \int_B \left[\frac{\partial}{\partial p}\ln(f(x; p)) + \frac{\partial}{\partial p}\ln(g(x; p)) \right] \times f(x; p)\,dG. \quad (3.12)$$

The variance of $\Delta f/\Delta p$ can now be expressed as

$$V(\Delta f/\Delta p) = \int_B \left[\frac{\partial}{\partial p} f(x; p) + f(x; p)\frac{\partial}{\partial p}\ln[g(x; p)] \right]^2 dG - [E(\Delta f/\Delta p)]^2. \quad (3.13)$$

This expression proves that a correlated sampling procedure has indeed a finite relative variance for $\Delta p \to 0$, as long as f and g and their *derivatives* remain finite within the integration interval. Experience has shown that this condition is satisfied in many practical cases. In situations where a finite integral is not guaranteed, it is possible to transform the problem as discussed below.

If only one of the functions $f(..)$ or $g(..)$ depends upon p then Eq. (3.9) becomes either

$$E(\Delta f) = \int_B \left[\frac{g(x; p+\Delta p)}{g(x; p)} - 1 \right] f(x)\,dG \quad (3.14)$$

or

$$E(\Delta f) = \int_B \left[\frac{f(x; p+\Delta p)}{f(x; p)} - 1 \right] f(x, p)\,dG. \quad (3.15)$$

For both special cases a finite relative variance depends again on the condition that the derivatives of f or g be finite in the whole integration space.

In this perturbation scheme it is an essential requirement that the likelihood ratio $g(x; p+\Delta p)/g(x; p)$ in Eq. (3.9) remains finite in the whole integration space. Possibly it should be close to unity to keep the variance of the differential effect small.

In cases where these conditions cannot be satisfied the author proposes a technique called *multiple correlation.* It consists in the construction of a *reference function* $g_0(x; p)$ from which the random variables are sampled. The reference function g_0 must be finite in the whole integration domain of both $g(x; p)$ and $g(x; p+\Delta p)$ so that

$$\frac{g(x; p)}{g_0(x; p)} < \infty, \qquad \frac{g(x; p+\Delta p)}{g_0(x; p)} < \infty.$$

We define a new expectation,

$$E(f_0) = \int_B f(x; p)\, dG_0,$$

where $\varphi(x; p) = f(x; p)$ and $dG_0 = g_0(x; p)\, dx$. It is subtracted from and added to Eq. (3.3):

$$[E(f^*) - E(\varphi)] + [E(\varphi) - E(f)] = E(\Delta f_\alpha) + E(\Delta f_\beta).$$

The uncorrelated estimate of Δf_α is

$$E(\Delta f_\alpha) = E(f^*) - E(\varphi) = \int_B f(x; p+\Delta p)\, dG^* - \int_B f(x; p)\, dG_0.$$

Similar to (3.8), it follows that

$$dG^* = \frac{g(x; p+\Delta p)}{g_0(x; p)}\, dG_0$$

and in analogy with (3.9), the correlated estimate becomes

$$E(\Delta f_\alpha) = \int_B \left[f(x; p+\Delta p)\, \frac{g(x; p+\Delta p)}{g_0(x; p)} - f(x; p) \right] dG_0.$$

In the same manner Δf_β can be evaluated:

$$E(\Delta f_\beta) = E(\varphi) - E(f) = \int_B f(x; p)\, dG_0 - \int_B f(x; p)\, dG.$$

Replacing dG by dG_0 multiplied by the likelihood ratio g/g_0,

$$dG = \frac{g(x; p)}{g_0(x; p)}\, dG_0,$$

the estimate of the correlated Δf_β becomes

$$E(\Delta f_\beta) = \int_B \left[f(x; p) - f(x; p)\, \frac{g(x; p)}{g_0(x; p)} \right] dG_0.$$

Adding up the two estimates of Δf_α and Δf_β results in

$$E(\Delta f) = \int_B \left[f(x; p+\Delta p)\, \frac{g(x; p+\Delta p)}{g_0(x; p)} - f(x; p)\, \frac{g(x; p)}{g_0(x; p)} \right] dG_0. \tag{3.16}$$

The function $g_0(x; p)$ can be chosen freely, but it will depend to a large extend upon the problem to be solved; g_0 determines the variance of the procedure and may therefore be the subject of optimisation. The general conditions for g_0 are:

— g_0 is a probability density function satisfying the condition $g_0(x; p) > 0$ for all x,

— $G_0(x; p)$, the integral of $g_0(x; p)$ is known analytically and increases monotonically from zero to one,

— either function $G_0(x; p)$ can be inverted analytically or a g_0-distributed random variable is available,

— both ratios $g(x; p+\Delta p)/g_0(x; p)$ and $g(x; p)/g_0(x; p)$ should be as constant as possible for the whole integration space.

Typical, but not necessarily optimized, examples for choosing g_0 are the *arithmetic mean*

$$g_0(x; p) = [g(x; p+\Delta p) + g(x; p)]/2 \tag{3.17}$$

or the *maximum* of either function g^* and g:

$$g_0(x; p) = \frac{\max[g(x; p+\Delta p),\, g(x; p)]}{\int_B dx \max[g(x; p+\Delta p),\, g(x; p)]}. \tag{3.18}$$

This latter reference function looks particularly attractive, since it leads to likelihood ratios close to unity. It can easily be generated if piecewise constant functions such as probability tables are used. A specific example of this function is described in Section 5.4.

4. DIFFERENTIAL OPERATOR SAMPLING

4.1. *Basic Considerations*

It is the scope of differential Monte Carlo to calculate linear and, if required by the problem, higher order derivatives. These derivatives of the target quantities refer, as in the case of correlated tracking, to the parameters **p**.

The derivative of the expectation will be formulated for the general case, in which the mean-value of $f(\mathbf{x};\mathbf{p})$ is calculated for a variate with the pdf $g(\mathbf{x};\mathbf{p})$:

$$r = E(f) = \int_{-\infty}^{\infty} f(\mathbf{x};\mathbf{p}) \cdot g(\mathbf{x};\mathbf{p})\, dx = \int_0^1 f(\mathbf{x};\mathbf{p})\, d\mathbf{G}. \tag{4.1}$$

For the derivative with respect to p_i it follows:

$$\frac{\partial r}{\partial p_i} = \frac{\partial}{\partial p} E(f) = \int_B \frac{\partial}{\partial p_i} f(\mathbf{x};\mathbf{p})\, d\mathbf{G} + \int_B f(\mathbf{x};\mathbf{p})\, d\left(\frac{\partial \mathbf{G}}{\partial p_i}\right). \tag{4.2}$$

This expression will now be interpreted in terms of a (stochastic) sampling procedure. In a first attempt one would probably intend to integrate the two terms of the right-hand side independently. But in a simulation analysis of a complicated multidimensional problem (for which Monte Carlo is actually made) it would lead to a substantial increase of the programming effort and computer time. In most cases it would, for these reasons, clearly not be practical.

An easy transformation makes it, however, possible to express the estimate of the derivative in a form which facilitates the sampling procedure. To this end we replace $\partial \mathbf{G}/\partial p_i$ by

$$\frac{\partial G}{\partial p_i} = \frac{1}{g(\mathbf{x};\mathbf{p})} \frac{\partial G}{\partial p_i} g(\mathbf{x};\mathbf{p}) = \frac{1}{g(\mathbf{x};\mathbf{p})} \frac{\partial}{\partial p_i} g(\mathbf{x};\mathbf{p})\, g(\mathbf{x};\mathbf{p})\, dx = \frac{\partial}{\partial p_i} \ln(g(\mathbf{x};\mathbf{p}))\, d\mathbf{G},$$

and introduce this expression into (4.2):

$$\frac{\partial r}{\partial p_i} = \int_B \left(\frac{\partial}{\partial p_i} \ln[f(\mathbf{x},\mathbf{p})] + \frac{\partial}{\partial p_i} \ln[g(\mathbf{x},\mathbf{p})] \right) f(\mathbf{x},\mathbf{p})\, d\mathbf{G}. \tag{4.3}$$

The advantage of this transformation is the fact that the term $f(\mathbf{x};\mathbf{p})$ appears explicitly in the formula. This allows for the simultaneous calculation of the integral (4.1) and its derivative (4.3). In the scoring procedure it is only necessary to multiply the function value of $f(x;p)$ by the function value of the derivatives in the parenthesis. Furthermore, it follows that for sampling the derivatives the same variates are used as for the calculations of the original response r. A Monte Carlo estimate of $\partial \hat{r}/\partial p_i$ is then calculated as

$$\frac{\partial \hat{r}}{\partial p_i} = \frac{1}{N} \sum_{\nu=1}^{N} \left(\frac{\partial}{\partial p_i} \ln[f_\nu(\mathbf{x},\mathbf{p})] + \frac{\partial}{\partial p_i} \ln[g_\nu(\mathbf{x},\mathbf{p})] \right) f_\nu(\mathbf{x},\mathbf{p}),$$

where f_ν stands for the value of the function f and $\partial \ln[f_\nu(x;p)]/\partial p_i$ for the value of the derivative in the νth trial.

5. PERTURBATION ANALYSIS OF LINEAR MATRIX EQUATIONS

5.1. *Solutions of Linear Matrix Equations by Random Walk Techniques*

A set of linear equations

$$\mathbf{r} = \mathbf{P}\mathbf{r} + \mathbf{a} \tag{5.1}$$

of order m can be interpreted as a random walk process if the matrix **P** describes transition probabilities in a system which can assume a set of states from one to m. If the probability for transition from state i to the next state j does not depend on the past; i.e., on the path by which j was reached, the system is said to be Markovian. A typical Markov chain will start in some state l_0 with subsequent transitions through a sequence of states $l_1, l_2, \ldots$ and finally a termination in state l_k yielding a series of integers $\gamma_k = (l_0, l_1, l_2, \ldots, l_k)$, where $l_n \in \{1, 2, \ldots, m\}$. In Monte Carlo terminology such a chain is often called a *history*.

After a start with probability

$$P(l_0 = i) = a_i,$$

the successive states are connected by the transition probabilities

$$P(l_{n+1} = i \mid l_n = j, n < k) = p_{ij},$$

and the termination probability

$$P(k = n \mid l_n = i) = p_i.$$

The $p_{i \leftarrow j}$s being elements of the matrix **P** (dimension

$m \times m$) are the transition probabilities from state j to i, for which the following conditions must be satisfied:

$$p_{ij} \geqslant 0, \qquad \max_j \left(\sum_i p_{ij} \right) \leqslant 1,$$
$$\sum_i a_i = 1, \qquad p_j = 1 - \sum_i p_{ij} \leqslant 1. \tag{5.2}$$

The last condition is the terminating criterion for a chain of events. There must be at least one p_j for which $p_j < 1$ is met. Furthermore, only matrices $\mathbf{P}$ with spectral radius $\rho(\mathbf{P}) < 1$ are allowed. (The *spectral radius* of the matrix $\mathbf{P}$ is the largest modulus of all eigenvalues of $\mathbf{P}$. Thus it is the radius of the smallest circel in the complex plane which contains all eigenvalues of $\mathbf{P}$.)

For each state i we define a variable r_i which is the average number of entries into state i. The state i can be reached either by an event starting in i with probability a_i or by a transition to i from state j with probability $p_{i \leftarrow j}$. The mean number of transitions from j to i corresponds to the product $p_{i \leftarrow j} r_j$.

Note. The solution of this system of linear equations by the random walk algorithm described here requires that the p_{ij} are interpreted as transfer probabilities from j to i leading to the "backward" notation $p_{i \leftarrow j}$.

All events in state i are the sum

$$r_i = p_{i \leftarrow 1} r_1 + p_{i \leftarrow 2} r_2 + \cdots + p_{i \leftarrow m} r_m + a_i.$$

The expected number of transitions from all states to all other states are then taken into account by the matrix Eq. (5.1).

The solution of this system of m linear equations by a random walk procedure requires the following additional considerations. If $P_i^{(n)}$ denotes the probability that the nth transition (in the sequence) occurs to state i

$$P_i^{(n)} = P(l_n = i \mid n - 1 < k),$$

a recursion expression for $P_i^{(n)}$ can be formulated

$$P_i^{(n)} = \sum_j p_{ij} P_j^{(n-1)} \qquad \text{for} \quad n \geqslant 1, \qquad \mathbf{P}_i^{(0)} = \mathbf{P}(l_0 = i) = a_i.$$

It stands for the probability that transition n enters state i, provided $P_j^{(n-1)}$ is the probability that the $(n-1)$th transition occurs in j summed over all intermediate states j.

With this recursion the number of entries into state i denoted by r_i can now be interpreted as being the expectation $\langle x_i \rangle$, where

$$\langle x_i \rangle = a_i + \sum_{l_1=1}^{m} p_{il_1} a_{l_1} + \sum_{l_2=1}^{m} \sum_{l_1=1}^{m} p_{il_2} p_{l_2 l_1} a_{l_1} + \cdots \tag{5.3}$$
$$= a_i + \sum_{k=1}^{\infty} \sum_{l_k} \cdots \sum_{l_1} p_{il_k} \cdots p_{l_2 l_1} a_{l_1} \tag{5.4}$$
$$= a_i + (Pa)_i + (P^2 a)_i + \cdots$$
(corresponds to the Neumann series!)
$$= a_i + P[a_i + (Pa)_i + (P^2 a)_i \cdots]$$
$$= a_i + P\langle x_i \rangle. \tag{5.5}$$

Written in matrix notation,

$$\langle \mathbf{x} \rangle = \mathbf{P}\langle \mathbf{x} \rangle + \mathbf{a} \tag{5.6}$$

is equivalent to Eq. (5.1).

The *sample mean* $\hat{x}_i$ of all chains of events (histories) $\gamma_k = (l_0, l_1, l_2, \ldots, l_k)$ is the normalized sum of all entries into state i:

$$\hat{x}_i = \frac{1}{H} \sum_{h=1}^{H} \sum_{n=0}^{k(h)} \delta_{il_n}, \tag{5.7}$$

with Kronecker's symbol $\delta_{ij} = 0$ for $i \neq j$ and $\delta_{ij} = 1$ for $i = j$. H is the total number of histories and $k(h) < \infty$ the length of the transition chain γ in the hth history.

The estimates of $\hat{x}_i$ converge to the exact solution of the matrix equation as the number of histories increases (see Fig. 5.1).

5.2. *Perturbation Algorithms Solving Linear Matrix Equations by Random Walk Techniques*

It has been shown [3 p. 85] that a matrix equation

$$\mathbf{x} = \mathbf{Q}\mathbf{x} + \mathbf{b}, \tag{5.8}$$

satisfying less stringent conditions than those required by (5.2) may be solved by "*mapping*" it onto a random walk as described above. A correlation technique which makes this possible, is also the base for the perturbation analysis of *linear matrix equations* discussed in this chapter.

Using the same notation as previously, we define two likelihood ratios

$$\alpha_i = b_i / a_i, \qquad \pi_{ij} = q_{ij} / p_{ij} \qquad \text{requiring} \quad \begin{cases} \neq 0 & \text{for} \quad p_{ij} \neq 0 \\ = 0 & \text{for} \quad p_{ij} = 0, \end{cases}$$

and

$$\|Q\| = \max_j \left(\sum_i |q_{ij}| \right) < 1.$$

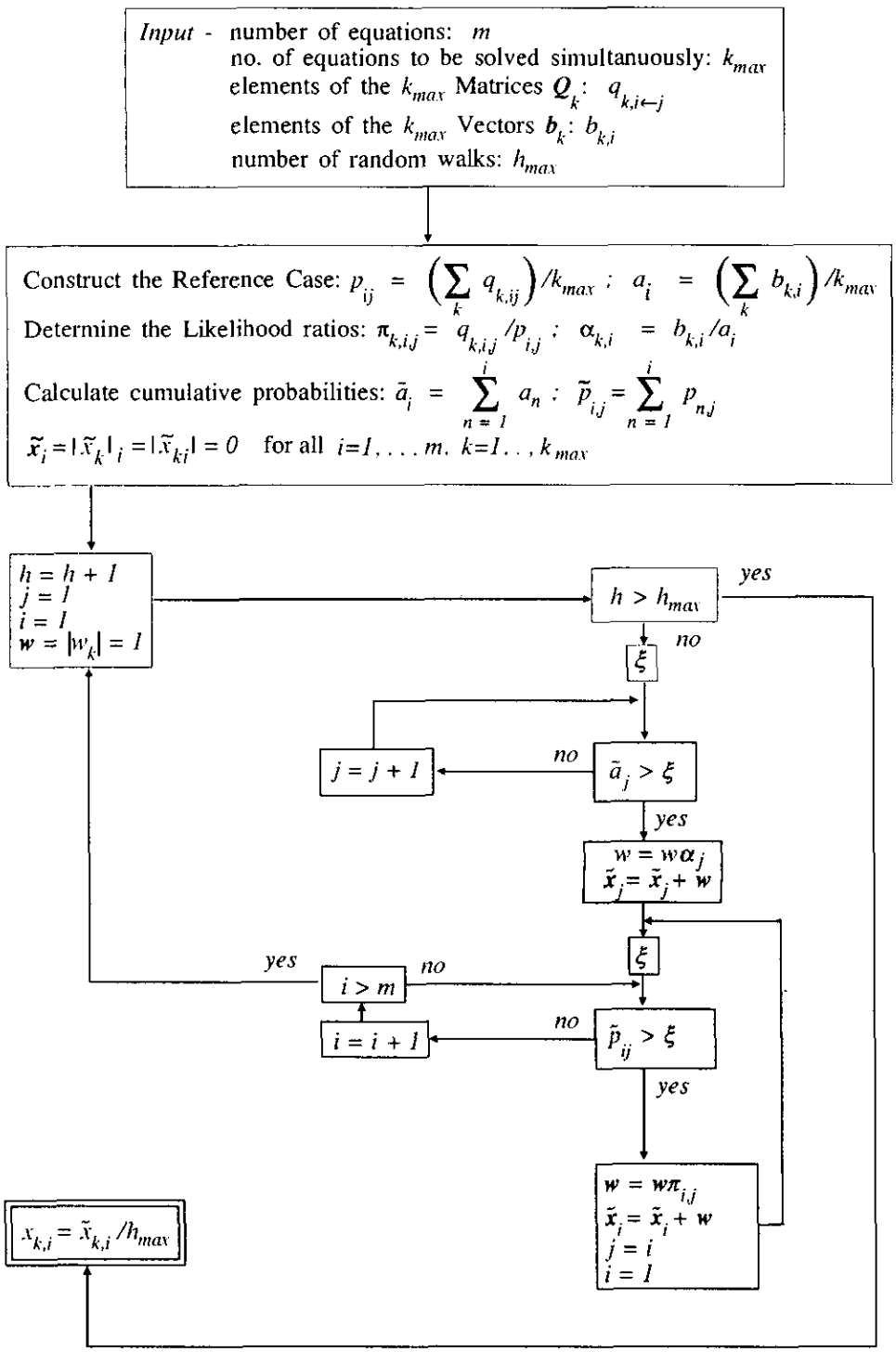

FIG. 5.1. Flow-chart for the simultanuous solution of systems of linear matrix equations by the *multiple correlation* algorithm. In constructing the reference case it is assumed that all $q_{k,i,j}$ are positive.

To solve Eq. (5.8) we again follow random walks $\gamma_k = (l_0, l_1, l_2, ..., l_n, ..., l_k)$ connected by the transition probabilities p_{ij}. But in this case the random walk is accompanied by a weight factor w_n being the product of all π_{ij} associated with the states it passed through before entering state n,

$$w_n = w_{n-1}\pi_{l_n l_{n-1}} = \pi_{l_n l_{n-1}} \cdots \pi_{l_3 l_2}\pi_{l_2 l_1}\alpha_{l_1} \qquad (w_0 \equiv \alpha_{l_1}).$$

Instead of just counting the transition events as in (5.7) the score of x_i up to the kth transition is the sum of thoose weight factors which refere to transitions into state i. It leads to the sample mean

$$\hat{x}_i = \frac{1}{H}\sum_{h=1}^{H}\sum_{n=0}^{k(h)} w_n \cdot \delta_{il_n}.$$

To calculate the second moment, the score of x_i must be summed up and squared for each history separately,

$$\widehat{x_i^2} = \frac{1}{H}\sum_{h=1}^{H}\left(\sum_{n=0}^{k(h)} w_{l_n}\delta_{il_n}\right)^2, \tag{5.9}$$

rendering the sample variance

$$\mathrm{var}(\hat{x}_i) = \widehat{x_i^2} - \hat{x}_i^2, \tag{5.10}$$

To facilitate the reading of the formulae we adopted the convention of indicating pdfs from which samples are taken by Roman letters (*italic*) and likelihood ratios (determining weight factors) by Greek letters.

The frequency with which these weight factors are chosen corresponds to the random walk distribution governed by the matrix **P** and the start distribution **a**. In analogy to expression 5.3 the expectation of x_i for all chains of events starting with the distribution $a_1, a_2, ..., a_m$ and reaching the final state k is

$$\begin{aligned}\langle x_i\rangle &= \alpha_i a_i + \sum_{l_1=1}^{m}\pi_{il_1}\alpha_{l_1}p_{il_1}a_{l_1} \\ &\quad + \sum_{l_2=1}^{m}\sum_{l_1=1}^{m}\pi_{il_2}\pi_{l_2l_1}\alpha_{l_1}p_{il_2}p_{l_2l_1}a_{l_1} + \cdots \\ &= \alpha_i a_i + \sum_{k=1}^{\infty}\sum_{l_k}\cdots\sum_{l_1}\pi_{il_k}\cdots \\ &\quad \pi_{l_2l_1}\alpha_{l_1}\cdot p_{il_k}\cdots p_{l_2l_1}a_{l_1}.\end{aligned} \tag{5.11}$$

In perturbation analysis the interest is focused on the difference of two (or more) estimates $\hat{x}_i^*$ and $\hat{x}_i$ being solutions of the two matrix equations

$$\hat{\mathbf{x}}^* = \mathbf{Q}\hat{\mathbf{x}}^* + \mathbf{b}^* \qquad \text{and} \qquad \hat{\mathbf{x}} = \mathbf{Q}\hat{\mathbf{x}} + \mathbf{b}, \tag{5.12}$$

which are of the same dimension and may differ only slightly from each other.

A straightforward approach would be to perform two independent calculations with $\mathbf{Q}^*$, $\mathbf{b}^*$ and $\mathbf{Q}$, $\mathbf{b}$, where $\mathbf{Q}^* = \mathbf{Q} + \Delta\mathbf{Q}$ and $\mathbf{b}^* = \mathbf{b} + \Delta\mathbf{b}$. In the case of two independent and uncorrelated computer runs the variance for the difference becomes

$$V(\Delta\mathbf{x}) = V(\mathbf{x}^*) + V(\mathbf{x}).$$

As $V(\mathbf{x}^*)$ and $V(\mathbf{x})$ are always positive quantities, the relative variance $V(\Delta\mathbf{x})/|\Delta\mathbf{Q}|$ is unbounded as $|\Delta\mathbf{Q}| \to 0$, similar to the considerations of (3.1).

Applying a correlation technique it is possible to avoid this shortcoming and to obtain a finite variance even for very small changes of the matrix **Q** and/or the start distribution **b**. Correlation is achieved by mapping both systems $\mathbf{Q}^*$, $\mathbf{b}^*$ and **Q**, **b** onto the same random walk process which we call the *reference system*. It is governed by a transition matrix **P** and a vector **a**.

The *reference system* can be chosen arbitrarily, provided it satisfies the condition (5.2). If all $q_{ij}^* \geqslant 0$ and $q_{ij} \geqslant 0$ it can be constructed as the arithmetic mean of two (or more) systems $\mathbf{Q}^*$, $\mathbf{b}^*$ and **Q**, **b**:

$$p_{ij} = (q_{ij}^* + q_{ij})/2, \tag{5.13}$$

or

$$p_{ij} = \frac{\max(|q_{ij}^*|, |q_{ij}|)}{\sum_i \max(|q_{ij}^*|, |q_{ij}|)}\sum_i \frac{1}{2}(|q_{ij}^*| + |q_{ij}|), \tag{5.14}$$

The *reference system* can also be identical with one of the systems $\mathbf{Q}^*$, $\mathbf{b}^*$ or **Q**, **b**.

The *likelihood ratios* determining the solution of Eqs. (5.11) are

$$\begin{aligned}\pi_{ij}^* &= q_{ij}^*/p_{ij}, \qquad \alpha_i^* = b_i^*/a_i \\ \pi_{ij} &= q_{ij}/p_{ij}, \qquad \alpha_i = b_i/a_i.\end{aligned}$$

As in (3.2) these ratios should be as close to unity as possible.

With this assumption both systems follow the same random walk, one with the product of weights

$$w_n^* = w_{n-1}^*\pi_{l_n l_{n-1}}^* = \pi_{l_n l_{n-1}}^*\pi_{l_{n-1}l_{n-2}}^* \cdots \pi_{l_2 l_1}^*\alpha_{l_1}^* = \prod_{i=1}^{n}\pi_{l_i l_{i-1}}^*, \tag{5.15a}$$

and the other with

$$w_n = w_{n-1}\pi_{l_n l_{n-1}} = \pi_{l_n l_{n-1}}\pi_{l_{n-1}l_{n-2}} \cdots \pi_{l_2 l_1}\alpha_{l_1} = \prod_{i=1}^{n}\pi_{l_i l_{i-1}}, \tag{5.15b}$$

where $\pi_{l_1 l_0} \equiv \alpha_{l_1}$ and $\pi_{l_1 l_0}^* \equiv \alpha_{l_1}^*$.

To simplify the further considerations we assume that in $\mathbf{Q}^*$ only one element $\iota\eta$,

$$q^*_{\iota\eta} = q_{\iota\eta} + \Delta q_{\iota\eta},$$

is different from $\mathbf{Q}$ and that $\mathbf{Q}$, $\mathbf{b}$ serves as the reference system. With these assumptions all matrix elements of $\mathbf{P}$ are $p_{ij} = q_{ij}$ and all likelihood ratios π_{ij} and a_j are 1 except

$$\pi^*_{\iota\eta} = \frac{q_{\iota\eta} + \Delta q_{\iota\eta}}{p_{\iota\eta}} = 1 + \frac{\Delta q_{\iota\eta}}{q_{\iota\eta}} = 1 + \Delta\rho_{\iota\eta}, \tag{5.16}$$

where $\Delta\rho_{\iota\eta}$ is the relative change of the element $q_{\iota\eta}$.

The score of the difference Δx_i in a *single* history (chain of events γ_k) is

$$\begin{aligned}\Delta\hat{x}_i(\gamma_k) = \sum_{n=1}^{k} \Bigg[\prod_{\nu=1}^{n} (1 + \Delta\rho_{\iota\eta}\,\delta_{\iota l_\nu}\,\delta_{\eta l_{\nu-1}})\,\pi_{l_\nu l_{\nu-1}} \\ - \prod_{\nu=1}^{n} \pi_{l_\nu l_{\nu-1}} \Bigg] \delta_{il_n}, \\ = \sum_{n=1}^{k} \Bigg\{ \Bigg[\prod_{\nu=1}^{n} (1 + \Delta\rho_{\iota\eta}\,\delta_{\iota l_\nu}\,\delta_{\eta l_{\nu-1}}) - 1 \Bigg] \\ \times \prod_{\nu=1}^{n} \pi_{l_\nu l_{\nu-1}} \Bigg\} \delta_{il_n}. \end{aligned} \tag{5.17}$$

If $\Delta\rho_{\iota\eta} \ll 1$ the following approximation is possible:

$$\prod_{\nu} (1 + \Delta\rho_{\iota\eta}\,\delta_{\iota l_\nu}\,\delta_{\eta l_{\nu-1}}) \approx 1 + \sum_{\nu} \Delta\boldsymbol{\rho}_{\iota\eta}\,\boldsymbol{\delta}_{\iota l_\nu}\,\boldsymbol{\delta}_{\eta l_{\nu-1}}. \tag{5.18}$$

Introduced into Eq. (5.17) one has

$$\Delta\hat{x}_i(\gamma_k) = \Delta\rho_{\iota\eta} \sum_{n=1}^{k} \Bigg[\Bigg(\sum_{\nu=1}^{n} \delta_{\iota l_\nu}\,\delta_{\eta l_{\nu-1}} \Bigg) \prod_{\nu=1}^{n} \pi_{l_\nu l_{\nu-1}} \Bigg] \delta_{il_n}. \tag{5.19}$$

After moving $\Delta\rho_{\iota\eta}$ to the left-hand side and summing over all histories H (where $k(h)$ is the total number of transitions in the hth history) the sample mean (Monte Carlo estimate) of $\Delta x_i/\Delta\rho_{\iota\eta}$ becomes

$$\lim_{\Delta\rho\to 0} \frac{\Delta\hat{x}_i}{\Delta\rho_{\iota\eta}} = \frac{1}{H} \sum_{h=1}^{H} \sum_{n=1}^{k(h)} \Bigg[\Bigg(\sum_{\nu=1}^{n} \delta_{\iota l_\nu}\,\delta_{\eta l_{\nu-1}} \Bigg) \prod_{\nu=1}^{n} \pi_{l_\nu l_{\nu-1}} \Bigg] \delta_{il_n}, \tag{5.20}$$

and in analogy to (5.9) its second moment is

$$\widehat{\Delta x_i^2} = \frac{1}{H} \sum_{h=1}^{H} \Bigg\{ \Delta\rho_{\iota\eta} \sum_{n=1}^{k(h)} \Bigg[\Bigg(\sum_{\nu=1}^{n} \delta_{\iota l_\nu}\,\delta_{\eta l_{\nu-1}} \Bigg) \prod_{\nu=1}^{n} \pi_{l_\nu l_{\nu-1}} \Bigg] \delta_{il_n} \Bigg\}^2.$$

With the assumption that in Eq. (5.20) all $\pi_{ij} = 1$ and all $\alpha_j = 1$ the variance of the relative perturbation is

$$\begin{aligned}\lim_{\Delta\rho\to 0} \frac{\operatorname{var}(\Delta\hat{x}_i)}{\Delta\rho^2_{\iota\eta}} = \frac{1}{H} \sum_{h=1}^{H} \Bigg(\sum_{n=1}^{k(h)} \sum_{\nu=1}^{n} \delta_{\iota l_\nu}\,\delta_{\eta l_{\nu-1}} \Bigg)^2 \\ - \Bigg(\frac{1}{H} \sum_{h=1}^{H} \sum_{n=1}^{k(h)} \sum_{\nu=1}^{n} \delta_{\iota l_\nu}\,\delta_{\eta l_{\nu-1}} \Bigg)^2. \end{aligned} \tag{5.22}$$

As this expression is finite for all terminating random walk procedures, it proves that the *multiple correlation technique*, proposed here, renders a bounded relative variance for arbitrarily small perturbations.

5.3. *Differential Operator Sampling: Gradients (Sensitivities) and Higher Order Derivatives*

Equation (5.20) yields another important result. The fact that

$$\frac{\partial\hat{x}_k}{\partial\rho_{ij}} = \lim_{\Delta\rho\to 0} \left(\frac{\Delta\hat{x}_k}{\Delta\rho_{ij}} \right), \qquad \text{where} \quad i, j, k \in \{1, ..., m\},$$

allows to calculate the elements of a *Jacobian* of the $\hat{x}_i$s with respect to all transition probabilities, q_{ij}s, by the unbiased random walk algorithm described above. A different and more general method of calculating simultaneously the $\hat{x}_i$s and the derivatives of the $\hat{x}_i$s is, however, shown in this section. To this end we express the expectation of the x_is of Eq. (5.8) by their Neumann series which describes the random walk procedure rendering the solution of the problem,

$$\langle x_i \rangle = b_i + \sum_{k=1}^{\infty} \sum_{l_k=1}^{m} \cdots \sum_{l_1=1}^{m} q_{il_k} \cdots q_{l_2 l_1} b_{l_1}. \tag{5.23a}$$

To simplify the further considerations this equation is rewritten in the form

$$\langle x_i \rangle = b_i + \sum_{k=1}^{\infty} \sum_{l_k} \cdots \sum_{l_1} \prod_{n=1}^{k+1} q_{l_n l_{n-1}}, \tag{5.23b}$$

where $q_{l_1 l_0} \equiv b_{l_1}$ and $q_{l_{k+1} l_k} \equiv q_{il_k}$.

Now we assume that μ_k transitions $\iota \leftarrow \eta$ were encountered in the chain $\gamma_k = (l_1, ..., l_k)$, where $\mu_k \in \{0, 1, 2, 3, ...\}$. The procedure of differentiating $\langle x_i \rangle$ with respect to the element $q_{\iota\eta}$ becomes more transparent if we take the $q_{\iota\eta}$s out of the product term so that only the qs remain for which $l_n \leftarrow l_{n-1} \neq \iota \leftarrow \eta$. This is formally expressed by the modified upper product boundary $k^* + l = k - \mu_k + 1$:

$$\langle x_i \rangle = b_i + \sum_{k=1}^{\infty} \sum_{l_k} \cdots \sum_{l_1} q_{\iota\eta}^{\mu_k} \prod_{n=1}^{k^*+1} q_{l_n l_{n-1}}.$$

Differentiation with respect to the element $q_{\iota\eta}$ leads to

$$\left\langle \frac{\partial x_i}{\partial q_{\iota\eta}} \right\rangle = \sum_{k=1}^{\infty} \sum_{l_k} \cdots \sum_{l_1} \mu_k q_{\iota\eta}^{\mu_k - 1} \prod_{n=1}^{k^*+1} q_{l_n l_{n-1}}. \quad (5.24)$$

Putting $q_{\iota\eta}^{\mu}$ back into the product this equation can be written as

$$\left\langle \frac{\partial x_i}{\partial q_{\iota\eta}} \right\rangle = \sum_{k=1}^{\infty} \sum_{l_k} \cdots \sum_{l_1} \frac{\mu_k}{q_{\iota\eta}} \prod_{n=1}^{k+1} q_{l_n l_{n-1}}. \quad (5.25)$$

Again, as in (5.2), we generalize the Markovian chain process by mapping it onto a random walk system governed by the matrix $\|p_{ij}\|$. Finally, with $\partial q/q = \partial\rho$ one obtains

$$\left\langle \frac{\partial x_i}{\partial \rho_{\iota\eta}} \right\rangle = \sum_{k=1}^{\infty} \sum_{l_k} \cdots \sum_{l_1} \mu_k \prod_{n=1}^{k+1} \pi_{l_n l_{n-1}} \prod_{n=1}^{k+1} p_{l_n l_{n-1}}, \quad (5.26)$$

where $p_{l_1 l_0} \equiv a_{l_1}$, $\pi_{l_1 l_0} \equiv \alpha_{l_1}$, $p_{l_{k+1} l_k} \equiv p_{\iota l_k}$, and $\pi_{l_{k+1} l_k} \equiv \pi_{\iota l_k}$.

This equation can be interpreted in terms of the same random walk process by which the expectation $\langle x_i \rangle$ is obtained. In the chain γ_k a counter μ_k must be increased by 1 each time a transition from η to ι occurs. Simultaneously the value of the counter is multiplied with the weight w_{k+1} (see (5.15)) and added to the score from which the sample mean of the derivative of x_i with respect to $q_{\iota\eta}$ is calculated:

$$\frac{\widehat{\partial x_i}}{\partial \rho_{\iota\eta}} = \frac{1}{H} \sum_{h=1}^{H} \sum_{n=1}^{k(h)} \left[\left(\sum_{\nu=1}^{n} \delta_{\iota l_\nu} \delta_{\eta l_{\nu-1}} \right) \times \prod_{\nu=1}^{n} \pi_{l_\nu l_{\nu-1}} \right] \delta_{\iota l_n}. \quad (5.27)$$

It says:

— if the first transition occurs from the source state $l_0 = \eta$ into $l_1 = \iota$ or from a later transition state $l_{\nu-1} = \eta$, $l_\nu = \iota$, and if at the same time $i = \iota$, then $1w_1$ is added to the score of $\partial \hat{x}_i / \partial q_{\iota\eta}$;

— if later on in the chain of events γ_k, a second transition from $l_{\nu-1} = \eta$ to $l_\nu = \iota$ occurs, and if at the same time $i = \iota$, then $2w_\nu$ is added to the score of $\partial \hat{x}_i / \partial q_{\iota\eta}$ an so on.

The expression is identical to (5.20).

Figure 5.2 shows a flow-chart for a Monte Carlo procedure by which the derivatives of all x_ks with respect to all π_{ij} (or p_{ij}) can be estimated. The results are listed in m two-dimensional matrices $\mathbf{G}_k = \|g_{ij}\|_k$, where $i, j, k = \{1, 2, ..., m\}$:

$$\frac{\partial \hat{x}_k}{\partial \rho_{ij}} = g_{kij}. \quad (5.28)$$

To obtain the *second-order derivative* expression (5.24) is again differentiated with respect to $q_{\iota\eta}$, leading to the expectation

$$\left\langle \frac{\partial^2 x_i}{\partial q_{\iota\eta}^2} \right\rangle = \sum_{k=1}^{\infty} \sum_{l_k} \cdots \sum_{l_1} \mu_k(\mu_k - 1)\, q_{\iota\eta}^{\mu_k - 2} \prod_{n=1}^{k^*+1} q_{l_n l_{n-1}}.$$

Replacing $\partial q^2/q^2$ by $\partial\rho^2$ and following the procedure described above we obtain

$$\left\langle \frac{\partial^2 x_i}{\partial \rho_{\iota\eta}^2} \right\rangle = \sum_{k=1}^{\infty} \sum_{l_k} \cdots \sum_{l_1} (\mu_k^2 - \mu_k) \prod_{n=1}^{k+1} \pi_{l_n l_{n-1}} \prod_{n=1}^{k+1} p_{l_n l_{n-1}} \quad (5.29)$$

and the corresponding sample mean

$$\frac{\widehat{\partial^2 x_i}}{\partial \rho_{\iota\eta}^2} = \frac{1}{H} \sum_{h=1}^{H} \sum_{n=1}^{k(h)} \left\{ \left[\left(\sum_{\nu=1}^{n} \delta_{\iota l_\nu} \delta_{\eta l_{\nu-1}} \right)^2 - \sum_{\nu=1}^{n} \delta_{\iota l_\nu} \delta_{\eta l_{\nu-1}} \right] \times \prod_{\nu=1}^{n} \pi_{l_\nu l_{\nu-1}} \right\} \delta_{\iota l_n}. \quad (5.30)$$

In a similar fashion the cross-terms can be determined if derivatives for two (or more) terms $q_{\iota\eta}$ and $q_{\kappa\lambda}$ are required:

$$\left\langle \frac{\partial^2 x_i}{\partial \rho_{\iota\eta}\, \partial \rho_{\kappa\lambda}} \right\rangle = \sum_{k=1}^{\infty} \sum_{l_k} \cdots \sum_{l_1} (\mu_k \nu_k) \prod_{n=1}^{k} \pi_{l_n l_{n-1}} \prod_{n=1}^{k} p_{l_n l_{n-1}}. \quad (5.31)$$

In this case the sample mean is

$$\frac{\widehat{\partial^2 x_i}}{\partial \rho_{\iota\eta}\, \partial \rho_{\kappa\lambda}} = \frac{1}{H} \sum_{h=1}^{H} \sum_{n=1}^{k(h)} \left\{ \left[\left(\sum_{\nu=1}^{n} \delta_{\iota l_\nu} \delta_{\eta l_{\nu-1}} \right) \times \left(\sum_{\nu=1}^{n} \delta_{\kappa l_\nu} \delta_{\lambda l_{\nu-1}} \right) \right] \prod_{\nu=1}^{n} \pi_{l_\nu l_{\nu-1}} \right\} \delta_{\iota l_n}. \quad (5.32)$$

5.4. *Examples*

To demonstrate the functioning of multiple correlation and differential operator sampling in 5.2 and 5.3, the author wrote two *Fortran* 77 programs following the schemes developed in Figs. 5.1 and 5.2. The programs use dynamic storage allocation so that the number of equations is only limited by the available memory space. Experience has shown that scores must be collected in double precision (64 bit) arrays. In the case of differential operator sampling the execution time can be substantially shortened if the score of the Jacobian is calculated by a vector processor.

EXAMPLE 1. A system of three linear equations. The random walk solution can easily be verified by a deterministic method such as *Cramer's rule*:

$$\begin{pmatrix} x_{1,1} \\ x_{1,2} \\ x_{1,3} \end{pmatrix} = \begin{pmatrix} 0.25, 0.50, 0.08 \\ 0.30, 0.30, 0.10 \\ 0.20, 0.20, 0.20 \end{pmatrix} \begin{pmatrix} x_{1,1} \\ x_{1,2} \\ x_{1,3} \end{pmatrix} + \begin{pmatrix} 0.30 \\ 0.30 \\ 0.40 \end{pmatrix}.$$

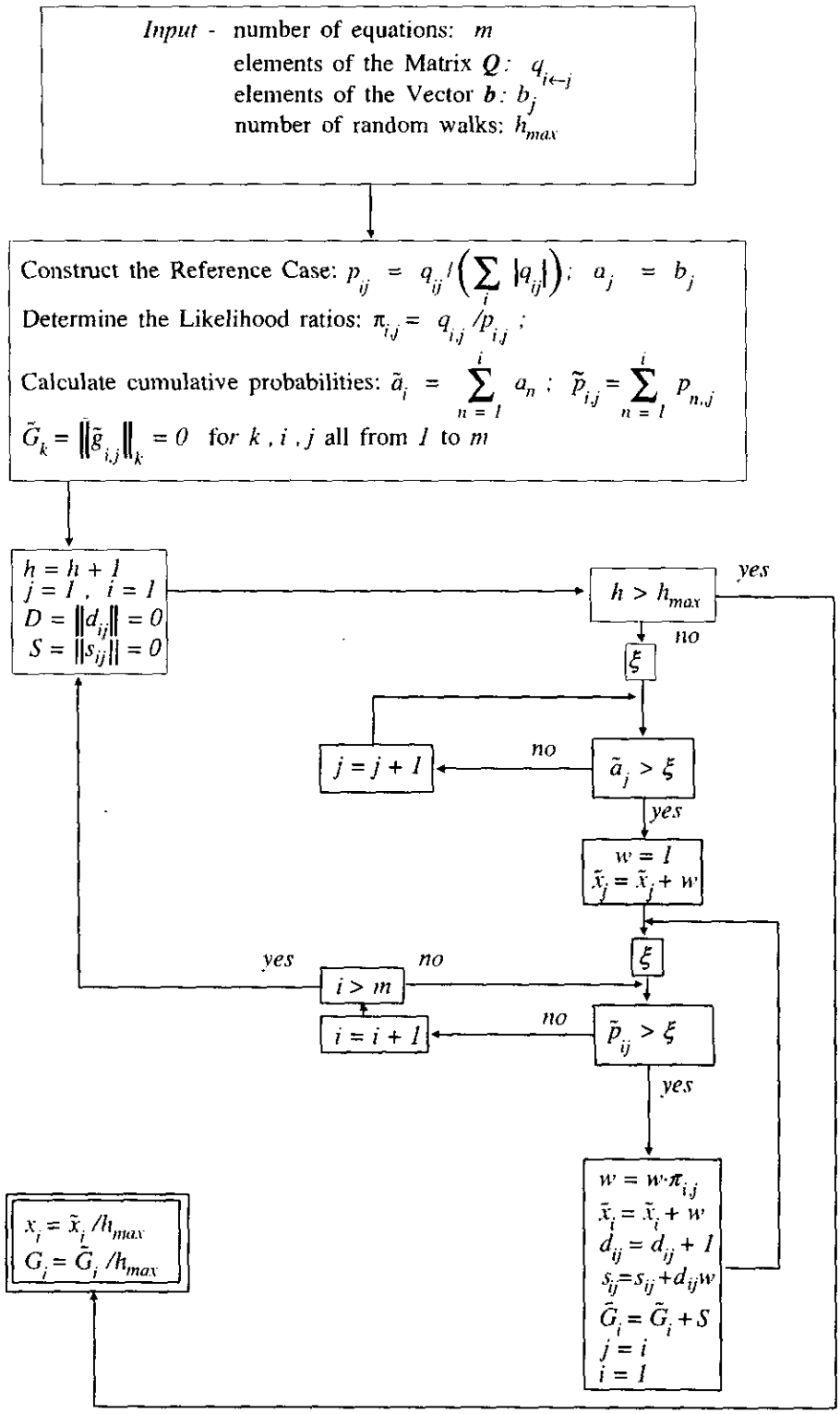

FIG. 5.2. Flow-chart for calculating a system of linear equations including the first-order derivatives (sensitivities) with respect to the elements q_{ij} of the matrix **Q** (note that $\tilde{G}_i = \tilde{G}_{k=i}$).

In the following perturbation analysis carried out either by *multiple correlation* (a) or a *first-order Taylor approximation* (b) two small parameter changes in q_{ij} are considered. In one case q_{11} is increased by 1% leading to the system

$$\begin{pmatrix} x_{2,1} \\ x_{2,2} \\ x_{2,3} \end{pmatrix} = \begin{pmatrix} 0.2525, & 0.50, & 0.08 \\ 0.30, & 0.30, & 0.10 \\ 0.20, & 0.20, & 0.20 \end{pmatrix} \begin{pmatrix} x_{2,1} \\ x_{2,2} \\ x_{2,3} \end{pmatrix} + \begin{pmatrix} 0.30 \\ 0.30 \\ 0.40 \end{pmatrix};$$

and in the other case, q_{32} is decreased by 1% leading to

$$\begin{pmatrix} x_{3,1} \\ x_{3,2} \\ x_{3,3} \end{pmatrix} = \begin{pmatrix} 0.25, & 0.50, & 0.08 \\ 0.30, & 0.30, & 0.10 \\ 0.20, & 0.198, & 0.20 \end{pmatrix} \begin{pmatrix} x_{3,1} \\ x_{3,2} \\ x_{3,3} \end{pmatrix} + \begin{pmatrix} 0.30 \\ 0.30 \\ 0.40 \end{pmatrix}.$$

(a) *Multiple correlation.* The samples of 50,000 random walks provided for the three equations ($k = 1$ to 3), the results shown in Table I.

(b) *Differential operator sampling.* In this example simultaneously with the unknowns x_1, x_2, x_3 also their derivatives (Jacobians) with respect to the parameters q_{ij} are calculated. The perturbation of 1% is so small that a linear Taylor expansion using the sensitivity coefficients $\partial x_k/\partial q_{ij}$ provide good perturbation estimates. Applying the scheme of Fig. 5.2 the results listed in Table II were obtained.

Discussion of results. In the second and third columns of Table III the results from *multiple correlation* ($\Delta x_{ki} = x_{ki} - x_{1i}$) and a linear Taylor approximation ($\Delta\rho = 0.01$), based on *differential operator sampling* are listed. Comparing these values with each other and a deterministic solution listed in column four shows excellent agreement of the perturbation effect. On the other hand, it can be seen that the results from two independent Monte Carlo runs listed in the last column have a standard deviation which is much larger

TABLE I

Perturbation Analysis of Three Linear Equations Carried Out by Multiple Correlation

LISTING OF OUTPUT

NO. OF RAND. WALKS: 50000
EXEC. TIME = 8.39600E + 01 [sec]

THE UNKNOWNS $X(K, I)$ ARE:

	$X(K, 1)$	$X(K, 2)$	$X(K, 3)$
$K = 1$	1.2715E + 00	1.1360E + 00	1.1962E + 00
S.D.	6.9857E − 03	7.3807E − 03	4.2029E − 03
$K = 2$	1.2782E + 00	1.1392E + 00	1.1086E + 00
$K = 3$	1.2706E + 00	1.1351E + 00	1.1029E + 00

THE DIFFERENCES $DX(K, I) = X(K, I) - X(1, I)$ ARE:

	$DX(K, 1)$	$DX(K, 2)$	$DX(K, 3)$
$K = 2$	6.7048E − 03	3.2691E − 03	2.4344E − 03
S.D.	1.1089E − 04	7.4829E − 05	4.4329E − 05
$K = 3$	−9.0842E − 04	−8.8541E − 04	−3.2822E − 03
S.D.	2.6021E − 05	2.9303E − 05	3.7437E − 05

TABLE II

Differental operator sampling in a System of Three Linear Equations

EXEC. TIME = 7.31800E + 01 [sec]
NO. OF RAND. WALKS: 50000

$X(1) = 1.272$ $X(2) = 1.137$ $X(3) = 1.106$

JACOBIAN OF $X(1)$ WITH RESPECT TO:

	$RHO(I, 1)$	$RHO(I, 2)$	$RHO(I, 3)$
$I = 1$	6.6798E − 01	1.1941E + 00	1.8848E − 01
$I = 2$	6.1888E − 01	5.4806E − 01	1.8842E − 01
$I = 3$	1.1176E − 01	9.0800E − 02	8.7580E − 02

JACOBIAN OF $X(2)$ WITH RESPECT TO:

	$RHO(I, 1)$	$RHO(I, 2)$	$RHO(I, 3)$
$I = 1$	3.2682E − 01	5.8052E − 01	9.2300E − 02
$I = 2$	8.6800E − 01	7.7624E − 01	2.5632E − 01
$I = 3$	1.0516E − 01	8.8660E − 02	8.2900E − 02

JACOBIAN OF $X(3)$ WITH RESPECT TO:

	$RHO(I, 1)$	$RHO(I, 2)$	$RHO(I, 3)$
$I = 1$	2.4232E − 01	4.3694E − 01	7.0400E − 02
$I = 2$	3.6398E − 01	3.3066E − 01	1.0954E − 01
$I = 3$	3.7564E − 01	3.2856E − 01	3.1990E − 01

TABLE III

Comparison of Perturbation Estimates

i	Δx_{2i} mult. corr.	$(\partial x_i/\partial\rho_{11})\,\Delta\rho$	Δx_{2i} analytical	Δx^*_{2i} 2 indep. M.C. runs
1	6.70E − 3	6.68E − 3	6.66E − 3	0.9E − 3
	±0.11E − 3			±9.9E − 3
2	3.27E − 3	3.27E − 3	3.20E − 3	−2.1E − 3
	±0.07E − 3			±10.4E − 3
3	2.43E − 3	2.42E − 3	2.46E − 3	0.6E − 3
	±0.04E − 3			±5.9E − 3
i	Δx_{3i}	$(\partial x_i/\partial\rho_{32})(-\Delta\rho)$		Δx^*_{3i}
1	−9.08E − 4	−9.08E − 4	−9.25E − 4	2.3E − 3
	±0.26E − 4			±9.9E − 3
2	−8.85E − 4	−8.87E − 4	−8.64E − 4	5.6E − 3
	±0.29E − 4			±10.5E − 3
3	−3.28E − 3	−3.29E − 3	−3.28E − 3	−9.1E − 3
	±0.04E − 3			±5.9E − 3

than the differental effect and up to two orders of magnitude greater than that of the correlated calculation.

EXAMPLE 2. Multiple correlation estimates using *reference equations* constructed in two different manners. The first is based on the *arithmetic mean* as in expression (3.11) and the other, on the slightly modified procedure (5.13). (Note the difference in the *likelihood ratios* in Tables IV and V.) In both cases we solve the first equation of the previous example, assuming that in the perturbed case $q(1, 3)$ changes from 0.08 to 0.0.

With the *reference equation* based on the arithmetic mean 50,000 random walks render the results listed in Table IV.

TABLE IV

Perturbation Analysis of Three Linear Equations Carried Out by Multiple Correlation and a Reference Case Based on the Arithmetic Mean

INPUT DATA

CUMULATIVE PROBABILITIES OF $P(I, J)$:

	$P(I, 1)$	$P(I, 2)$	$P(I, 3)$
$I=1$	2.5000E−01	5.0000E−01	4.0000E−02
$I=2$	5.5000E−01	8.0000E−01	1.4000E−01
$I=3$	7.5000E−01	1.0000E+00	3.4000E−01

LIKELIHOOD RATIOS $PI(1, I, J)$:

	$PI(1, I, 1)$	$PI(1, I, 2)$	$PI(1, I, 3)$
$I=1$	1.0000E+00	1.0000E+00	2.0000E+00
$I=2$	1.0000E+00	1.0000E+00	1.0000E+00
$I=3$	1.0000E+00	1.0000E+00	1.0000E+00

LIKELIHOOD RATIOS $PI(2, I, J)$:

	$PI(2, I, 1)$	$PI(2, I, 2)$	$PI(2, I, 3)$
$I=1$	1.0000E+00	1.0000E+00	0.0000E+00
$I=2$	1.0000E+00	1.0000E+00	1.0000E+00
$I=3$	1.0000E+00	1.0000E+00	1.0000E+00

LISTING OF OUTPUT

NO. OF RAND. WALKS: 50000
EXEC. TIME = 4.47600E+01 [sec]

THE UNKNOWNS $X(K, I)$ ARE:

	$X(K, 1)$	$X(K, 2)$	$X(K, 3)$
$K=1$	1.2699E+00	1.1364E+00	1.1087E+00
S.D.	±7.8462E−03	±7.8264E−03	±4.5533E−03
$K=2$	1.0972E+00	1.0550E+00	1.0451E+00

THE DIFFERENCES $DX(K, I) = X(K, I) - X(1, I)$ ARE:

	$DX(K, 1)$	$DX(K, 2)$	$DX(K, 3)$
$K=2$	−1.7268E−01	−8.1400E−02	−6.3560E−02
S.D.	±4.8356E−03	±3.5014E−03	±2.2833E−03

The same problem calculated with a *reference equation* based on the procedure (5.14) leads to the results of Table V.

It is interesting to note, that the second type of *reference equation* improves the efficiency of calculating the perturbation effects $[DX(K, I)]$ almost by a factor of two (e.g., $2.28^2/1.74^2 = 1.7$).

For completeness the same perturbation effects are also calculated by a linear Taylor expansion using $\Delta\rho = -1$ and

TABLE V

Perturbation Analysis of Three Linear Equations Carried Out by Multiple Correlation and a Reference Case Based on the Procedure (3.12)

INPUT DATA

CUMULATIVE PROBABILITIES OF $P(I, J)$:

	$P(I, 1)$	$P(I, 2)$	$P(I, 3)$
$I=1$	2.5000E−01	5.0000E−01	7.1579E−02
$I=2$	5.5000E−01	8.0000E−01	1.6105E−01
$I=3$	7.5000E−01	1.0000E+00	3.4000E−01

LIKELIHOOD RATIOS $PI(1, I, J)$:

	$PI(1, I, 1)$	$PI(1, I, 2)$	$PI(1, I, 3)$
$I=1$	1.0000E+00	1.0000E+00	1.1176E+00
$I=2$	1.0000E+00	1.0000E+00	1.1176E+00
$I=3$	1.0000E+00	1.0000E+00	1.1176E+00

LIKELIHOOD RATIOS $PI(2, I, J)$:

	$PI(2, I, 1)$	$PI(2, I, 2)$	$PI(2, I, 3)$
$I=1$	1.0000E+00	1.0000E+00	0.0000E+00
$I=2$	1.0000E+00	1.0000E+00	1.1176E+00
$I=3$	1.0000E+00	1.0000E+00	1.1176E+00

LISTING OF OUTPUT

NO. OF RAND. WALKS: 50000
EXEC. TIME = 4.46700E+01 [sec]

THE UNKNOWNS $X(K, I)$ ARE:

	$X(K, 1)$	$X(K, 2)$	$X(K, 3)$
$K=1$	1.2720E+00	1.1369E+00	1.1108E+00
S.D.	±7.2779E−03	±7.6264E−03	±4.5107E−03
$K=2$	1.0955E+00	1.0498E+00	1.0450E+00

THE DIFFERENCES $DX(K, I) = X(K, I) - X(1, I)$ ARE:

	$DX(K, 1)$	$DX(K, 2)$	$DX(K, 3)$
$K=2$	−1.7644E−01	−8.7145E−02	−6.5822E−02
S.D.	±3.6295E−03	±2.7564E−03	±1.7438E−03

the corresponding elements of the Jacobian calculated in Example 1:

$$(\partial x_1/\partial \rho_{1,1,3})\,\Delta\rho = -1.885\text{E}-1,$$
$$(\partial x_2/\partial \rho_{2,1,3})\,\Delta\rho = -9.23\text{E}-2,$$
$$(\partial x_3/\partial \rho_{3,1,3})\,\Delta\rho = -7.04\text{E}-2.$$

These approximations show a small, but systematic deviation from the "exact" solution obtained by the correlation technique.

The examples shown above were calculated on an Archimedes 440 (ACORN Computers Ltd., UK) home computer with an ARM2 processor and a FP co-processor.

6. FREDHOLM TYPE INTEGRAL EQUATIONS AND PERTURBATION ANALYSIS

6.1. *General Considerations*

The previous section dealt with discrete transition probabilities $p_{i\leftarrow j}$ and source distributions a_i. If the matrix $\mathbf{P}$ is replaced by continuous functions of the form $K(x \leftarrow y)$ and the source vector $\mathbf{a}$, by a distribution function $a(x)$ then Eq. (5.2) becomes a Fredholm type integral equation of the second kind:

$$f(\mathbf{x}) = \int_R \mathrm{d}\mathbf{y}\, K(\mathbf{x} \leftarrow \mathbf{y})\, f(\mathbf{y}) + a(\mathbf{x}), \tag{6.1}$$

where $K(x, y)$ describes the transition properties in the (multidimensional) integration space R.

This integral equation can be interpreted as an infinite matrix equation for which the random walk procedure, described in Section 5, approximates the exact solution as the number of trials increases, provided the criteria of convergence are satisfied.

The random walk procedure can be interpreted in terms of the Neumann series being a solution of this integral equation,

$$f(\mathbf{x}) = \sum_{n=0}^{\infty} f_n(\mathbf{x}) = \sum_{n=0}^{\infty} \int_R \cdots \int_R K(\mathbf{x} \leftarrow \mathbf{u}_n)\, K(\mathbf{u}_n \leftarrow \mathbf{u}_{n-1}) \cdots K(\mathbf{u}_2 \leftarrow \mathbf{u}_1)\, K(\mathbf{u}_1 \leftarrow \mathbf{u}_0)\, d\mathbf{u}_n \cdots d\mathbf{u}_0, \tag{6.2}$$

or in shorter notation

$$f(\mathbf{x}) = \sum_{n=0}^{\infty} \int_R \cdots \int_R \prod_{i=0}^{n} d\mathbf{u}_i\, K(\mathbf{u}_{i+1} \leftarrow \mathbf{u}_i) \equiv \sum_{n=0}^{\infty} \int_R \cdots \int_R \prod_{i=0}^{n} d\mathbf{u}_i\, K_i,$$

where $u_{n+1} = x$ and

$$f_0(\mathbf{x}) = \int_R d\mathbf{u}_0 a(\mathbf{x} \leftarrow \mathbf{u}_0)$$

is the sum of all direct contributions (transitions) from the source,

$$f_1(\mathbf{x}) = \int_R \int_R d\mathbf{u}_1\, d\mathbf{u}_0\, K(\mathbf{x} \leftarrow \mathbf{u}_1)\, a(\mathbf{u}_1 \leftarrow \mathbf{u}_0),$$

the sum of all contributions undergoing an intermediate transition, and so on.

Since there is no essential difference between solving Eqs. (6.1), (6.2), and (5.2) by random walk techniques, the perturbation algorithms developed in Section 5 can readily be applied to Fredholm type integral equations.

6.2. *Correlated Sampling Procedures*

We assume that the kernel K_i depends on a set of parameters represented by the elements of the vector ρ which undergoes a possibly small change $\Delta\rho$. The difference

$$\Delta f(x) \equiv f^*(x) - f(x) \tag{6.3}$$

where

$$f(\mathbf{x}) = \sum_{n=0}^{\infty} \int_R \cdots \int_R \prod_{i=0}^{n} d\mathbf{u}_i\, K(\mathbf{u}_{i+1} \leftarrow \mathbf{u}_i; \rho)$$

and

$$f^*(\mathbf{x}) = \sum_{n=0}^{\infty} \int_R \cdots \int_R \prod_{i=0}^{n} d\mathbf{u}_i\, K(\mathbf{u}_{i+1} \leftarrow \mathbf{u}_i; \rho + \Delta\rho)$$

can be calculated by a correlation technique as described in the previous section.

To this end, again, a reference case characterized by the kernel $C_i = C(\mathbf{u}_{i+1}, \mathbf{u}_i; \rho_0)$ is correlated with the kernels $K_i^* = K(\mathbf{u}_{i+1}, \mathbf{u}_i; \rho + \Delta\rho)$ and $K_i = K(\mathbf{u}_{i+1}, \mathbf{u}_i; \rho)$. As outlined above, the equation with the reference kernel C_i can either be identical to the unperturbed or the perturbed case or it can be constructed such that the variance of the differential is minimized. This choice depends on the problem to be treated. Usually it will be in favour of a reference kernel constructed according to the examples (3.11) or (3.12). Defining the *likelihood ratios* $w_i = K_i/C_i$ and $w_i^* = K_i^*/C_i$ we obtain

$$f(\mathbf{x}) = \sum_{n=0}^{\infty} \int_R \cdots \int_R \left\{\prod_{i=0}^{n} w_i\right\}\left(\prod_{i=0}^{n} C_i\, du_i\right) \tag{6.4a}$$

$$f^*(\mathbf{x}) = \sum_{n=0}^{\infty} \int_R \cdots \int_R \left\{\prod_{i=0}^{n} w_i^*\right\}\left(\prod_{i=0}^{n} C_i\, du_i\right). \tag{6.4b}$$

The expression $\prod C_i\, d\mathbf{u}_i$ describes the random walk procedure of the *reference* system, while $\prod w_i$ and $\prod w_i^*$ refers to the product of the likelihood ratios correlating the unperturbed and the perturbed equation with the *reference* equation.

In the most general case the w_is and the w_i^*s must be determined for each transition separately. Usually they are simple analytical expressions, which can be computed with little effort.

6.3. *Differential Operator Sampling*

Another, quite different way of estimating perturbation effects, is the differential Monte Carlo method. It is based on a general or multivariate Taylor series expansion of $f(\mathbf{x})$:

$$f(x;\rho+\Delta\rho_k)=f(x;\rho)+\frac{\partial f(x;\rho)}{\partial\rho_k}\Delta\rho_k + \frac{1}{2}\frac{\partial^2 f(x;\rho)}{\partial\rho_k^2}(\Delta\rho_k)^2\cdots. \tag{6.5}$$

This method implies the estimation of parametric derivatives. The estimators for the various derivatives of $f(\mathbf{x})$ can be derived directly from the Neumann series. Equation (6.2) differentiated once with respect to one element ρ_k of ρ leads to

$$\frac{\partial f}{\partial\rho_k}=\sum_{n=0}^{\infty}\int_R d\mathbf{u}_n\cdots\int_R d\mathbf{u}_1\left\{\sum_{i=0}^{n}\frac{\partial K_i}{K_i\,\partial\rho_k}\right\}\left(\prod_{i=0}^{n}K_i\right), \tag{6.6}$$

where the brackets (...) stand for the regular Monte Carlo game and {...} for the estimator of the scores rendering the first-order derivatives of $f(\mathbf{x})$.

Expression (6.6) is very similar to (6.4). Only the multiplicative term $\prod w_i$ generating the weights of the perturbed history has been replaced by an additive one containing a sum of first-order derivatives of the transition kernels. The two terms are, however, closely related. Inserting (6.4) into (6.3) gives

$$\Delta f(x)=\sum_{n=0}^{\infty}\int_R\cdots\int_R\left\{\prod_{i=0}^{n}w_i^*-\prod_{i=0}^{n}w_i\right\}\left(\prod_{i=0}^{n}C_i\,du_i\right).$$

Assuming that $C_i=K_i$ and therefore $w_i^*=K_i^*/K_i$ and $w_i=1$ we obtain

$$\Delta f(x)=\sum_{n=0}^{\infty}\int_R\cdots\int_R\left\{\prod_{i=0}^{n}(K_i^*/K_i)-1\right\}\left(\prod_{i=0}^{n}K_i\,du_i\right). \tag{6.7}$$

For a small $\Delta\rho_k$ the kernel K_i^* can be approximated by a linear Taylor series expansion,

$$K_i^*\approx K_i+\frac{\partial K_i}{\partial\rho_k}\Delta\rho_k$$

which allows for a further approximation if $\Delta\rho_k$ is small:

$$\prod_{i=0}^{n}(K_i^*/K_i)-1\approx\prod_{i=0}^{n}\left(1+\frac{\partial K_i}{K_i\,\partial\rho_k}\Delta\rho_k\right)-1$$

$$\approx\Delta\rho_k\sum_{i=0}^{n}\frac{\partial K_i}{K_i\,\partial\rho_k}.$$

Introduced into (6.5) we obtain

$$\lim_{\Delta\rho_k\to 0}\frac{\Delta f(x)}{\Delta\rho_k}=\sum_{n=0}^{\infty}\int_R\cdots\int_R\left\{\sum_{i=0}^{n}\frac{\partial K_i}{K_i\,\partial\rho_k}\right\}\left(\prod_{i=0}^{n}K_i\,du_i\right), \tag{6.8}$$

which is the same as (6.6).

It is interesting to note that the estimator of the first derivative is the sum of derivative operators taken at each transition point in phase space independently of the actual size of the perturbation $\Delta\rho_k$. This is quite different from correlated sampling where in (6.4) the perturbation $\Delta\rho_k$ is an essential parameter of the estimator itself.

The second-order derivative of $f(\mathbf{x})$ in Expression (6.2) is given by

$$\frac{\partial^2 f}{\partial\rho_k^2}=\sum_{n=0}^{\infty}\int_R du_n\cdots\int_R du_1\left\{\left(\sum_{i=0}^{n}\frac{\partial K_i}{K_i\,\partial\rho_k}\right)^2 + \sum_{i=0}^{n}\frac{\partial}{\partial\rho_k}\left(\frac{\partial K_i}{K_i\,\partial\rho_k}\right)\right\}\left(\prod_{i=0}^{n}K_i\right), \tag{6.9a}$$

or

$$\frac{\partial^2 f}{\partial\rho_k^2}=\sum_{n=0}^{\infty}\int_R du_n\cdots\int_R du_1\left\{\left(\sum_{i=0}^{n}\frac{\partial K_i}{K_i\,\partial\rho_k}\right)^2 + \sum_{i=0}^{n}\left[\frac{\partial^2 K_i}{K_i\,\partial\rho_k^2}-\left(\frac{\partial K_i}{K_i\,\partial\rho_k}\right)^2\right]\right\}\left(\prod_{i=0}^{n}K_i\right), \tag{6.9b}$$

where the expression in the brackets {...} stands for its estimator. It consists of the square of the estimate for the first-order derivatives (of the transition kernel) plus the sum of the estimators of the second-order derivatives calculated for all transition points.

Similarly the estimator for cross-terms can be derived,

$$\frac{\partial^2 f}{\partial \rho_\alpha \partial \rho_\beta} = \sum_{n=0}^{\infty} \int_R d\mathbf{u}_n \cdots \int_R d\mathbf{u}_1 \times \left\{ \sum_{i=0}^{n} \left[\frac{\partial^2 K_i}{K_i \partial \rho_\alpha \partial \rho_\beta} - \frac{\partial K_i}{K_i \partial \rho_\alpha} \frac{\partial K_i}{K_i \partial \rho_\beta} \right] + \sum_{i=0}^{n} \frac{\partial K_i}{K_i \partial \rho_\alpha} \times \sum_{i=0}^{n} \frac{\partial K_i}{K_i \partial \rho_\beta} \right\} \left(\prod_{i=0}^{n} K_i \right), \tag{6.8}$$

as well as higher-order derivatives of $f(\mathbf{x})$.

7. CONCLUSIONS

Most stochastic simulation applications belong to one of the categories treated here and are therefore candidates for perturbation analysis. The algorithms outlined in Sections 3 and 4 can be applied to any multidimensional integration carried out by Monte Carlo techniques. The method discussed in Section 5 is applicable to matrix equations describing discrete parameter Markov chains playing a dominant role in operations research problems. Section 6 deals with continuous parameter Markov chains which allows for the simulation of queuing models as well as many equations of physics, such as the Boltzmann particle transport equation, the Kolmogorov differential equation, models of statistical mechanics. In all these cases adding to the classical Monte Carlo procedure the algorithms described above provide a wealth of additional information at the expense of a relatively modest supplementary programming and computation effort.

The correlation option requires the construction of a *reference case* to limit the variance of the perturbation effect. Unfortunately there is no general rule by which reference equations can be constructed. This task depends to a large extent on the problem to be treated and will therefore be one of the more important subjects to be studied in Monte Carlo perturbation analysis. "Null event" models (or δ-scattering in the case of particle transport) will probably play a substantial role in this context.

ACKNOWLEDGMENTS

The author is indebted to his co-worker Dr. W. Matthes for numerous clarifying discussions and valuable advice in preparing this work. He is also very grateful to the referees of *J. Comput. Phys.* for pointing out inconsistencies in the formulation and printing errors. In particular, Dr. Booth from LASL had spent much effort in drawing the authors attention to some weaknesses of the mathematical representation in the first part of the paper. His comments helped a great deal to make the mathematics more readable.

REFERENCES

1. R. Frühwirth and M. Regler, *Monte-Carlo-Methoden* (Bibliographisches Institut, Zürich, 1983).
2. M. C. C. Hall, "Monte Carlo Perturbation Theory in Neutron Transport Calculations," in *Proceedings, Seminar Workshop, ORNL-RSIC-44 (1980)*, p. 47.
3. J. M. Hammersley and D. C. Handscomb, *Monte Carlo Methods* (Methuen, London, 1964).
4. F. James, *Rep. Prog. Phys.* **43**(9), 1145 (1980).
5. I. Lux, *Int. J. Sci. Eng.* (1984).
6. I. Lux and L. Koblinger, *Monte Carlo Transport Methods: Neutron and Photon Calculations* (CRC Press, Boca Raton, FL, 1991).
7. G. A. Mikhailov, *Zh. Vych. Mat. Fiz.* **7**, 915 (1967). [Russian]
8. L. B. Miller, Argonne National Laboratory ANL-7307 (TID-4500), 1967.
9. H. Rief, *Ann. Nucl. Energy* **11**, 455 (1984).
10. H. Rief, "Monte Carlo Uncertainty Analysis," in *Uncertainty Analysis*, edited by Y. Ronen (CRC Press, Boca Raton, FL, 1988).
11. R. Y. Rubinstein, *Monte Carlo Optimization, Simulation and Sensitivity of Queueing Networks* (Wiley, New York, 1986).
12. I. M. Sobol, *Die Monte-Carlo-Methode* (VEB, Berlin, 1983); original in Russian.
13. J. Spanier and E. M. Gelbard, *Monte Carlo Principles and Neutron Transport Problems* (Addison–Wesley, Reading, MA, 1969).