MONTE CARLO TECHNIQUES FOR FILTERING AND PREDICTION OF NONLINEAR STOCHASTIC PROCESSES

Johannes Edmund Handschin

A thesis submitted for the degree of Doctor of Philosophy

December 1968

Centre for Computing and Automation,
Imperial College of Science and Technology,
University of London.

ABSTRACT

In this thesis, contributions are made to the solution of the filtering and prediction problems for nonlinear, stochastic, discrete-time processes. The new feature of the approach presented is that a combination of analytic and numerical methods yields a statistical solution to evaluate parameter estimates with reduced sampling variance.

Our Monte Carlo applications are concerned with the evaluation of integrals. A new multi-stage control variate estimation procedure is developed in which the control variate function parameters are iteratively improved, approximating a minimum variance estimator. Both first and second order stochastic approximation procedures to update the parameters are described.

For the prediction problem, a direct evaluation of the Chapman-Kolmogorov equation may be avoided if the Monte Carlo approach is adopted. The system is simulated and relevant data collected in order to evaluate some estimates describing the probability density an arbitrary number of steps ahead. The conjecture of inefficiency inherent in Monte Carlo methods is invalidated because it is shown that judiciously designed variance reduction techniques improve the accuracy of the estimates. Various modifications of two variance reduction methods - the antithetic variate method and the control variate method - are introduced. The resulting predictor algorithms are tested by application to scalar and multivariable systems.

The nonlinear filtering problem is discussed within the framework of the Bayesian approach. It is not possible to express the posterior density in a closed-form in most cases, so the usual recursion of Bayes' theorem is replaced by an expression which is amenable to Monte Carlo integration. For systems where the plant and measurement noise is assumed to be additive and Gaussian, variance reduction techniques are introduced to improve the sampling procedures. In connection with the single-stage case, the properties of importance sampling densities are examined. The control variate method, however, is found to be better suited for the multi-stage filtering problem. The statistical linearization procedure provides a good starting point for the introduction of an efficient Monte Carlo filtering algorithm. It yields a zero sampling variance estimator when applied to a linear Gaussian system. A second control variate method shows that the Monte Carlo approach can successfully be adapted to estimate the approximation error of existing nonlinear filter equations. A number of examples is computed to demonstrate the relative effectiveness of the algorithms.

ACKNOWLEDGEMENTS

The author is deeply indebted to his supervisor, Dr. D. Q. Mayne, for his guidance, encouragement and advice throughout the course of the research work described in this thesis. Professor J. H. Westcott's interest is gratefully acknowledged. Mention must be made of the many useful and stimulating discussions with fellow students in the Control Group.

This work was made possible by the financial support of AG Brown, Boveri & Co. Baden, Switzerland. In particular, the author is very grateful to Dr. H. Meyer and Professor Dr. A. P. Speiser for their interest and personal encouragement.

J. E. Handschin

INDEX

MONTE CARLO TECHNIQUES FOR FILTERING AND PREDICTION OF NONLINEAR STOCHASTIC PROCESSES

PRINCIPAL SYMBOLS AND NOTATION

In general, capital letters refer to matrices, while lower-case letters denote vectors when underlined and scalars otherwise. Other notations are given as

$\underline{x}_k$	n-dimensional state vector
$\underline{y}_k$	m-dimensional observation vector
$\underline{w}_k$	p-dimensional noise vector acting upon the plant (1.1)
$\underline{v}_k$	r-dimensional noise vector acting upon the observation system (1.4).
k	time argument
$\underline{f}$	function that relates past and current states; eqn.(1.1)
$\underline{g}$	function that relates measurement and state; eqn.(1.4)
$\underline{m}_x$	expected value of initial state; eqn.(1.10)
Σ_x	covariance matrix of the initial condition P.D.F.; eqn.(1.10)
Σ_{w_k}	covariance matrix of the plant noise P.D.F.; eqn.(1.7)
Σ_{v_k}	covariance matrix of the measurement noise P.D.F.; eqn.(1.8)
$p(\underline{x})$	probability density function of $\underline{x}$
$P(\underline{x})$	probability distribution of $\underline{x}$
θ	value of scalar integral (2.1)
A	transformation matrix; eqn.(2.23)
$\phi(x), Y(x)$	control variate function, eqns.(2.24) and (2.56)
$\underline{a}$	parameter vector; eqns.(2.31) and (3.88)
$F(\underline{a})$	cost function in $\underline{a}$, eqn.(2.33)

ρ	correlation coefficient; eqns.(2.48), (3.53) and (3.54)
η_v	variance reduction factor; eqn.(2.71)
η_L	labour ratio; eqn.(2.72)
η	efficiency gain; eqn.(2.73)
$\underline{c}$	auxiliary constant; eqn.(2.64)
N	sample size
$\underline{\mu}_k$	mean of $p(\underline{x}_k)$; eqn.(3.27)
M_k	matrix of second order moments of $p(\underline{x}_k)$; eqn.(3.28)
Σ_k	covariance matrix of $p(\underline{x}_k)$; eqn.(3.29)
$p_a(\underline{x}_k^*)$	P.D.F. of model state $\underline{x}_k^*$; eqn.(3.64)
$\underline{m}_k, P_k$	mean and covariance matrix of $p_a(\underline{x}_k^*)$; eqn.(3.73)
$\underline{a}_k, B_k$	parameter of linear control variate model of plant; eqn.(3.72)
$\theta_k^{(2)}$	matrix of second order moments defined for linear control variate model; eqn.(3.84)
$\underline{\omega}_k$	random vector; eqn.(3.90)
$\underline{a}, \Sigma_\omega$	mean and covariance matrix of $p(\underline{\omega}_k)$; eqn.(3.102)
a_k	parameter matrix of linear weighting sequence model; eqn.(3.118)
$p(\underline{x}\|\underline{y})$	conditional P.D.F.; eqn.(4.2)
$E[\underline{x}\|\underline{y}]$	conditional mean; eqn.(4.2)
$\underline{\theta}_n, \theta_d$	numerator and denominator of conditional mean $E[\underline{x}\|\underline{y}]$; eqn.(4.8)
$E_y[\cdot]$	expectation w.r.t. an unconditional P.D.F. while $\underline{y}$ is kept constant ; eqns.(4.10) and (4.11)

$\hat{\underline{x}}_y$	Monte Carlo estimate of conditional mean $E[\underline{x}\|\underline{y}]$; eqn.(4.15)
$\underline{x}_{bias}$	approximate bias of $\hat{\underline{x}}_y$; eqn.(4.17)
V	sampling covariance matrix of $\hat{\underline{x}}_y$; eqn.(4.19)
$h(\underline{x})$	auxiliary P.D.F. for importance sampling; eqn.(4.23)
$h^o(\underline{x})$	optimal Gaussian sampling density for linear case; eqn.(4.34
$\underline{\mu}$, Σ	mean $\underline{\mu}$ (eqn.(4.38)) and covariance matrix Σ (eqn. (4.39)) for $h^o(\underline{x})$; mean μ (eqn.(4.58)) and variance Σ (eqn.(4.59)) for $h(x)$ in nonlinear case.
$p(\underline{x}_k\|\underline{y}^k)$	conditional P.D.F. at time k; eqn.(5.6)
$E[\underline{x}_k\|\underline{y}^k]$	conditional mean; eqn.(5.11)
$\underline{\theta}_{n,k}$, $\theta_{d,k}$	numerator and denominator of conditional mean; eqn.(5.11)
$\hat{\underline{x}}_{k\|k}$	Monte Carlo estimate of conditional mean $E[\underline{x}_k\|\underline{y}^k]$; eqn.(5.18)
V_k	sampling covariance matrix of $\hat{\underline{x}}_{k\|k}$; eqn.(5.20)
$\theta^{(2)}_{k\|k}$	numerator of conditional second order moment $M_{k\|k}$; eqn.(5.22)
$p_a(\underline{x}^*_k\|\underline{y}^k)$	approximate Gaussian posterior P.D.F. with mean $\underline{\mu}_{k\|k}$ and covariance matrix $\Sigma_{k\|k}$; eqn.(5.24) and eqn.(5.66)
$\underline{a}_k$, B_k	parameters of control variate model of plant; eqn.(5.25)
$\underline{c}_k$, D_k	parameters of control variate model of observation system; eqn.(5.26)
$p_a(\underline{x}^*_k\|\underline{y}^{k-1})$	approximate Gaussian P.D.F. prior to the occurrence of observation $\underline{y}_k$. The mean is $\underline{\mu}_{k\|k-1}$ and the covariance matrix $\Sigma_{k\|k-1}$; eqn.(5.27)

$\underline{\theta}_{an,k}, \theta_{ad,k}$	numerator and denominator of conditional mean defined for linear control variate model; eqns.(5.42) and (5.43)
$\vartheta_{ad,k}$	m-variate normal P.D.F. (eqns.(5.47) and (5.76)) with mean $\underline{m}_k$ and covariance matrix R_k defined by eqns. (5.55) and (5.77)
$\underline{m}_{k\|k}$, $P_{k\|k}$	mean and covariance matrix of posterior P.D.F. defined for control variate model (5.71) where all parameters are found by statistical linearization w.r.t. $p_a(\underline{x}_k^* \| \underline{y}^k)$ of eqn.(5.66)
A^T	transpose of A
$\|A\|$	determinant of A
A^{-1}	inverse of A
$\|\underline{x}\|^2$	Euclidean norm $\underline{x}^T \underline{x}$
$x(i)$	i:th component of $\underline{x}$
$\underline{x}_j$	j:th realization of random vector $\underline{x}$
$\{\underline{x}\}_j$	random sample of size N
$\underline{x}_j^-$	antithetic variate to $\underline{x}_j^+$
$\underline{x}_j^*$	control variate of $\underline{x}_j$
$\hat{\underline{a}}$	statistical estimate of $\underline{a}$
$\mathrm{var}(\hat{\underline{a}})$	sampling covariance matrix of $\hat{\underline{a}}$
$E[\underline{x}]$	expected value of $\underline{x}$
$\mathrm{var}(\underline{x})$	covariance matrix of $\underline{x}$
$\sum_j^N$	summation operator with an implied lower limit of one
$\prod_{i=1}^k a_i b c$	product operator; affects only term immediately following it, i.e. $\prod_{i=1}^k a_i b c \triangleq (a_1 a_2 .. a_k) b c$

The following abbreviations are used in the text:

$n(\underline{x};\ \underline{\mu}, \Sigma)$	normal P.D.F. with mean $\underline{\mu}$ and covariance matrix Σ
P.D.F.	probability density function
L.H.S.	left hand side
R.H.S.	right hand side
eqn.	equation
w.r.t.	with respect to

Further notational details are described in the text, when required.

CHAPTER ONE

GENERAL DISCUSSION AND PROBLEM STATEMENT

1.1 Introduction

The estimation of the states of an arbitrary n-dimensional nonlinear system subject to stochastic disturbances is called the nonlinear prediction problem. The nonlinear filtering problem pertains to the state estimation from $m \leq n$ noisy observable outputs, given as nonlinear functions of the n state variables.

Modern estimation theory has passed through three consecutive periods. Period I started with the Wiener-Kolmogorov theory[1,2] of steady-state filtering for stationary stochastic processes with known statistical characteristics. The solution obtained in the frequency domain yields a linear time-invariant physically realizable filter. Transformation techniques are required as the solution is given by the Wiener-Hopf equation, an integral equation. An extensive bibliography of the papers which have generalized, modified, interpreted and extended the original Wiener-Kolmogorov theory is given by Zadeh[3].

The work of period II is based on the Kalman-Bucy theory[4,5]. It deals with linear, finite-time filtering problems for nonstationary processes and results in a set of five differential or difference equations using the concept of state variables. The main advantages of this approach are a complete and successful resolution of the problem of synthesis of the optimal filter and an effective error analysis of

the filter performance by the determination of the error covariance matrix. This theory has found wide application and heuristic techniques have been developed to apply it in certain nonlinear situations. A summary of such applications is given by Sorenson[6].

Finally, in period III the research work is directed toward a nonlinear estimation theory. The primary impetus for the current activity in nonlinear estimation stems from the work of Stratonovitch[7], Wonham[8], and Kushner[9] - to mention but a few of the early contributors in this field. As truly optimal nonlinear filters are of infinite dimensionality, see Kushner[60], finite dimensional approximations have been proposed by Wonham[64], Bass et al.[12], Kushner[10] and Fisher[11]. But as Bucy[13] points out, examples of nonlinear filtering problems for which the performance could be established are non-existent. Indeed, the comparison of two estimators for a given set of data is not necessarily meaningful since one estimator may be much more sensitive to the particular parameter configuration than the other. Because a general nonlinear estimation theory is yet unknown, one has no alternative but to resort to simulation techniques. This is one of the motivations for the research work contained in this thesis.

The significance of the estimation problem for engineering applications is suggested by the variety of applications concerned with aerospace guidance[62], design of state-vector control[65], plant identification[66], etc. Indeed, fundamental to the implementation of optimal and adaptive feedback control is the necessity of determining, or estimating, present and future values of the system state variables and/or parameters.

Optimal stochastic controllers for linear systems with a quadratic

error criterion and subject to Gaussian white noise inputs are synthesized by cascading an optimal estimator with a deterministic optimal control. This result is known as the separation principle which was originally formulated by Gunckel[14].

No analogous result can be anticipated for nonlinear stochastic system optimization. Indeed, the problem of joint estimation and optimization for a nonlinear system belongs to a class of extremely difficult unsolved statistical optimization problems. From the viewpoint of applications one has therefore no alternative but to use estimates of state variables. Here, a second comment on the motivation for the present work is appropriate. Instead of concentrating on approximations, the nonlinear estimation problem as a whole is put into a statistical framework and Monte Carlo methods are designed for an effective resolution of the problems of filtering and prediction of nonlinear stochastic processes. No claim is made that this approach leads to a rigorous mathematical nonlinear estimation theory but it is felt that it offers a meaningful alternative to and improvement on existing approximation techniques based on analytic manipulations. Our prime interest centers around specific quantities associated with the posterior probability density function (P.D.F.) which can be more easily calculated than the entire P.D.F. and which may have some potential use to generate control strategies.

1.2 The mathematical model

The object of this section is to specify a mathematical model; the choice must be made carefully since the whole analysis depends upon the characteristics of the model. The problem considered in this thesis is the estimation of the state variables of a k-stage, discrete-time nonlinear system which is excited by a sequence of independent random vectors. The state variables $\underline{x}_k$ of the dynamical stochastic system are continuous and are defined in some subset of a Euclidean space. They are assumed to evolve according to the stochastic nonlinear difference equation

$$\underline{x}_{k+1} = \underline{f}(\underline{x}_k, \underline{w}_k, k) \tag{1.1}$$

where the state vector $\underline{x}_k$ is n-dimensional. At each time k the system is disturbed by the random noise $\underline{w}_k$, a $p \leq n$-dimensional vector.

The meaning of 'stochastic' as it is used for eqn. (1.1) implies not only the probabilistic nature of such a system but also the complete knowledge of the prior P.D.F. of all random quantities involved in the system description. That is, we assume we know the P.D.F. $p(\underline{w}_k)$ of the sequence $\underline{w}^k$ defined by

$$\underline{w}^k \triangleq \underline{w}_1, \underline{w}_2, \cdots, \underline{w}_k \quad . \tag{1.2}$$

The independence of $\underline{w}_k$ from one sampling time to the next is an additional assumption which implies

$$p(\underline{w}_1, \underline{w}_2, \cdots, \underline{w}_k) = \prod_{i=1}^{k} p(\underline{w}_i) \tag{1.3}$$

Furthermore, the statistical properties of the state $\underline{x}_k$ at time k=1,

the initial condition, are also completely specified as the P.D.F. $p(\underline{x}_1)$ is assumed to be known. The random vector $\underline{x}_1$ is assumed to be independent of the noise sequences $\underline{w}_k$ in eqn. (1.1) and $\underline{v}_k$ in eqn. (1.4).

The vector valued function $\underline{f}(.,.,.)$ is considered to be known and is referred to as the plant. Any known forcing functions, such as test signals or control inputs, are accounted for by the explicit dependence of the plant on the time parameter k.

The behaviour of the plant (1.1) is generally observed imperfectly through the $m \leq n$-dimensional vector $\underline{y}_k$ which is functionally related to the state variables and which contains random errors. These observations are assumed to be subject to the following known relationship

$$\underline{y}_k = \underline{g}(\underline{x}_k, \underline{v}_k, k). \tag{1.4}$$

The measurement noise $\underline{v}_k$ is an $r \leq m$-dimensional vector and is a member of a white noise sequence with known P.D.F. $p(\underline{v}_k)$. It is independent of $\underline{x}_1$ and $\underline{w}_k$. A schematic diagram of the general stochastic process to be considered is shown in fig. 1.1.

An unnecessary but often useful assumption is to state that $\underline{w}^k$ and $\underline{v}^k$ are white Gaussian sequences. In this case we use the notation

$$p(\underline{w}_k) = n(\underline{w}_k; \underline{0}, \Sigma_{w_k}) \tag{1.5}$$

and

$$p(\underline{v}_k) = n(\underline{v}_k; \underline{0}, \Sigma_{v_k}) \tag{1.6}$$

to indicate that $\underline{w}_k$ and $\underline{v}_k$ possess normal P.D.F. with zero means and

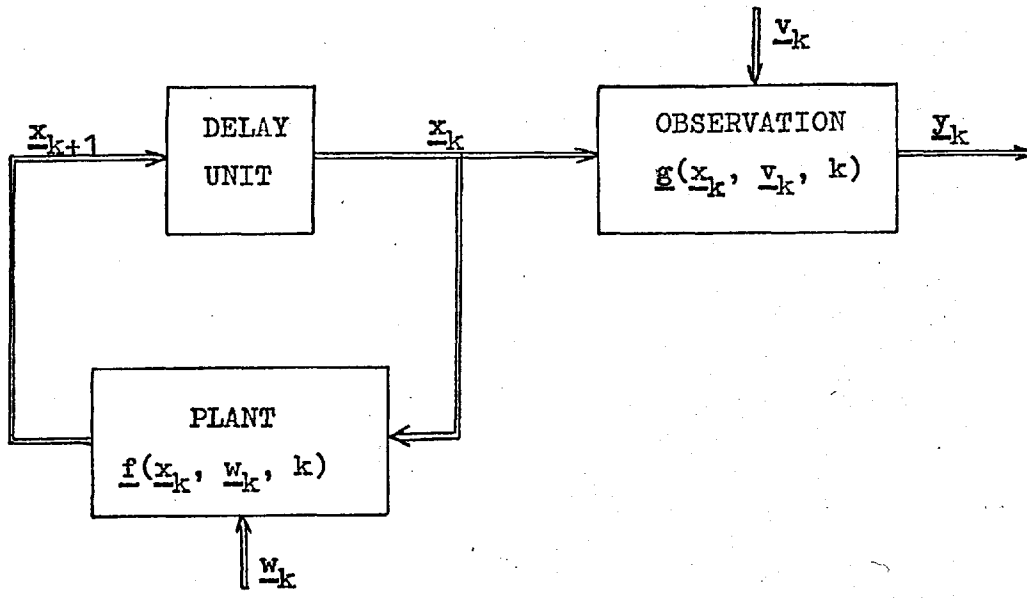

Fig. 1.1 Schematic diagram of the stochastic process considered.

covariance matrices Σ_{w_k} and Σ_{v_k}, respectively defined by

$$E[\underline{w}_i \underline{w}_k^T] \triangleq \Sigma_{w_k} \delta_{ik} \tag{1.7}$$

$$E[\underline{v}_i \underline{v}_k^T] \triangleq \Sigma_{v_k} \delta_{ik} \tag{1.8}$$

for all integers i and k. $E[.]$ is the expectation operator and δ_{ik} is the Kronecker delta (=1 if i=k, and =0 if i≠k). The independence between $\underline{w}_k$ and $\underline{v}_k$ implies

$$E[\underline{w}_i \underline{v}_k^T] = 0.$$

If the initial condition P.D.F. $p(\underline{x}_1)$ is Gaussian we use the notation

$$p(\underline{x}_1) = n(\underline{x}_1; \underline{m}_x, \Sigma_x) \tag{1.10}$$

to indicate that $p(\underline{x}_1)$ has the mean $\underline{m}_x$ and the covariance matrix Σ_x.

The special case $\Sigma_x = 0$ corresponds to a deterministically specified initial condition.

Problems such as the following then arise and are considered in this thesis:

(1) Given that the states $\underline{x}_k$ of the plant (1.1) are all accessible, estimate for a particular time k some relevant parameters belonging to the P.D.F. $p(\underline{x}_k)$ of the state $\underline{x}_k$. This is the state prediction problem.

(2) Given a set of observations $\underline{Y}^k$ related to the states $\underline{x}_k$ by eqn. (1.4), estimate for a particular time k some relevant parameters of the posterior P.D.F. $p(\underline{x}_k|\underline{Y}^k)$. This is the state filtering problem.

(3) If the time argument k is considered to be varying the previous two problems become the trajectory prediction problem and the trajectory filtering problem respectively.

1.3 Previous results

In the Kalman-Bucy theory[4,5] the distinction between continuous-time and discrete-time systems is not fundamental. Unfortunately this is not true for nonlinear estimation theory. Since physical systems are usually described by differential equations the mathematical description to be considered could have reasonably well be stated as a continuous-time system where stochastic differential equations replace eqn. (1.1). Likewise eqn. (1.4) would be stated as a continuous-time

measurement process. From the mathematical point of view this approach allows rigorous derivations of approximations. Recent results have been published, among others, by Bucy[15], Bass et al.[12], Kushner[10] and Sunahara[16]. Fisher[11] studied the nonlinear filtering problem from the point of view of approximating the posterior P.D.F. Schwartz[17] considered the rigorous validation of approximate filter equations. Jazwinski [18] investigated the estimation problem for continuous-time plant dynamics where the states were observed at discrete-time intervals.

The mathematical model to be used in our work assumes a discrete-time formulation. Such systems originate in various ways. If the state variables of a continuous-time system are allowed to change only at discrete-times the system is described by a vector difference equation of the type of eqn.(1.1). But this formulation may also arise from other considerations where a continuous-time dynamic system is not involved, see Brown[67]. It is quite a difficult problem to reduce a continuous-time system to the form of eqn. (1.1). However, it is believed that the present discrete-time formulation provides a realistic basis for several reasons:

(1) For minimal variance filtering, which is a probabilistic criterion of optimality, the manipulations leading to the nonlinear filter equations are simplified by assuming that the random processes are white noises. But this assumption complicates the interpretation of the mathematical results in the light of reality as such noise is physically unrealizable. As a white noise sequence differs considerably from a white noise process this complication does not arise in our formulation.

(2) The general solution of the continuous-time nonlinear estimation problem yields systems of complicated partial differential-integral equations that must be solved. The difficulties inherent in obtaining numerical solutions to practical problems using this formulation appear to be excessive.

(3) Finally, measurement data are usually available only at discrete-time intervals. Also, the control signals based on estimates are usually applied at discrete-time instants.

However, we would like to point out that the solutions presented in this thesis are applicable to continuous-time stochastic processes provided the disturbances acting on the plant eqn. (1.1) and the observation system (1.4) can be modelled as a discrete-time white noise sequence. The plant is then given by an ordinary differential equation which is integrated in the usual way. The noise affects the integration only at discrete intervals. Thus, from the simulation point of view the only difference between the discrete-time formulation adopted in this thesis and a continuous-time system of the type just mentioned is given by the fact that we do not directly proceed from time k to k+1 but have to integrate over the unit interval with a suitable step length h such that 1/h is an integer.

The important problem of optimal control and estimation given noisy measurements $\underline{y}^k$ is best solved using the P.D.F. of the state $\underline{x}_k$ where this function is conditioned (in the technical sense of probability theory) on the knowledge of all available observations $\underline{y}^k$. The reasons for that are twofold. Firstly, this conditional or posterior

P.D.F. sums up all available knowledge about the state of the plant. Secondly, it can - at least in principle - be calculated. The general concept of dealing with the conditional P.D.F. is referred to as the Bayesian approach to estimation and control. Ho and Lee[19] elaborated on the difficulties in the steps leading to the computation of the posterior P.D.F. $p(\underline{x}_k | \underline{y}^k)$. Cox[20] and Larson and Peschon[61] developed a recursive equation for the trajectory filtering problem and then applying a dynamic programming technique to this equation they deduced the modal trajectory by a series of minimizations. Sage and Masters[21] extended the work of Detchmendy and Sridhar[22] to discrete-time systems without requiring detailed statistical information concerning system disturbances. Aoki[23] considered both the estimation and control problem. Sorenson[24] suggested methods that allow the extension of the range of applicability of the linear Kalman-Bucy theory and derived a procedure to approximate the posterior P.D.F. by a truncated Edgeworth series expansion that included the fourth central moments. Our approach[68,69] differs from these analytically oriented methods; the new feature of the solutions presented is the introduction of Monte Carlo techniques to obtain efficient solutions of the nonlinear filtering and prediction problem; that is we are primarily concerned with the design of sampling methods which make direct use of the nonlinear system formulation.

1.4 Outline and contributions of this thesis

In order to list the main topics of our thesis this section presents the outline of, and some reflections on, the more interesting of the points discussed hereafter. In this way we state the originality of the ideas presented and point out which topics are considered to be of greatest interest.

As already indicated the main emphasis in this thesis is put on the statistical formulation of the nonlinear estimation problem and not on the development of analytic approximations. There are four distinct reasons for this approach:

(1) There is no theory available to establish what filter configuration yields the 'best' results and hence it is believed that a solution based on the actual nonlinear system is more realistic.

(2) As a consequence of the previous remark the nonlinear estimation problem has to be tackled from a statistical point of view. This, however, allows us to estimate any order moments of the P.D.F. in question. The accuracy of these estimators is known and depends on the effort put into the sampling procedure. No dimensionality restriction limits this approach.

(3) Monte Carlo techniques provide powerful methods to compare various approximate filtering equations.

(4) Finally, Monte Carlo techniques may be used to carry out a study of trade-off between truncation and accuracy with possible applications to stability theory.

Chapter two deals with those aspects of sampling procedures from

given P.D.F. which will be used to develop Monte Carlo techniques for the nonlinear prediction and filtering problem. Two fundamental theorems provide the mathematical foundation of these methods. The necessity of variance reduction techniques becomes apparent from a first investigation of crude Monte Carlo methods; section 2.2. Sampling from a known P.D.F. is shown to be feasible under fairly general conditions for both the scalar and the multivariate case; section 2.3. Several variance reduction techniques have been used in the past. The concept of multi-stage sampling procedures (section 2.4) is applied to the control variate method. The first order algorithm is an extension based on some results previously derived for importance sampling. The second order method is new and has very interesting properties compared with ordinary linear regression. In order to avoid the problem of combining correlated estimates a new two-stage estimator is proposed. A standard example is used to illustrate the potential power of the new procedure; section 2.5.

The state and trajectory prediction problem is dealt with in chapter three. The observation process (1.4) is not considered; we only refer to the plant eqn. (1.1). The system is assumed to start at time k= 1. The random initial condition $\underline{x}_1$ possesses a known P.D.F. $p(\underline{x}_1)$. Some of the difficulties we are faced with in attempting an analytic solution are discussed in section 3.1. The crude Monte Carlo predictor of section 3.2 is based on the concept of sample moments. The application of the antithetic variate method to the prediction problem may reduce the sampling error considerably. The extension

to several antithetic estimators is new in the context of the n-variate Gaussian P.D.F. and yields a further improvement of the sampling procedure. The control variate method, section 3.4, is a combination of analytic and sampling techniques. It leads to three versatile Monte Carlo methods for predicting the states $\underline{x}_k$. Although the statistical linearization procedure itself is well known its incorporation in a Monte Carlo method is new. A two-stage adaptive control variate method using a gradient technique is derived for the scalar and multivariable state prediction problem. Numerical examples are contained in section 3.5. They illustrate the feasibility of the proposed methods. A comparison between the various techniques allows us to draw a number of conclusions.

The evaluation of conditional expectation parameter estimates for one observation (static case) is contained in chapter four. A crude Monte Carlo solution is presented in section 4.2 for the evaluation of Bayes' theorem. The investigation of the resulting ratio estimator leads to some interesting conclusions for the successful solution of nonlinear filtering via Monte Carlo techniques. The concept of importance sampling is developed in section 4.3 for improving the crude estimator. The properties of the new estimators are examined in section 4.3.1 and it is shown that there always exist - at least in principle - functions such that the sampling error is zero. The optimal zero sampling variance estimator is derived in sections 4.3.2 and 4.3.3 for the linear Gaussian problem. For nonlinear observation systems one possible approximation for the sampling P.D.F. of the

denominator in Bayes' theorem is obtained in section 4.3.4. Possible disadvantages of importance sampling are also indicated. In order to compare our new methods in connection with conditional P.D.F. section 4.4 contains a concise comparison between our work and the original conditional Monte Carlo.

The multi-stage, nonlinear filtering problem is discussed in chapter five. Using the Bayesian approach, section 5.2, a general Monte Carlo procedure is obtained for the nonlinear filtering problem; section 5.3. The necessity of variance reduction techniques is again underlined by the low accuracy of the crude procedure. Statistical linearization has been used to obtain approximate nonlinear filtering equations. It is used in a new manner to implement the control variate method, section 5.4, and found extremely useful in improving not only the crude Monte Carlo method but also the analytic approximation. The optimal but infinite dimensional nonlinear filter is often replaced by a set of expressions for the mean and the covariance matrix of an approximate posterior P.D.F. It is shown in section 5.5 that such a set could be used to design an alternative control variate method. This application is important because it shows that our improved Monte Carlo procedures are versatile enough to compare various structures of approximate filter equations.

Chapter six contains the numerical results for nonlinear filtering techniques. As there is no fundamental difference between the single- and multi-stage filtering problem, this chapter contains examples for chapters four and five. A variety of scalar observation

systems for the static case is studied in section 6.2. A comparison between different methods reveals the superiority of the control variate method over importance sampling. The transition to multi-stage filtering proceeds via the example reported in section 6.3. The results are compared with those previously obtained by analytic approximation techniques. The nonlinear scalar example of section 6.4 allows us to draw a variety of conclusions and the multidimensional example of section 6.5 shows how our method may be used for the combined state and parameter estimation problem.

Conclusions based on the experience with Monte Carlo techniques for filtering and prediction of nonlinear stochastic processes and suggestions for further research work are given in chapter seven.

CHAPTER TWO

MONTE CARLO METHODS

2.1 Introduction

The object of this chapter is to discuss those aspects of Monte Carlo methods which will be used in this thesis for the filtering and prediction of nonlinear stochastic processes. Monte Carlo methods are concerned with the application of random sampling to problems of applied mathematics. Table 2.1 shows one possible classification of different numerical methods[25]. Monte Carlo techniques are shown to be stochastic solutions to exactly formulated problems.

		PROBLEM FORMULATION	
		exact	stochastic
METHOD OF SOLUTION	exact	Numerical analysis	Probability theory
	stochastic	Monte Carlo methods	Simulation methods

Table 2.1 Classification of different numerical methods.

The use of variance reduction techniques as well as simulation is implicit in the definition of Monte Carlo methods. Thus, they

generally include the following three topics:

(1) Choice or definition of the probability process; see section 2.2.

(2) Generation of sample values of the random variables; see section 2.3.

(3) Design and use of variance reduction techniques; see section 2.4.

For a more detailed discussion of the principles of Monte Carlo techniques we refer to the books by Kahn[26], Shreider[27], and Hammersley and Handscomb[36].

The application of Monte Carlo methods in nuclear physics, statistical mechanics and operational research is certainly well established to-day. It is equally true that the specific advantages of Monte Carlo methods have not yet been fully exploited for designing optimal estimation and control policies in nonlinear static and dynamic systems. It is believed that an extension towards such applications would provide a useful alternative to existing analytic approximations. Indeed, some applications in this field have already shown promising results. They can be put into three classes: Kwakernaak[28] and Mayne[29] use sampling techniques to solve optimal stochastic control problems. Little[30], Handler[31] and Cumming[32] propose general procedures to solve parabolic partial differential equations using Monte Carlo methods. Finally, there are a number of papers where crude Monte Carlo methods are used to measure the relative accuracy and computational efficiency of various estimators; e.g. Carney and

Goldwyn[33] or LeMay et al.[56].

The study of partial differential equations is motivated by the filtering and prediction problem for continuous-time systems. Here, the main difficulties in the application of Monte Carlo techniques arise from the modelling of the system dynamics as a continuous Markov process; i.e. as the solution of a stochastic differential equation. The important field of modelling has been treated in many contributions; see Clark[54] and Kailath and Frost[55] to mention but two among many others.

Only Mayne[29] uses variance reduction techniques in his approach to nonlinear stochastic control problems whereas in the remaining applications mentioned above these techniques are left out. However, we feel that the opinion that Monte Carlo is a last resort solution stems from the omission of variance reduction techniques. This thesis is devoted to the design of Monte Carlo methods for filtering and prediction of nonlinear stochastic processes. In our approach we discuss all of the three topics mentioned above. Since we are concerned with system dynamics represented in discrete-time formulation we can avoid the admittedly important but fairly developed field of stochastic modelling of continuous-time systems and concentrate on designing and applying efficient Monte Carlo procedures.

2.2 Fundamental principles of the Monte Carlo method

2.2.1 Evaluation of a scalar integral

A Monte Carlo technique starts with the specification of a sampling procedure for the estimation of some characteristic of a stochastic process. As pointed out by Hammersley and Handscomb[36] almost every Monte Carlo computation may be regarded as estimating the value of an integral

$$\theta = \int_X f(x)\, p(x)\, dx \tag{2.1}$$

where $f(x)$ is some known function of x and $p(x)$ is a known P.D.F. defined over the space X. It is desired to specify a sampling procedure which results in an unbiased estimate $\hat{\theta}$ of θ.

Using the definition of the expectation operator E we can interpret θ in eqn. (2.1) as

$$\theta = E\,[f(x)] \tag{2.2}$$

where the expectation E is w.r.t. $p(x)$. The mean θ in eqn. (2.2) is estimated by the sampling mean $\hat{\theta}_N$, defined by

$$\hat{\theta}_N \triangleq N^{-1} \sum_{j}^{N} f(x_j). \tag{2.3}$$

Let us use the notation $\{x\}_j \triangleq [x_j;\ j=1,2,\ldots,N]$ to denote a random sample of size N drawn from the P.D.F. $p(x)$. The generation of such a random sample is essential for all Monte Carlo work and is discussed in section 2.3. The subscript of $\hat{\theta}_N$ in eqn. (2.3) indicates that this estimate is based on a finite sample of size N. The random variable $\hat{\theta}_N$ is one stochastic solution to the exactly formulated integral of eqn. (2.1).

2.2.2 Two fundamental theorems

The mathematical foundation of Monte Carlo methods is provided by the following two theorems. For the sake of completeness they are both stated here. For their proofs we refer to the literature; e.g. Wilks[34].

1. The strong law of large numbers

Let $\{x\}_j$ denote a random sample of size N drawn from the P.D.F. p(x) to compute the random variable $\hat{\theta}_N$ using eqn. (2.3). If the independent variates $(f(x_j);\ j=1,2,\ldots N)$ have a common distribution and if the integral of eqn. (2.1) exists in the ordinary sense, then $\hat{\theta}_N$ converges with probability one to θ as N tends to infinity.

2. The central limit theorem

On the premises that the second order moment $\theta^{(2)} \triangleq E[f^2(x)]$, defined by

$$E[f^2(x)] \triangleq \int_X f^2(x)\, p(x)\, dx, \qquad (2.4)$$

exists and that the sample size N is large, the probability that the event E, defined by $E:\ \theta-\delta \leq \hat{\theta}_N \leq \theta+\delta$, occurs is asymptotically independent of the exact nature of f(x) or p(x). Indeed, the probability depends only on N and the variance of f(x), defined by

$$\mathrm{var}(f(x)) \triangleq E[(f(x) - \theta)^2] = \int_X (f(x)-\theta)^2 p(x)\, dx. \qquad (2.5)$$

Combining these two theorems, an estimate of the amount that $\hat{\theta}_N$

deviates from θ for a certain N is approximately given by

$$\text{Prob}\{\hat{\theta}_N \leq \theta+\delta\} \doteq \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{\delta\sqrt{N}}{(\text{var}(f(x)))^{1/2}}} \exp(-\frac{x^2}{2})dx. \qquad (2.6)$$

The probability that the deviation of $\hat{\theta}_N$ from θ will exceed $\pm \lambda\,[\text{var}(f(x))/N]^{1/2}$ is given in the following abbreviated table 2.2 of $\sqrt{2/\pi} \int_{\lambda}^{\infty} \exp(-\frac{x^2}{2})\,dx$.

λ	0.6745	1.0	2.0	3.0	4.0
P	0.5	0.3173	0.0455	0.0027	0.0001

Table 2.2 Multiples λ of standard deviations to determine $\alpha = 100\ P\ \%$ confidence limits of $\hat{\theta}_N$.

The sampling variance of the estimate $\hat{\theta}_N$, $\text{var}(\hat{\theta}_N)$, is given by

$$\text{var}(\hat{\theta}_N) = \frac{\text{var}(f(x))}{N} . \qquad (2.7)$$

Taking the square root on both sides of eqn.(2.7) yields the standard deviation or sampling error of $\hat{\theta}_N$. Using table 2.2 it is now possible to state that the probability of the event E, e.g. defined by $\{\theta - 2[\text{var}(\hat{\theta}_N)]^{1/2} \leq \hat{\theta}_N \leq \theta + 2[\text{var}(\hat{\theta}_N)]^{1/2}\}$, is about 0.95; or in other words, we can say with 95% confidence that $\hat{\theta}_N$ is within twice $(\text{var}(\hat{\theta}_N))^{1/2}$ around the mean θ. But we cannot assume to know θ and hence our statement implies that the repeated application of the Monte

Carlo experiment yields in the long run estimates $\hat{\theta}_N$ of which 95% lie within the interval $\hat{\theta}_N \pm 2\,[\mathrm{var}(\hat{\theta}_N)]^{1/2}$.

The results of the central limit theorem depend on N being large and the variance var(f(x)) being finite and known. However, since eqn.(2.5) implies the knowledge of θ, var(f(x)) is usually unknown too and is estimated by

$$\hat{\mathrm{var}}(f(x))_N = N^{-1} \sum_j^N (f(x_j) - \hat{\theta}_N)^2. \qquad (2.8)$$

In general, whenever it is desired to estimate values of formulas in which $\theta^{(2)}$ or θ are involved the expected values can be replaced by $\hat{\theta}_N^{(2)}$ and $\hat{\theta}_N$ respectively, where

$$\hat{\theta}_N^{(2)} \triangleq N^{-1} \sum_j^N f^2(x_j) . \qquad (2.9)$$

While the estimates will almost always be biased, the amount of the bias is usually proportional to N^{-1} and can be ignored if N is reasonably large, i.e. $N > 30$.

2.2.3 The crude Monte Carlo estimator

An estimator for θ as defined by eqn.(2.3) will be referred to as a crude Monte Carlo estimator. Its accuracy is proportional to $[\mathrm{var}(f(x))/N]^{1/2}$ and an improvement is only achieved by increasing the sample size N. However, a useful Monte Carlo method should be based on the general principle that if, at any point of the sampling

procedure, an estimate can be replaced by an exact value, the sampling error in the final result will be reduced[36]. A new variance reduction technique is introduced in section 2.4.

The following remarks are intended to complement the discussion of the crude Monte Carlo estimator.

(1) As $N \to \infty$, the sampling distribution of the estimate $\hat{\theta}_N$ tends asymptotically towards a normal P.D.F. whose mean and variance are given by

$$E[\hat{\theta}_N] = \theta = E[f(x)] \ , \ var(\hat{\theta}_N) = N^{-1} \, var(f(x)) . \qquad (2.10)$$

(2) Similarly as $N \to \infty$, the sampling distribution of the estimated variance $\hat{var}(f(x))_N$ tends asymptotically towards a normal P.D.F. As shown by Wilks[34] if the fourth central moment $\theta_c^{(4)}$, defined by

$$\theta_c^{(4)} \triangleq E[(f(x)-\theta)^4] , \qquad (2.11)$$

is finite, mean and variance of the sampling distribution are given by

$$E[\hat{var}(f(x))_N] = \frac{N}{N-1} \, var(f(x)) \qquad (2.12)$$

and

$$var(\hat{var}(f(x))_N) \doteq \frac{1}{N} [\theta_c^{(4)} - \frac{N-3}{N-1} \, var^2(f(x))] \qquad (2.12a)$$

respectively.

(3) For the sake of simplicity the discussion so far has been restricted to a scalar integral. While in actual practice it is rarely useful or efficient to apply crude Monte Carlo methods unless the integral is rather highly dimensional the results obtained for the

scalar case can readily be extended if the variable of integration x is considered to represent an n-dimensional vector $\underline{x}$.

(4) The reason that sampling is useful in evaluating multiple integrals is that neither of the theorems presented in section 2.2.2 depends on the dimensionality of the integral. The number of points required to evaluate a multidimensional integral to a fixed level of accuracy depends only on $(\mathrm{var}(f(x)))^{1/2}$ once there are enough samples so that the central limit theorem is reliable. By contrast, in almost all standard integration techniques the number of points required to evaluate an integral goes up in geometrical progression with its dimensionality. This exponential increase does not occur if the integration is done by Monte Carlo methods.

(5) Frequently, the characteristic of the problem that makes Monte Carlo integration preferable to numerical analysis is the complexity of the P.D.F. $p(\underline{x})$. Often it is not possible or convenient to write down $p(\underline{x})$ in a closed-form but one is able to sample from it.

2.3 Sampling from a given probability distribution

The problems discussed in this section are concerned with generating a random sample $\{\underline{x}\}_j$ of size N from a given P.D.F. $p(\underline{x})$. The term 'random sample' implies that the joint P.D.F. $p(\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_N)$ can be written as

$$p(\underline{x}_1, \underline{x}_2, \cdots, \underline{x}_N) = \prod_{j=1}^{N} p(\underline{x}_j) \; ; \qquad (2.13)$$

i.e. $\underline{x}_i$ and $\underline{x}_j$, for $i \neq j$, are independent.

2.3.1 A method of generating random samples from arbitrary P.D.F.

It is the object of this section to show that it is sufficient to employ a random number generator which produces variates uniformly distributed in the interval (0,1) in order to produce a scalar random sample $\{x\}_j$ with an arbitrary probability distribution F(x).

We recall the general definition of a probability distribution P(x) as

$$dP(x) \triangleq p(x)\, dx \qquad (2.14)$$

where p(x) is the P.D.F. of x.

The space of the experiment, using a uniform random number generator, consists of all points in the unit interval (0,1). All intervals are events with probability equal to their length

$$\text{Prob}\{ 0 \leq \zeta_j \leq \zeta \} = \zeta \; . \qquad (2.15)$$

Because F(x) is assumed to be a probability distribution it must be a continuous, monotone increasing function $\zeta = F(x)$ such that $F(-\infty) = 0$ and $F(+\infty) = 1$. With $x = F^{-1}(\zeta)$ its inverse, we have $F[F^{-1}(\zeta)] = \zeta$ for any ζ in the interval (0,1). A variate x_j is so defined that

$$x_j = F^{-1}(\zeta_j) \qquad (2.16)$$

where ζ_j is an outcome of the preceding experiment. We now have to show that the distribution $P(x)$ of x_j equals $F(x)$.

By definition, the distribution function of the variate x_j is given by $P(x) = \text{Prob}\{x_j \leq x\}$ defined for any number $-\infty \leq x \leq +\infty$. If we replace x_j from eqn.(2.16) we have $P(x) = \text{Prob}\{F^{-1}(\zeta_j) \leq x\}$. The inequality may be replaced by $\zeta_j \leq F(x)$ and therefore

$$P(x) = \text{Prob}\{\zeta_j \leq F(x)\}. \tag{2.17}$$

But ζ_j is uniformly distributed over the interval (0,1) and hence from eqns.(2.14) and (2.15)

$$P(x) = \int_{-\infty}^{F(x)} p(\zeta)\, d\zeta = \int_{0}^{F(x)} d\zeta = F(x). \tag{2.18}$$

This completes the proof that x_j possesses the distribution $P(x) = F(x)$ and shows that there always exists a procedure such that for an arbitrary realization ζ_j there is a x_j whose distribution is $P(x)$.

We shall be mainly concerned with random samples from the normal P.D.F. denoted by

$$p(x) = n(x; \mu, \Sigma). \tag{2.19}$$

The mean μ and the variance Σ specify the normal P.D.F. entirely. The uniformly distributed variate ζ_j is transformed to x_j' using an approximation to the inverse normal integral, see eqn. (2.16). A subroutine, called GMIST(K), is used in the available computer system (IBM 7090) to perform this transformation. Advantage of the symmetry of the normal distribution is taken such that only the inverse normal integral curve need be fitted in the range $0.5 \leq \zeta \leq 1$. To get a uniform accuracy the interval [0.5 ... 1.0] is subdivided into three

segments, $[0.5 \ldots 0.75]$, $[0.75 \ldots 1. - 2^{-10}]$ and $[1. - 2^{-10} \ldots 1. - 2^{-27}]$, and different interpolation methods are used in each of these sub-intervals. Finally we set $x_j = \mu + \Sigma^{1/2} x'_j$ implying that $\{x\}_j$ is drawn from $p(x)$ of eqn.(2.19).

The particular form of $p(x)$ may allow a more efficient procedure to generate $\{x\}_j$ than the general method we mentioned. Butler[39] elaborates on this subject.

2.3.2 Generation of an n-variate normal random sample $\{\underline{x}\}_j$.

An important topic to be discussed here and used in subsequent chapters is the generation of a vector random sample $\{\underline{x}\}_j$ where $\underline{x}_j$ is an n-dimensional random vector from the n-variate normal P.D.F. $p(\underline{x}) = n(\underline{x}; \underline{\mu}, \Sigma)$.

First, we consider the random vector $\underline{\xi}_j$ whose P.D.F. is

$$p(\underline{\xi}) = n(\underline{\xi}; \underline{0}, I) \tag{2.20}$$

where I is the identity matrix. The components of the random vector $\underline{\xi}_j$ are obtained by the n-fold use of GMIST(K). Then, the random vector $\underline{x}_j$ is found through the following transformation

$$\underline{x}_j = \underline{\mu} + A \ \underline{\xi}_j \ . \tag{2.21}$$

Because of the symmetry property of the covariance matrix Σ the elements of A above the main diagonal are set equal to zero. The remaining $\frac{1}{2} n(n+1)$ elements are determined from Σ as follows. Using eqn.(2.21) in the defining equation of the covariance matrix Σ ,

$$\Sigma \triangleq E[\underline{x}\,\underline{x}^T] - E[\underline{x}]\;\; E^T[\underline{x}] \quad , \tag{2.22}$$

yields

$$\Sigma = A\,A^T. \tag{2.23}$$

From eqn.(2.23) ensues a system of $\frac{1}{2}n(n+1)$ quadratic equations for the unknown, nonzero elements of A.

Supplementary to remark (4) of section 2.2.3 the evaluation of a multidimensional integral using Monte Carlo techniques does not involve any additional difficulties compared with the scalar case as far as the generation of a vector random sample $\{\underline{x}\}_j$ is concerned.

2.4 Multi-stage Monte Carlo techniques using adaptive control variate estimators

An important step in the design of the sampling experiment is concerned with variance reduction techniques. They are usually classified within the following six groups:

(1) Stratified sampling

(2) Importance sampling

(3) Control variate methods

(4) Antithetic variate methods

(5) Regression methods

(6) Methods using orthonormal functions.

A description of how each one is used by itself in connection

with the evaluation of a scalar integral of the form of eqn. (2.1) is contained in references [26], [27] and [36]. Because these methods are quite general the greatest gains in variance reduction are only obtained after a careful adaptation of these techniques to the specific properties of the problem considered. That is, these methods cannot be applied in a routine manner.

Here, a new multi-stage control variate estimator is developed for the evaluation of a scalar integral. The concept of a multi-stage sampling scheme was originally proposed by Kahn[37]. Marshall[38] took up this idea in the context of importance sampling. More recently Clark[41] and Pugh[42] pursued this work again in connection with importance sampling. The limited scope for applying importance sampling in our work with nonlinear stochastic processes as well as some open questions such as the combination of various estimates, the determination of sample sizes, accuracy problems and convergence properties, motivated our own work in this field.

The following section 2.4.1 serves as a brief introduction into the control variate method. The first order method of section 2.4.2 is an extension of reference[42]. The problem of combining several correlated estimates is contained in section 2.4.3. A new second order method for finding the optimal values of the parameters in the control variate function is introduced in section 2.4.4. The computational procedures for implementing both methods are summarized in section 2.4.5. Finally, a two-stage estimator is discussed in section 2.4.6.

2.4.1 The control variate method

The problem to be discussed is again the estimation of θ as explained in connection with eqn. (2.1). Rather than estimating $\hat{\theta}_N$ with eqn. (2.3) the concept of the control variate method is to break eqn. (2.1) into two parts such that

$$\theta = \int_X [f(x) - \phi(x)]\ p(x)\ dx + \int_X \phi(x)\ p(x)\ dx\ . \qquad (2.24)$$

The first integral is estimated by the crude Monte Carlo method while the second one is found by analytical integration. For a successful variance reduction of the new estimate $\hat{\theta}_N$, see eqn. (2.26), the control variate function $\phi(x)$ must satisfy the following two conditions:

(C1) $\phi(x)$ must be a close approximation to $f(x)$.

(C2) $\phi(x)$ must be amenable to exact integration such that

$$\theta_a = \int \phi(x)\ p(x)\ dx \qquad (2.25)$$

is known analytically.

The control variate estimator based on eqn. (2.24) now takes the form

$$\hat{\theta}_N = N^{-1} \sum_j^N [f(x_j) - \phi(x_j)] + \theta_a\ . \qquad (2.26)$$

The random sample $\{x\}_j$ is drawn from the P.D.F. $p(x)$. The middle term in eqn. (2.26), $\phi(x_j)$, is referred to as control variate of $f(x_j)$. If condition (C1) is satisfied the difference to be estimated by crude Monte Carlo is small and hence the sampling variance of the new estimate $\hat{\theta}_N$, $\mathrm{var}(\hat{\theta}_N)$, is reduced compared with eqn. (2.7). Indeed, the reduced sampling variance is found from eqn. (2.26) to be

$$\mathrm{var}(\hat{\theta}_N) = \mathrm{var}(\hat{\theta}_{1,N}) + \mathrm{var}(\hat{\theta}_{2,N}) - 2\,\mathrm{cov}(\hat{\theta}_{1,N}, \hat{\theta}_{2,N}) \tag{2.27}$$

where we use the definitions

$$\hat{\theta}_{1,N} \triangleq N^{-1} \sum_{j}^{N} f(x_j) \tag{2.28}$$

and

$$\hat{\theta}_{2,N} \triangleq N^{-1} \sum_{j}^{N} \phi(x_j) \,. \tag{2.29}$$

For a successful variance reduction the following condition, derived from eqn. (2.27), must be satisfied

$$2\,\mathrm{cov}(\hat{\theta}_{1,N}, \hat{\theta}_{2,N}) > \mathrm{var}(\hat{\theta}_{2,N}) \,. \tag{2.30}$$

Ideally, condition (C1) leads to $f(x_j) \equiv \phi(x_j)$. This implies zero sampling variance for the estimator (2.26). Since none of the terms in eqn. (2.27) is known we have to replace them by their estimates as explained in connection with eqns. (2.7) and (2.8) in order to calculate confidence limits for the improved estimate $\hat{\theta}_N$ based on eqn. (2.26).

2.4.2 A first order gradient method of adaptive Monte Carlo using control variates

The difficulty inherent in the control variate method discussed so far is that one is often not able to specify a priori what a good control variate function $\phi(x)$ is, or even if a good choice can be made with regard to a specific class of functions to be used, the best parameter values are difficult to determine. In this situation an adaptive

sampling procedure may significantly improve the control variate method. We use the specification 'adaptive' to express that use is made of the information contained in the sample as it is collected to sequentially improve the efficiency of the sampling procedure itself. The smaller the sampling variance of $\hat{\theta}_N$ for a given sample size N the better the procedure.

The technique presented in this section is a multi-stage sampling procedure. Instead of the fixed control variate function $\phi(x)$ used in eqn. (2.24) we now consider a sequence of p control variate functions $\phi(x, \underline{\alpha}_p)$, $(p = 1, 2, \ldots)$, of some parametric family of functions. The sampling procedure starts at $p = 1$ where we choose an arbitrary value for the m-dimensional parameter vector $\underline{\alpha}_1$.

At each stage p we have to solve the following three problems:

(1) Estimate the value θ of the integral (2.1) by

$$\hat{\theta}_p = N^{-1} \sum_j^N [f(x_j) - \phi(x_j, \underline{\alpha}_p)] + \int_X \phi(x, \underline{\alpha}_p)\, p(x)\, dx \qquad (2.31)$$

with a random sample $\{x\}_j$ of size N drawn from p(x).
(For notational convenience we omit the subscript N and use $\hat{\theta}_p$ instead of $\hat{\theta}_{p,N}$.)

(2) Find a new value $\underline{\alpha}_{p+1}$ such that the sampling variance of $\hat{\theta}_{p+1}$, $\mathrm{var}(\hat{\theta}_{p+1})$, is less than $\mathrm{var}(\hat{\theta}_p)$.

(3) Determine a suitable weighting sequence $a_i, (i=1,2,\ldots,p)$, in order to obtain the combined estimate $\hat{\hat{\theta}}_p = \sum_i^p a_i \hat{\theta}_i$ such that $\mathrm{var}(\hat{\hat{\theta}}_p)$ is minimal. The discussion of this topic is contained in section 2.4.3.

According to our criterion of an efficient sampling procedure the optimal value of $\underline{\alpha}_p$, denoted by $\underline{\alpha}^o$, minimizes the sampling variance $\text{var}(\hat{\theta}_p)$ of the estimate $\hat{\theta}_p$; that is $\underline{\alpha}^o$ minimizes the expression

$$\text{var}(\hat{\theta}_p) = N^{-1} \quad [\text{var}(f(x)-\phi(x,\underline{\alpha}_p))] \ . \tag{2.32}$$

Omitting the terms in eqn. (2.32) which are independent of $\underline{\alpha}_p$ the following functional $F(\underline{\alpha}_p)$ has to be minimized w.r.t. $\underline{\alpha}_p$:

$$F(\underline{\alpha}_p) \triangleq E[\phi^2(x, \underline{\alpha}_p)] \ - \ (E[\phi(x, \underline{\alpha}_p)])^2 \tag{2.33}$$

$$-2E[\phi(x,\underline{\alpha}_p) f(x)] \ + \ 2E[\phi(x, \underline{\alpha}_p)] \ E[f(x)] \ .$$

A close examination of $F(\underline{\alpha}_p)$ shows that the computation of $\underline{\alpha}^o$ based on minimizing eqn. (2.33) presupposes the knowledge of θ which is to be estimated. To avoid an arbitrary guess of $\underline{\alpha}_p$, the proposed sequential estimation procedure based on estimating the gradient of $F(\underline{\alpha}_p)$ follows the path of steepest descent. This steepest descent is the direction of the negative of the gradient vector $\underline{F}_{\alpha_p}$ of $F(\underline{\alpha}_p)$, defined by

$$\underline{F}_{\alpha_p} \triangleq \left[\frac{\partial F(\underline{\alpha}_p)}{\partial \alpha_p(1)} \ , \ \ldots \ , \ \frac{\partial F(\underline{\alpha}_p)}{\partial \alpha_p(m)} \right] \ . \tag{2.34}$$

Now, in eqn. (2.33) $\underline{\alpha}_p$ is assumed to be known and we seek the $\underline{\alpha}_{p+1}$ which reduces the value of $F(\underline{\alpha}_{p+1})$ compared with $F(\underline{\alpha}_p)$. Hence the direction to proceed from the known point $\underline{\alpha}_p$ in the m-dimensional parameter space is the direction $-\underline{F}_{\alpha_p}$ at that point. Assuming regularity conditions the r:th component of the gradient $\underline{F}_{\alpha_p}$, denoted by $F_{\alpha_p}(r)$,

is obtained from eqns. (2.33) and (2.34) by

$$F_{\alpha_p}(r) = 2\,E[\phi(x,\underline{\alpha}_p)\,\frac{\partial\phi(x,\underline{\alpha}_p)}{\partial\alpha_p(r)}] - 2E[\phi(x,\underline{\alpha}_p)]\,E[\frac{\partial\phi(x,\underline{\alpha}_p)}{\partial\alpha_p(r)}]$$

$$- 2\,E[f(x)\,\frac{\partial\phi(x,\underline{\alpha}_p)}{\partial\alpha_p(r)}] + 2\,E[f(x)]\;E\,[\frac{\partial\phi(x,\underline{\alpha}_p)}{\partial\alpha_p(r)}] \quad . \tag{2.35}$$

As the integrals in eqn. (2.35) not involving f(x) are amenable to analytic integration it is reasonable to estimate only the remaining terms by a sample $\{x\}_j$ of size N. The estimator for the gradient $\underline{F}_{\alpha_p}$ takes the form

$$\hat{F}_{\alpha_p}(r) = 2E[\phi(x,\underline{\alpha}_p)\,\frac{\partial\phi(x,\underline{\alpha}_p)}{\partial\alpha_p(r)}] - 2E[\phi(x,\underline{\alpha}_p)]\;E[\frac{\partial\phi(x,\underline{\alpha}_p)}{\partial\alpha_p(r)}] \tag{2.36}$$

$$- \frac{2}{N}\sum_{j}^{N}[\frac{\partial\phi(x_j,\underline{\alpha}_p)}{\partial\alpha_p(r)}\;f(x_j)] + [\frac{2}{N}\sum_{j}^{N}f(x_j)]\,E[\frac{\partial\phi(x,\underline{\alpha}_p)}{\partial\alpha_p(r)}] \quad .$$

A close estimate of the optimal vector $\underline{\alpha}^{o}$ can be found by applying the stochastic approximation algorithm

$$\hat{\underline{\alpha}}_{p+1} = \hat{\underline{\alpha}}_p - \varepsilon_p\;\hat{\underline{F}}_{\alpha_p} \quad . \tag{2.37}$$

The convergence conditions for this algorithm are given by Dvoretzky[43]. The important condition for convergence with probability one is that the step size ε_p should progressively decrease in such a way that the stochastic variation in the correction term dies out.

If in actual practice it is not possible to verify whether the necessary and sufficient conditions for convergence with probability one of algorithm (2.37) are met it seems to be necessary to resort to the following updating procedure, proposed by Pugh[42]:

$$\hat{\underline{a}}_{p+1} = \hat{\underline{a}}_p - \lambda \frac{\hat{\underline{F}}_{\underline{a}_p}}{\| \hat{\underline{F}}_{\underline{a}_p} \|} \tag{2.38}$$

where

$$\| \hat{\underline{F}}_{\underline{a}_p} \| \triangleq \left[\sum_r^m \hat{F}_{a_p}(r)^2 \right]^{1/2} \tag{2.39}$$

and λ is a suitable constant length of step in the parameter space. A very important problem is the selection of the step size λ. Moreover, as in any gradient technique, we may only achieve a relative minimum with algorithm (2.38). Indeed, we may only 'probably' achieve any minimum since the gradient is not known exactly but only estimated at each stage p.

No result is available to prove that the stochastic approximation procedure (2.38) converges. However, by the strong law of large numbers we have

$$\text{Prob} \left\{ \lim_{p \to \infty} (\hat{\underline{F}}_{\underline{a}_p} - \underline{F}_{\underline{a}}) = 0 \right\} = 1 \tag{2.40}$$

if in each stage p the gradient $\hat{\underline{F}}_{\underline{a}_p}$ is estimated with a sample $\{x\}_{j,p}$ of increasing size $N_p = pN$. ~~Eqn. (2.40) then follows because~~ Also

$$\text{var}(\hat{\underline{F}}_{\underline{a}_p}) \longrightarrow 0 \tag{2.41}$$

for $p \to \infty$. In section 2.4.4 we introduce a new second order algorithm

where convergence with probability one follows from eqn.(2.40).

The estimator (2.36) for the gradient $\hat{\underline{F}}_{\underline{a}_p}$ may be improved by means of the antithetic variate method. A detailed discussion of this method follows in chapter three in the context of the nonlinear state prediction problem. The basic idea of this method is to use instead of the random sample $\{x\}_j$, a linear combination of two negatively correlated samples such that the expectation remains unchanged. If x_j^+ is a variate from a symmetric P.D.F. $p(x)$ with mean μ then the antithetic variate x_j^- is given by

$$(x_j^- - \mu) = -(x_j^+ - \mu) \tag{2.42}$$

and the new estimator for the r:th component of $\hat{\underline{F}}_{\underline{a}_p}$ is given by

$$\hat{F}_{a_p}(r) = 2E[\phi(x,\underline{a}_p) \frac{\partial\phi(x,\underline{a}_p)}{\partial a_p(r)}] - 2E[\phi(x,\underline{a}_p)]\ E[\frac{\partial\phi(x,\underline{a}_p)}{\partial a_p(r)}]$$

$$-N^{-1} \sum_j^N [\frac{\partial\phi(x_j^+,\underline{a}_p)}{\partial a_p(r)} f(x_j^+) + \frac{\partial\phi(x_j^-,\underline{a}_p)}{\partial a_p(r)} f(x_j^-)]$$

$$+N^{-1} \sum_j^N [f(x_j^+) + f(x_j^-)]\ E[\frac{\partial\phi(x,\underline{a}_p)}{\partial a_p(r)}] \quad . \tag{2.43}$$

This estimator (2.43) is to be preferred to eqn.(2.36) if the terms to be estimated are predominantly odd in x.

2.4.3 Averaging several correlated estimates

The multi-stage control variate procedure introduced in section 2.4.2 yields a sequence of p scalar estimates $\hat{\theta}_1, \hat{\theta}_2, \ldots \hat{\theta}_p$. Lindley[58] proves the theorem that the combined, minimum variance estimate $\hat{\hat{\theta}}_p$ is given by

$$\hat{\hat{\theta}}_p = \frac{\sum_i^p w_i \hat{\theta}_i}{\sum_i^p w_i} = \sum_i^p a_i \hat{\theta}_i \qquad (2.44)$$

where $w_i^{-1} = \text{var}(\hat{\theta}_i)$, provided

(1) the P.D.F. of the random sample $\{\theta_i\}_j$ used to estimate $\hat{\theta}_i$ is normal with known variance, and

(2) $\hat{\theta}_i$ and $\hat{\theta}_j$ $(i \neq j)$ are uncorrelated.

This result is now extended in two steps. First we drop condition 2 and show that the combined, minimum variance estimate $\hat{\hat{\theta}}_p$ is given by

$$\hat{\hat{\theta}}_p = \frac{\underline{z}_p^T \, V_p^{-1}}{\underline{z}_p^T \, V_p^{-1} \, \underline{z}_p} \, \hat{\underline{\theta}}_p = \underline{a}_p^T \, \hat{\underline{\theta}}_p \qquad (2.45)$$

where

(1) $$\hat{\underline{\theta}}_p^T \triangleq [\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_p] \qquad (2.46)$$

is a p-dimensional row vector whose components are the p scalar estimates $\hat{\theta}_i$ $(i = 1,2,\ldots p)$;

(2) $$\underline{z}_p^T \triangleq [1,1,\ldots,1] \qquad (2.47)$$

is considered here as a p-dimensional identity vector;

(3) the sampling covariance matrix $V_p \triangleq \text{var}(\hat{\underline{\theta}}_p)$ is assumed to be

known. If $\hat{\theta}_r$ and $\hat{\theta}_p$ $(r \neq p)$ are uncorrelated; i.e. if the correlation factor ρ_{rp}, defined by

$$\rho_{rp} \triangleq \frac{\mathrm{cov}(\hat{\theta}_r, \hat{\theta}_p)}{[\mathrm{var}(\hat{\theta}_r)\mathrm{var}(\hat{\theta}_p)]^{1/2}} \quad , \tag{2.48}$$

is equal to zero then (2.45) is equivalent to eqn.(2.44).

In order to prove that the weighting vector $\underline{a}_p^T \triangleq [\, a_1, \ldots, a_p]$ is given by eqn.(2.45) we have to minimize $\mathrm{var}(\hat{\hat{\theta}}_p)$ w.r.t. $\underline{a}_p$. Since each $\hat{\theta}_i, (i=1,2,\ldots,p)$, is an unbiased estimate of θ the solution has to satisfy the condition

$$\underline{z}_p^T \underline{a}_p = 1 \ . \tag{2.49}$$

Thus, we have to minimize the function $J(\underline{a}_p)$, defined by

$$J(\underline{a}_p) \triangleq \underline{a}_p^T V_p \underline{a}_p - \lambda(\underline{z}_p^T \underline{a}_p - 1 \ ), \tag{2.50}$$

w.r.t. $\underline{a}_p$. The differentiation of eqn. (2.50) w.r.t. $\underline{a}_p$ yields

$$V_p \underline{a}_p = \lambda \underline{z}_p . \tag{2.51}$$

Premultiplying both sides with $\underline{z}_p^T V_p^{-1}$ yields

$$\underline{z}_p^T \underline{a}_p = \lambda \underline{z}_p^T V_p^{-1} \underline{z}_p = 1 \tag{2.52}$$

where the last equality follows from eqn.(2.49). Therefore

$$\lambda = (\underline{z}_p^T V_p^{-1} \underline{z}_p)^{-1} \tag{2.53}$$

and hence from eqn.(2.51) follows

$$\underline{a}_p = \frac{V_p^{-1} \underline{z}_p}{\underline{z}_p^T V_p^{-1} \underline{z}_p} \tag{2.54}$$

as postulated in eqn.(2.45). It is easy to show that the sampling variance of the combined estimate $\hat{\hat{\theta}}_p$ is given by

$$\text{var}(\hat{\hat{\theta}}_p) = (\underline{z}_p^T V_p^{-1} \underline{z}_p)^{-1} . \qquad (2.55)$$

The computation of $\hat{\hat{\theta}}_p$ with eqn.(2.45) requires the assumption that V_p is known. Since V_p is usually unknown the weighting vector $\underline{a}_p$ based on eqn.(2.54) only yields meaningful results if we can estimate the elements of V_p in the usual way (see eqns.(2.7) and (2.8) for the scalar case), replace V_p by its estimate $\hat{V}_p$ and ignore distributional changes that might result. This however is only likely to be satisfactory if the sample size N of each iteration is large enough so that the error in replacing V_p by $\hat{V}_p$ is small. This requirement may help to decide upon the sample size N in the multi-stage sampling procedure introduced in section 2.4.2.

2.4.4 A second order gradient method of adaptive Monte Carlo using control variates

To the conditions introduced in section 2.4.1 for the control variate function $\phi(\underline{x}, \underline{\alpha})$, we now add a third one restricting $\phi(x, \underline{\alpha})$ to depend linearly on the components $\alpha(r)$ of $\underline{\alpha}$, as in eqn.(2.56). $\gamma_r(x)$ are appropriate basis functions.

(C3) $$\phi(x, \underline{\alpha}) = \sum_r^m \alpha(r)\, \gamma_r(x). \qquad (2.56)$$

Substituting eqn.(2.56) into eqn. (2.33), the cost functional $F(\underline{\alpha})$ becomes a quadratic expression in $\underline{\alpha}$. Hence, the displacement between

an arbitrary point $\underline{\alpha}$ and the optimum $\underline{\alpha}^{\circ}$ where $F(\underline{\alpha})$, and the sampling variance $\mathrm{var}(\hat{\theta})$, takes its minimum w.r.t. $\underline{\alpha}$, is given by

$$\underline{\alpha}^{\circ} - \underline{\alpha} = - F_{\alpha\alpha}^{-1} \underline{F}_{\alpha} \, . \tag{2.57}$$

$\underline{F}_{\alpha}$ is again the m-dimensional gradient defined by eqn.(2.34). Using eqn.(2.56) its r:th component is given by

$$F_{\alpha}(r) = 2E[\phi(x,\underline{\alpha})\, \gamma_r(x)] - 2E[\phi(x,\underline{\alpha})]\; E[\gamma_r(x)]$$

$$- 2E[f(x)\, \gamma_r(x)] + 2E[f(x)]\; E[\gamma_r(x)] \, . \tag{2.58}$$

$F_{\alpha\alpha}$ is the matrix of second order derivatives of $F(\underline{\alpha})$ w.r.t. $\underline{\alpha}$. The (r,p):th element is obtained after differentiating eqn.(2.58) w.r.t. $\alpha(p)$ as

$$F_{\alpha\alpha}(r,p) = 2E[\gamma_p(x)\, \gamma_r(x)] - 2E[\gamma_p(x)]\; E[\gamma_r(x)] \, . \tag{2.59}$$

The determination of $\underline{F}_{\alpha}$ requires the integration of f(x). In general, this can be done by Monte Carlo methods; i.e. $\underline{F}_{\alpha}$ has to be replaced by its estimate $\hat{\underline{F}}_{\alpha}$ which is obtained by an estimator similar to eqn.(2.36) or (2.43). Thus, the one step convergence indicated by eqn.(2.57) can no longer be maintained and has to be replaced by the following multi-stage updating procedure for $\hat{\underline{\alpha}}_p$

$$\hat{\underline{\alpha}}_{p+1} = \hat{\underline{\alpha}}_p - F_{\alpha_p \alpha_p}^{-1} \hat{\underline{F}}_{\alpha_p} \, . \tag{2.60}$$

The subscript p refers again to the number of iterations performed.

The novel feature of eqn.(2.60) is the deterministic property of the matrix $F_{\alpha_p \alpha_p}$ given by eqn.(2.59). Since its elements are

independent of $\underline{\alpha}_p$ and do not involve f(x) they can be determined analytically before the sampling procedure starts.

An estimate of $\underline{\alpha}^o$ can also be found by linear regression. Unlike our procedure this approach requires the inversion of a matrix which is itself composed of random variables.

Let us denote the estimate of $\underline{F}_\alpha$ for the time being by $(\hat{\underline{F}}_\alpha)_N$. By the strong law of large numbers, discussed in section 2.2.2, the estimate $(\hat{\underline{F}}_\alpha)_N$ tends with probability one to $\underline{F}_\alpha$ as $N \to \infty$; that is

$$\text{Prob} \left\{ \lim_{N \to \infty} \left[(\hat{\underline{F}}_\alpha)_N - \underline{F}_\alpha \right] = 0 \right\} = 1 \quad . \tag{2.61}$$

In order to keep the sample size N of the first iteration finite eqn. (2.61) is not very practical. Choosing a basic sample size N the gradient $\underline{F}_\alpha$ is estimated in an iterative manner. Using a series of p random samples $\{x\}_{j,p}$ of increasing size $N_p = p\,N$ implies, after a sufficient number of iterations, that $(\hat{\underline{F}}_{\alpha_p})_{N_p}$ is arbitrarily close to $\underline{F}_{\alpha_p}$ and hence using eqns.(2.57) and (2.61)

$$\text{Prob} \left\{ \lim_{p \to \infty} \left[\underline{\alpha}^o - \hat{\underline{\alpha}}_p = - F_{\alpha_p \alpha_p}^{-1} (\hat{\underline{F}}_{\alpha_p})_{N_p} \right] = \underline{\alpha}^o - \underline{\alpha} \right\} = 1 \, . \tag{2.62}$$

Thus the series of estimates $\hat{\underline{\alpha}}_1$, $\hat{\underline{\alpha}}_2$, ..., $\hat{\underline{\alpha}}_p$ converges with probability one to $\underline{\alpha}^o$ if $p \to \infty$.

For computational reasons the concept of an increasing sample space N_p is not very attractive. We therefore propose an alternative method where each iteration is based on a sample of constant size N. Here, convergence is obtained through the introduction of an auxiliary, m-dimensional vector $\underline{c}$. In the p:th iteration we estimate $\hat{\underline{c}}_p$ and then

combine it with all previous estimates $\hat{\underline{c}}_1, \hat{\underline{c}}_2, \ldots, \hat{\underline{c}}_{p-1}$ to get a weighted average $\hat{\hat{\underline{c}}}_p$ such that $\mathrm{var}(\hat{\hat{\underline{c}}}_p) \rightarrow 0$ as $p \rightarrow \infty$.

Solving eqn.(2.60) for $\hat{\underline{F}}_{\underline{\alpha}_p}$ yields

$$\hat{\underline{F}}_{\underline{\alpha}_p} = F_{\alpha_p \alpha_p} (\hat{\underline{\alpha}}_p - \hat{\underline{\alpha}}_{p+1}) . \tag{2.63}$$

The vector $\hat{\underline{c}}_p$ is defined by

$$\hat{\underline{c}}_p \triangleq F_{\alpha_p \alpha_p} \hat{\underline{\alpha}}_{p+1} = F_{\alpha_p \alpha_p} \hat{\underline{\alpha}}_p - \hat{\underline{F}}_{\underline{\alpha}_p} . \tag{2.64}$$

As $\underline{c}$ is invariant w.r.t. $\underline{\alpha}$ (see fig. 2.1) it is feasible to combine the p estimates $\hat{\underline{c}}_i$, (i=1,2,...,p) to ensure convergence of the weighted estimate $\hat{\hat{\underline{c}}}_p$.

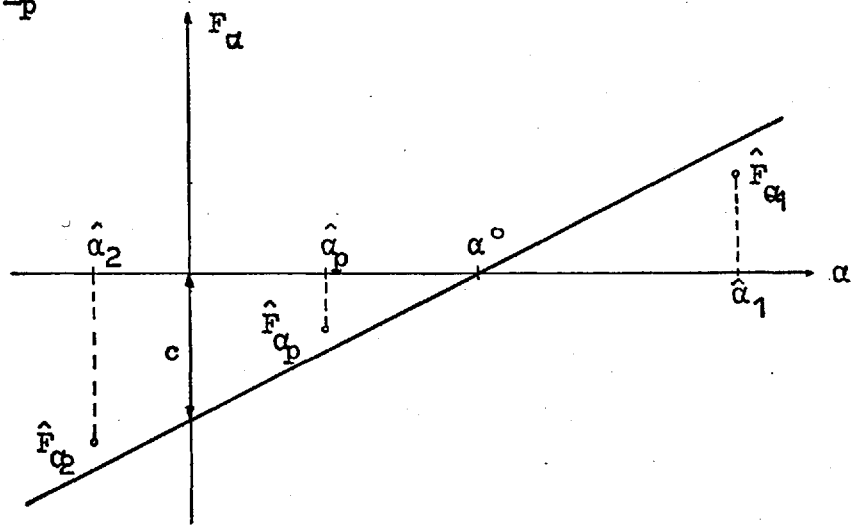

Fig. 2.1 Definition of the auxiliary constant c in the scalar case.

Let us denote the estimated sampling covariance matrix of the estimated gradient $\hat{\underline{F}}_{\underline{\alpha}_p}$ by $\hat{\mathrm{var}}(\hat{\underline{F}}_{\underline{\alpha}_p})$. Neglecting the correlation between $\hat{\underline{c}}_r$ and $\hat{\underline{c}}_p$, $(r \neq p)$, allows us to use eqn.(2.44). The combined

estimate $\hat{\underline{c}}_p$ is computed by

$$\hat{\underline{c}}_p = \left[\sum_i^p \hat{W}_i \right]^{-1} \sum_i^p \hat{W}_i \, \hat{\underline{c}}_i \tag{2.65}$$

where the matrix $\hat{W}_i$ is defined by

$$\hat{W}_i^{-1} = \hat{\mathrm{var}}(\hat{\underline{F}}_{\underline{a}_i}) \tag{2.66}$$

The sample size N in each stage p is constant but it has to be chosen sufficiently large such that the replacement of $\mathrm{var}(\hat{\underline{F}}_{\underline{a}_i})$ by its estimate $\hat{\mathrm{var}}(\hat{\underline{F}}_{\underline{a}_i})$ is admissible.

The increase in computing time due to the evaluation of eqn.(2.65) may be avoided if we use a deterministic weighting sequence. Denoting the r:th component of $\underline{c}_p$ by $c_p(r)$ the combined estimate $\hat{\hat{c}}_p(r)$ is computed by

$$\hat{\hat{c}}_p(r) = \sum_i^p a_i \, \hat{c}_i(r) \; ; \; r = 1,2,\ldots,m. \tag{2.67}$$

The weighting sequence $a_1, a_2, \ldots$ must satisfy eqn.(2.49). An admissible choice is

$$a_i = \frac{2i}{p(p+1)} \; , \quad i=1,2,\ldots,p. \tag{2.68}$$

Using the weighted estimate $\hat{\hat{\underline{c}}}_p$ the updated parameter $\hat{\underline{a}}_{p+1}$ is found with eqn.(2.64) to be

$$\hat{\underline{a}}_{p+1} = F_{a_p a_p}^{-1} \, \hat{\hat{\underline{c}}}_p. \tag{2.69}$$

The combination of the sequence $\hat{\underline{c}}_1, \hat{\underline{c}}_2, \ldots$ by means of eqn.(2.65) or (2.67) implies that $\mathrm{var}(\hat{\hat{\underline{c}}}_p) \to 0$ as $p \to \infty$ because $\hat{\hat{\underline{c}}}_p$ is estimated with a sample of increasing size $N_p = pN$. Therefore, by the strong law

of large numbers we have

$$\text{Prob} \left\{ \lim_{p \to \infty} (\hat{\underline{c}}_p - \underline{c}) = 0 \right\} = 1 \quad . \tag{2.70}$$

Since the sequence $\hat{\underline{c}}_1, \hat{\underline{c}}_2, \ldots$ converges with probability one to $\underline{c}$ and the matrix $F_{\alpha\alpha}^{-1}$ is a deterministic multiplier we conclude that the sequence $\hat{\underline{\alpha}}_1, \hat{\underline{\alpha}}_2, \ldots$ obtained from eqn.(2.69) converges with probability one to $\underline{\alpha}^o$.

2.4.5 Computational procedures

The object of this section is to summarize the various implementations of the multi-stage, adaptive control variate method.

Algorithm 1: First order method; section 2.4.2

(1) Choose a suitable class of control variate functions satisfying (C1) and (C2) of section 2.4.1. Set $\underline{\alpha}_1$ to an arbitrary value.

(2) In the p:th iteration, generate a random sample $\{x\}_{j,p}$ of size N from the given P.D.F. p(x).

(3) Estimate $\hat{\theta}_p$ using eqn.(2.31) and its sampling variance, $\text{var}(\hat{\theta}_p)$, using eqn.(2.32).

(4) Increase the sample size to $N_p = pN$, estimate the gradient $\hat{\underline{F}}_{\alpha_p}$ with eqn.(2.36) or (2.43) and update $\underline{\alpha}_p$ using eqn.(2.38).

(5) Combine $\hat{\theta}_p$ with the previous p-1 estimates $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_{p-1}$ as described in section 2.4.3 to compute $\hat{\hat{\theta}}_p$ by eqn.(2.45) and $\text{var}(\hat{\hat{\theta}}_p)$ by eqn.(2.55).

(6) Increase p to p+1 and repeat steps (2) ...(5) until the confidence

interval for $\hat{\hat{\theta}}_p$ is less than a preset accuracy level.

The second order method of section 2.4.4 is summarized in algorithm 2. Since we found various methods to ensure convergence of eqn.(2.60) we shall refer to them as algorithms 2a, 2b, and 2c respectively.

Algorithm 2a

(1) Choose a suitable class of control variate functions satisfying (C1), (C2) of section 2.4.1 and (C3) of section 2.4.4. Set $\underline{a}_1$ to an arbitrary value and compute $F_{\alpha\alpha}^{-1}$ analytically using eqn.(2.59).

(2) In the p:th iteration, generate a sample $\{x\}_{j,p}$ from the given p(x) of fixed size N.

(3) Estimate $\hat{\theta}_p$ using eqn.(2.31) and its sampling variance, $\mathrm{var}(\hat{\theta}_p)$, using eqn.(2.32).

(4) Increase the sample size to $N_p = pN$, estimate the gradient $\hat{\underline{F}}_{\underline{\alpha}_p}$ by Monte Carlo integration of eqn.(2.58) and update $\hat{\underline{a}}_p$ by eqn.(2.60).

(5) Combine $\hat{\theta}_p$ with the previous p-1 estimates $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_{p-1}$ as described in section 2.4.3, and compute $\hat{\hat{\theta}}_p$ by eqn.(2.45) and $\mathrm{var}(\hat{\hat{\theta}}_p)$ by eqn.(2.55).

(6) Increase p to p +1 and repeat steps (2) ...(5) until the confidence interval for $\hat{\hat{\theta}}_p$ is less than a preset value.

The increasing sample size N_p of step (4) is necessary to guarantee convergence of algorithm 2a. But it can be avoided if we use the auxiliary constant $\underline{c}$. Algorithms 2b and 2c differ from algorithm 2a

only in step (4). Different methods for this combination of the estimates $\hat{\underline{c}}_1 \ldots \hat{\underline{c}}_p$ are employed in each and we state these in the sequel.

Algorithm 2b

(4) Use the sample of step (2) to estimate $\hat{\underline{F}}_{\alpha_p}$ and $\hat{\text{var}}(\hat{\underline{F}}_{\alpha_p})$. Compute $\hat{\underline{c}}_p$ with eqn.(2.64) and combine it with the p-1 previous estimates to find $\hat{\hat{\underline{c}}}_p$ using eqn.(2.65). The updated value $\hat{\underline{\alpha}}_{p+1}$ is given by eqn.(2.69).

Algorithm 2c

(4) Use the sample of step (2) to estimate $\hat{\underline{F}}_{\alpha_p}$. Compute $\hat{\underline{c}}_p$ with eqn.(2.64) and combine it with the p-1 previous estimates $\hat{\underline{c}}_1, \ldots, \hat{\underline{c}}_{p-1}$ to find $\hat{\hat{\underline{c}}}_p$ using eqn.(2.67); i.e. a fixed weighting sequence. The updated value $\hat{\underline{\alpha}}_{p+1}$ is given by eqn.(2.69).

In order to compare the relative efficiency of two Monte Carlo methods let us introduce the following measures. The variance ratio η_v is defined by

$$\eta_v = \frac{\text{var}(\hat{\theta})_{\text{Method 1}}}{\text{var}(\hat{\theta})_{\text{Method 2}}} \quad ; \tag{2.71}$$

that is, the ratio of the sampling variances of the estimate $\hat{\theta}$ using methods 1 and 2 respectively. The labour ratio η_L is defined by

$$\eta_L = \frac{t_1}{t_2} \tag{2.72}$$

thus expressing that method 1 calls for t_1 and method 2 for t_2 units of computing time. The efficiency gain η, defined by

$$\eta = \eta_v \, \eta_L \quad , \tag{2.73}$$

measures the overall improvement of method 2 compared with method 1. The variance ratio depends on the problem and the Monte Carlo procedure used. In numerical examples η_v is usually easy to assess. The labour ratio depends not only on the Monte Carlo method but also on the computing machinery available. It is not necessarily meaningful to take t_1 and t_2 to be the respective numbers of times that the integral (2.1) has to be evaluated in each method. This is mainly due to the fact that the generation of random sample $\{x\}_j$ may call for considerably more computing time than basic operations like addition and multiplication. In particular, the generation of one variate x_j using GMIST(K), see section 2.3.1, takes on the IBM 7090 856 μsec compared with a multiplication requiring 20 μsec.

2.4.6 A two-stage control variate method

The problem of combining several estimates $\hat{\theta}_1,\ldots,\hat{\theta}_p$ requires the estimation of the sampling covariance matrix V_p between these variates; see section 2.4.3. This can be avoided by separating the 'learning' phase from the 'estimating' phase. Indeed, an arbitrary value of $\underline{a}_p$ in the control variate function $\phi(x,\underline{a}_p)$ yields an estimate $\hat{\theta}_p$ with a large sampling variance $\mathrm{var}(\hat{\theta}_p)$ compared with the value we obtain with the optimal $\underline{a}^o$. Hence, the first few iterations will usually give small contributions to the final weighted average $\hat{\hat{\theta}}_p$ and may therefore be neglected altogether. This observation leads quite naturally to a two-stage estimator. In the first stage, the learning

phase, we use the second order method of section 2.4.4 for p_1 iterations to find an estimate $\hat{\underline{\alpha}}_{p_1}$ of $\underline{\alpha}^o$ using eqn.(2.69). This value is subsequently used in stage two, the estimating phase, to determine $\hat{\theta}$ with eqn.(2.31) using $\hat{\underline{\alpha}}_{p_1}$ in the control variate function $\phi(x,\underline{\alpha}_p)$. Thus, the second stage consists of one iteration only. The loss of information during the first stage with respect to estimating $\hat{\theta}$ is offset by the reduced computing task.

2.5 A numerical example

Let us consider the following example used by Hammersley and Handscomb[36]. It deals with the evaluation of the scalar integral

$$\theta = \int_0^1 \frac{e^x - 1}{e - 1} p(x)\, dx = 0.418023 \qquad (2.74)$$

where $p(x)$ is the uniform P.D.F. defined in the interval $(0,1)$. If we choose the fixed control variate function in eqn.(2.26) to be

$$\phi(x) = x \qquad (2.75)$$

the variance reduction η_v, see eqn.(2.71), compared with the crude Monte Carlo estimator (2.3) is found to be

$$\eta_v = 60.4 \qquad (2.76)$$

An admissible parameterized control variate function of the form (2.56) is

$$\phi(x) = \alpha(1)\, x + \alpha(2) x^2 \qquad (2.77)$$

It is easily verified that conditions (C1), (C2) and (C3) of section 2.4 are satisfied by eqn.(2.77). The optimal values $\alpha^o(1)$ and $\alpha^o(2)$ are analytically found to be

$$\alpha^o(1) = 0.5 \qquad \alpha^o(2) = 0.485 \quad . \tag{2.78}$$

For the first order method of section 2.4.2 a suitable value of the convergence factor λ is found by trial and error to be $\lambda = 0.08$. The matrix $F_{\alpha\alpha}$ of eqn.(2.59) for the second order method of section 2.4.4 is

$$F_{\alpha\alpha} = \begin{bmatrix} 0.1667 & 0.1667 \\ 0.1667 & 0.1778 \end{bmatrix} \quad . \tag{2.79}$$

For the first iteration we set the components of $\underline{\alpha}$ as

$$\alpha(1) = 1.0 \qquad \alpha(2) = 1.0 \quad . \tag{2.80}$$

In the Monte Carlo computation each iteration is based on a sample of size N = 50. Starting at p =1 , iterations up to p=10 are performed and the sampling variances of $\hat{\hat{\theta}}_{10}$ compared. The gradient $\underline{F}_{\alpha}$ is computed with eqn.(2.43) because the integrand to be evaluated is predominantly odd. Unless otherwise stated the exhibited results are averages over ten ensembles. Thus, $\bar{\hat{\hat{\theta}}}_p$ denotes the average over ten values of $\hat{\hat{\theta}}_p$. Similarly $\overline{\hat{var}(\hat{\hat{\theta}}_p)}$ denotes the ensemble average of the sampling variances of $\hat{\hat{\theta}}_p$. The labour ratios stated have to be understood as approximate values. They are based on our numerical work executed on an IBM 7090. The numerator of the variance reduction factor (2.71) is always given by the sampling variance of the crude Monte Carlo estimator. In table 2.3 we compare the different

estimators for θ defined in eqn.(2.74).

	$\overline{\hat{\hat{\theta}}_{10}}$	$\overline{\hat{var}(\hat{\hat{\theta}}_{10})}$	η_V	η_L	η
Crude Monte Carlo; sec.2.2.3	0.412003	$1.601.10^{-4}$	-	-	-
First order method sec.2.4.2	0.417863	$1.737.10^{-7}$	920	0.1	92
Second order method; algorithm 2a;sec.2.4.4	0.418046	$2.560.10^{-8}$	6250	0.1	625
Algorithm 2b	0.418010	$3.029.10^{-8}$	5300	0.2	1060
Algorithm 2c	0.418036	$4.037.10^{-8}$	4000	0.4	1600

Table 2.3 Comparison of the multi-stage control variate estimators with the crude Monte Carlo method.

The stochastic convergence behaviour of the parameter $\alpha(1)$ in eqn.(2.77) is shown in fig. 2.2 for algorithm 1 and 2 for one particular experiment. The same sample $\{x\}_j$ is used for all four evaluations of $\hat{\alpha}_p(1)$.

In fig. 2.3 we plot the p ensemble estimates of the sampling variances, $\overline{\hat{var}(\hat{\theta}_p)}$, for the individual estimates $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_p$ after each iteration. The evaluation is based on eqn.(2.32). It can be

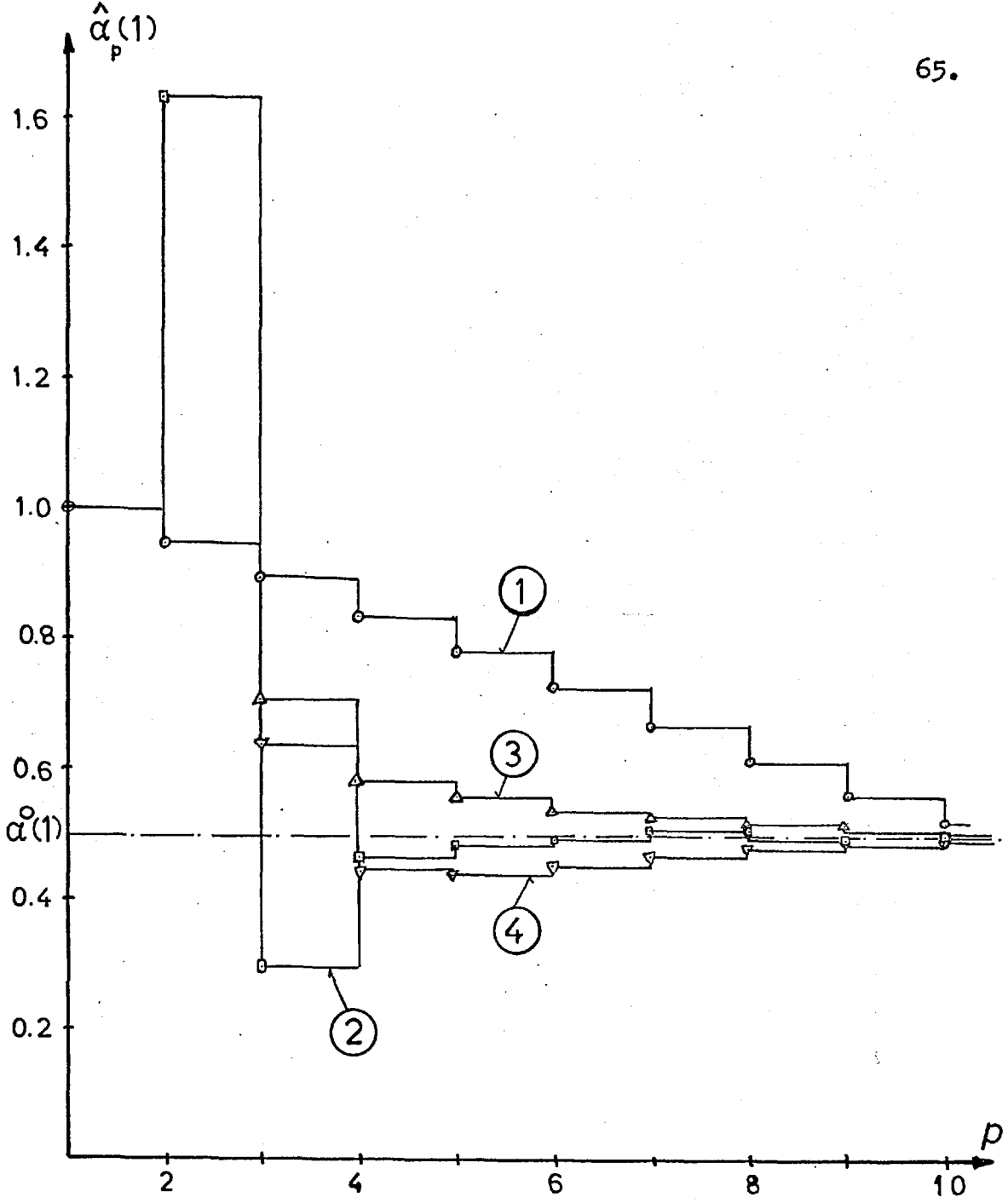

Fig.2.2 Convergence behaviour of the control variate function parameter $\hat{\alpha}_p(1)$ vs. the iteration number p of the multi-stage Monte Carlo estimation procedure.

1	algorithm 1	first order method
2	algorithm 2a	second order method
3	algorithm 2b	second order method
4	algorithm 2c	second order method

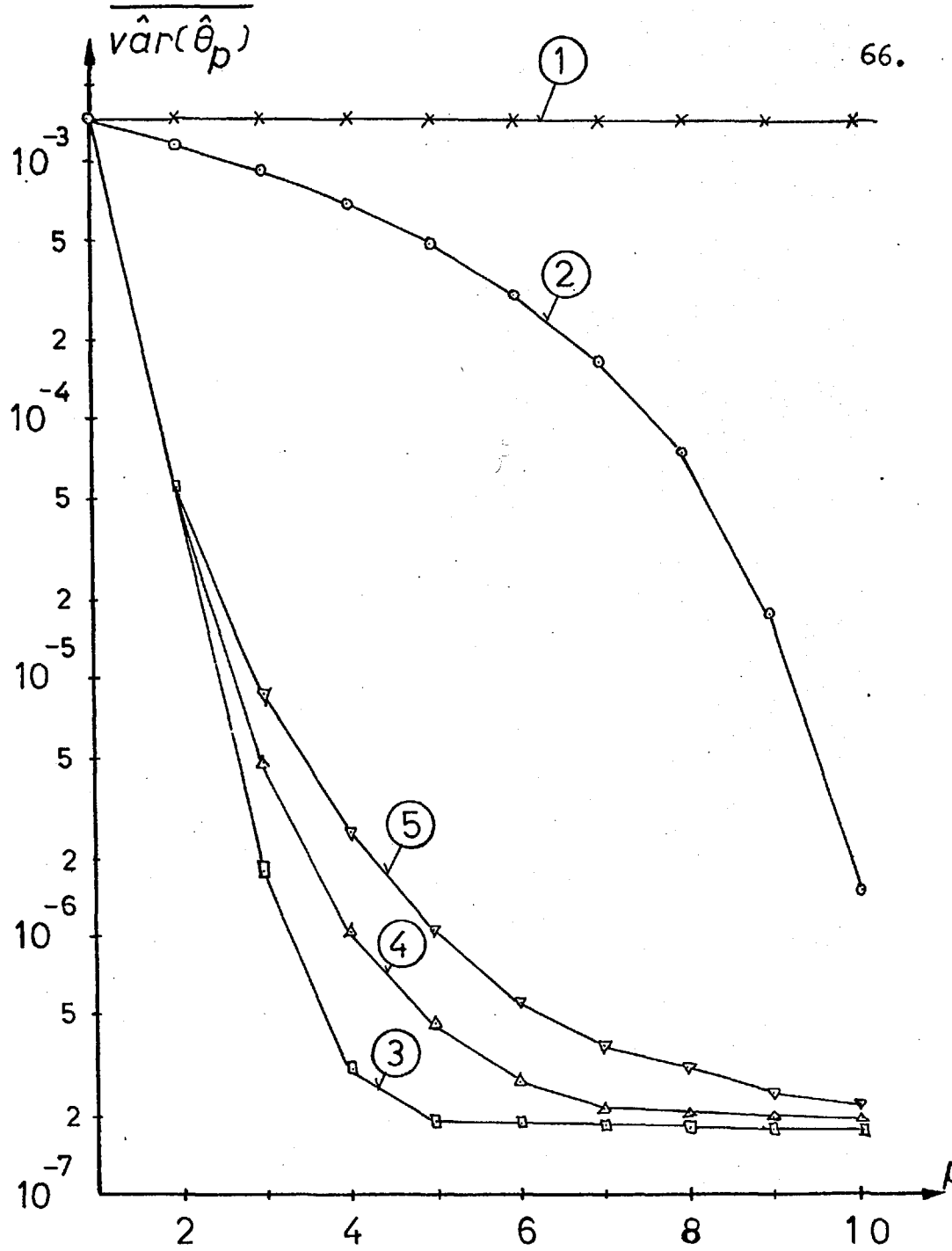

Fig.2.3 Ensemble values of the estimated sampling variance $\overline{\hat{var}(\hat{\theta}_p)}$ vs. the iteration number p of the multi-stage Monte Carlo estimation procedure.

1 crude Monte Carlo

2 algorithm 1 — first order method

3 algorithm 2a

4 algorithm 2b — second order method

5 algorithm 2c

seen that the second order method yields the minimum variance estimator after 3 to 4 iterations. This fast variance reduction makes it quite appropriate to use the two-stage estimator of section 2.4.6.

In order to compare the relative efficiency of the proposed algorithms we plot in fig. 2.4 a 'cost' function defined as product of computing time required per iteration times the ensemble value of the sampling variance of the combined estimate $\hat{\bar{\theta}}_p$, $\overline{\hat{\mathrm{var}}(\hat{\bar{\theta}}_p)}$, versus p. This graph allows us to establish the relative merits of the three modifications for the second order method. Increasing the sample space for estimating $\hat{\underline{F}}_{\alpha_p}$ is preferable only as long as the number of iterations is small. The same remark is true for algorithm 2b where the extra computing time to estimate $\hat{\mathrm{var}}(\hat{\underline{F}}_{\alpha_p})$ cannot offset the smaller accuracy of a constant weighting sequence used in algorithm 2c.

In fig. 2.5 we plot the sampling variance $\mathrm{var}(\hat{F}_\alpha(1))$ of the estimated gradient component $\hat{F}_\alpha(1)$. It is sharply dependent on the optimal value of $\alpha(1)$. This behaviour explains the better performance of algorithm 2b for small p because the optimal weighting to obtain $\hat{\bar{c}}_p$ differs considerably from the fixed weighting sequence (2.68).

Finally, the two-stage estimator of section 2.4.6 is applied to example (2.74). Based on the results of fig. 2.4 we use algorithm 2a where the first $p_1=3$ iterations of stage one with samples of size $N_p = pN$ (where $N = 50$) are used to compute $\hat{\underline{F}}_{\alpha_p}$ with eqn.(2.43). Updating the parameter $\hat{\underline{\alpha}}_p$ is based on eqn.(2.60). Thus, a total

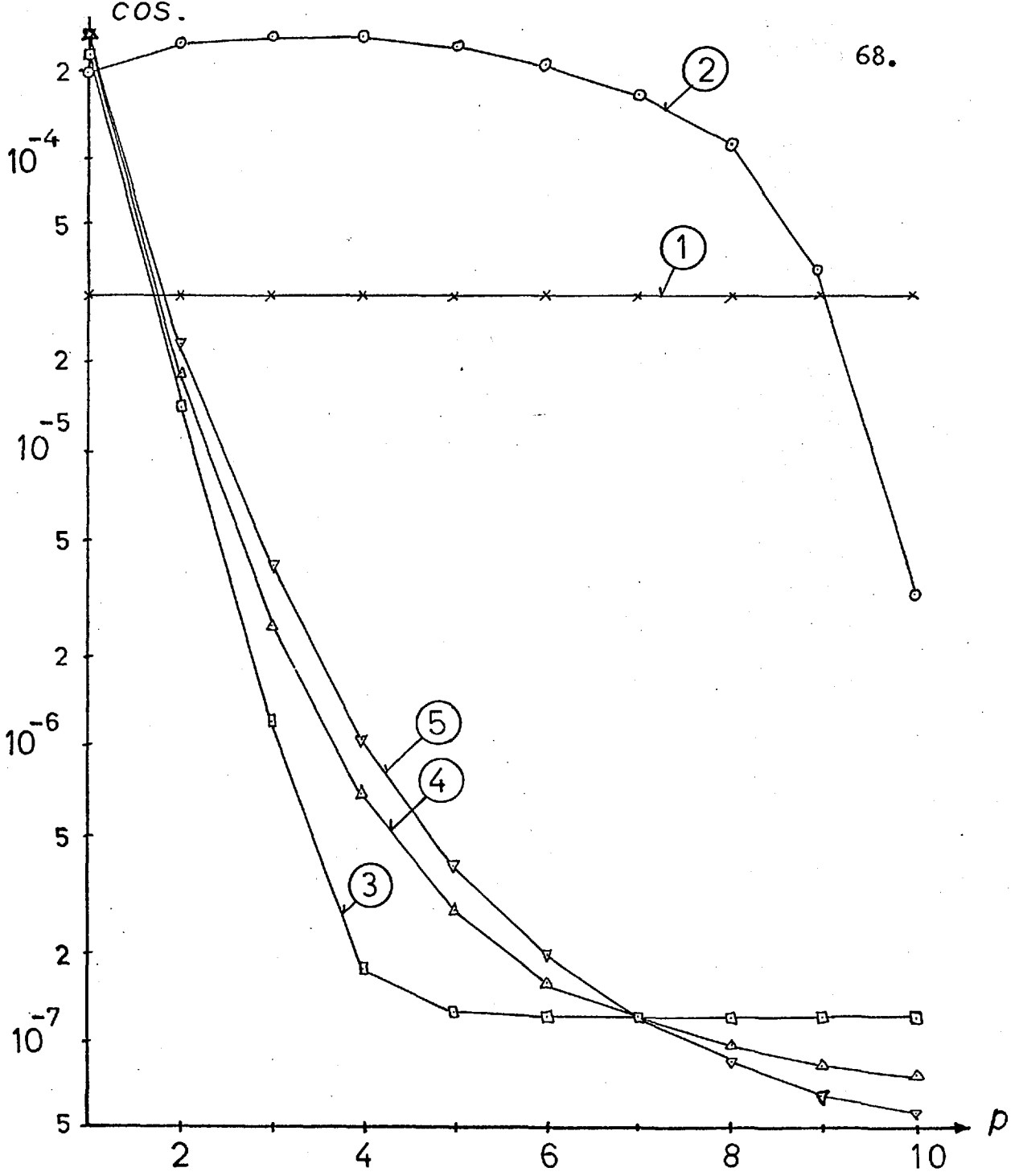

Fig.2.4 Efficiency comparison: 'Cost' ≙ computing time per iteration p times $\widehat{\text{var}}(\hat{\theta}_p)$ vs. iteration number p of the multi-stage Monte Carlo estimation procedure.

1 crude Monte Carlo

2 algorithm 1 — first order method

3 algorithm 2a

4 algorithm 2b — second order method

5 algorithm 2c

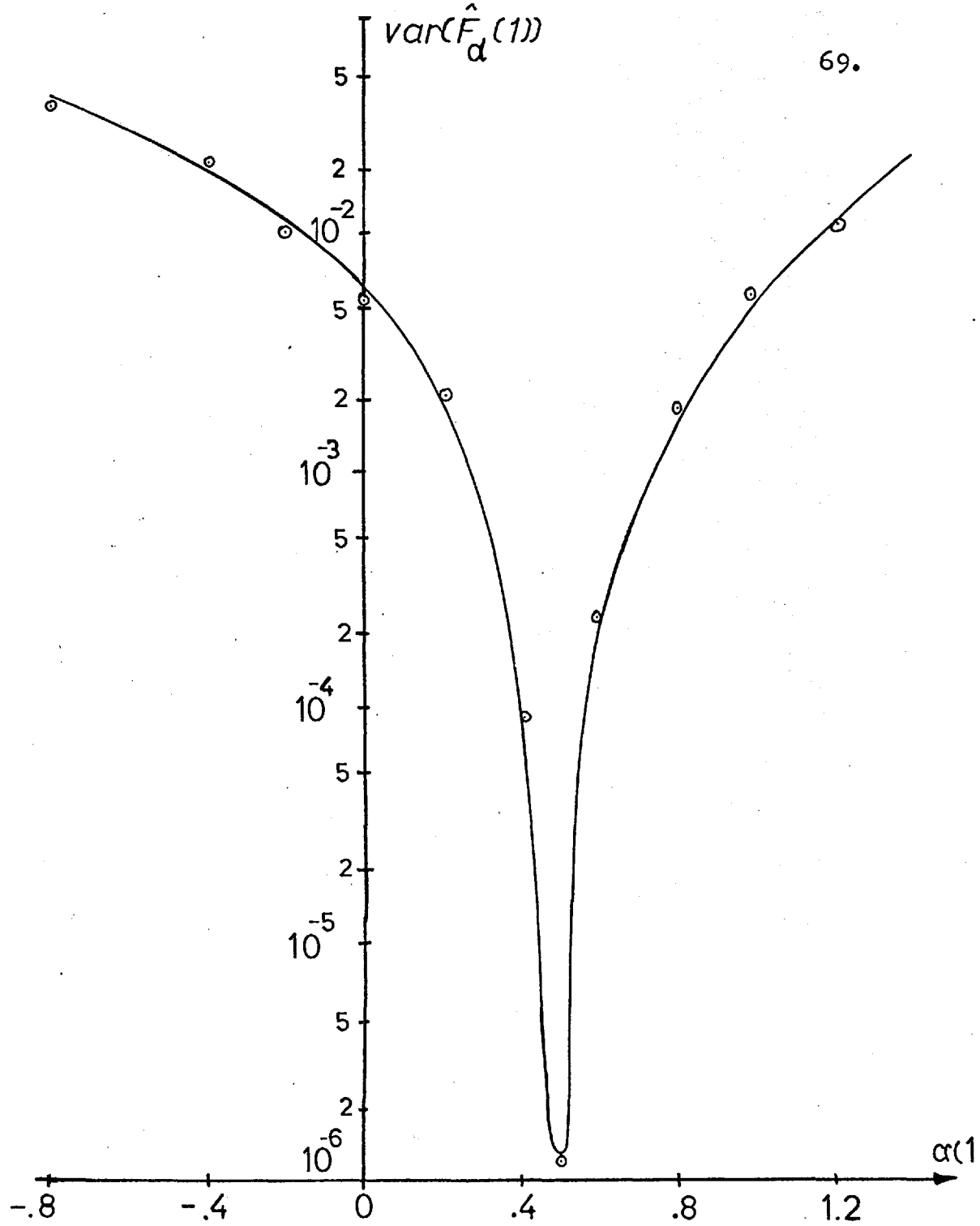

Fig.2.5 Sampling variance of the first component of the estimated gradient vector $\underline{\hat{F}}_\alpha$, $\mathrm{var}(\hat{F}_\alpha(1))$, vs. the first component $\alpha(1)$ of the parameter vector $\underline{\alpha}$. The estimated values ⊙ are ensemble averages over ten Monte Carlo computations of $\hat{\mathrm{var}}(\hat{F}_\alpha(1))$ using samples of size N = 50; $\alpha(2)$ is kept constant and set equal to $\alpha^\circ(2)$.

of $N_t = 300$ samples are used to estimate $\hat{\underline{\alpha}}_{p_1}$ approximating $\underline{\alpha}^o$. In the second stage, a sample of size $N_2 = 500$ is generated to estimate $\hat{\theta}$. Table 2.4 exhibits the results which are again ensemble averages. The parameter $\underline{\alpha}_1$ is initially set equal to one.

Stage	Iterations	$\overline{\hat{\alpha}(1)}$	$\overline{\hat{\alpha}(2)}$	$\overline{\hat{\theta}}$	$\overline{\hat{var}(\hat{\theta})}$
1	1	1.0	1.0		
	2	0.7598	0.1120		
	3	0.4754	0.5119		
2	1	0.4979	0.4909	0.418025	$1.928 \cdot 10^{-8}$

Table 2.4 Two-stage estimator to evaluate integral (2.74)

Comparing these results with the crude Monte Carlo estimator shown in table 2.3 we have to take the different sample sizes into consideration. Thus, the variance reduction η_v, defined by eqn.(2.71), is found to be

$$\eta_v = \frac{500}{800} \cdot \frac{1.601 \cdot 10^{-4}}{1.928 \cdot 10^{-8}} = 5200 . \qquad (2.81)$$

Working with a labour ratio of $\eta_L = 0.5$ the final efficiency gain of the two-stage estimator is found to be $\eta = 2600$.

2.6 Summary and discussion

The multi-stage adaptive control variate method is found to yield a significant improvement compared with the fixed control variate and the crude Monte Carlo method. The additional computational effort required to minimize the cost functional $F(\underline{a})$ in the parameter space is more than compensated by the accuracy gain of the new estimator. The second order method is always to be preferred to the first order method provided the additional condition (C3) does not conflict with (C1). The problem of deciding upon the structure of the control variate function $\phi(x,\underline{a})$ can only be treated for a given function $f(x)$. That is, in order to get a good procedure the specific properties of the problem considered have to be fully exploited. This attitude seems to be necessary in all types of Monte Carlo work and confirms that it is more rewarding than trying to apply some standard techniques.

If the parameter $\underline{a}$ of the control variate function $\phi(x,\underline{a})$ is found by linear regression technique then such a procedure corresponds to a two-stage estimator. However, our solution presented does not include a random matrix to be inverted but a deterministic matrix which can be computed before the sampling experiment starts.

As far as the efficiency gain is concerned the two-stage estimator of section 2.4.6 gives the best result for the example considered. This is largely due to the fact that

(1) the matrix F_{aa} is almost singular and

(2) the problem of combining estimates can be avoided.

Although the derivations in section 2.4 and the example of section

2.5 are based on a scalar case there is no reason why the same method should not be applicable in cases where $\underline{x}$ is an n-dimensional vector. The adaptive control variate method will be taken up in the following chapter when we consider the state variable prediction in systems where all states are accessible. There, the discussion will be extended to the n-dimensional case in the context of the state prediction problem.

CHAPTER THREE

STATE VARIABLE PREDICTION

3.1 Introduction

3.1.1 Stochastic difference equations

A discrete-time stochastic process is represented by the following vector difference equation,

$$\underline{x}_{k+1} = \underline{f}(\underline{x}_k, \underline{w}_k, k) , \tag{3.1}$$

where $\underline{x}_k$ is the n-dimensional state vector, $\underline{w}_k$ is the $p \leq n$ dimensional random disturbance and k denotes the time.

Using the notation

$$\underline{w}^k \triangleq \underline{w}_1, \underline{w}_2 \cdots , \underline{w}_k \tag{3.2}$$

to designate the collection of random vectors $\underline{w}_k$ we make the assumption that $\underline{w}^k$ is a white random sequence for which the conditional P.D.F. satisfies

$$p(\underline{w}_i | \underline{w}_k) = p(\underline{w}_i), \quad \text{for } i > k . \tag{3.3}$$

Moreover, if all $\underline{w}_k$ are jointly normally distributed we assume that $\underline{w}^k$ is a white Gaussian vector sequence. It is completely specified by the mean vector

$$E[\underline{w}_k] = \underline{0} , \quad \text{for all } k \tag{3.4}$$

and the covariance matrix Σ_{w_k}

$$E[\underline{w}_i \underline{w}_k^T] = \Sigma_{w_k} \delta_{ik}. \tag{3.5}$$

A white Gaussian random sequence is a good model for the noise in a discrete-time dynamical system where the noise is due to the superposition of a large number of small, completely random (independent) effects. Finally, it is emphasized that a white noise sequence must not be confused with a white noise process which has a considerably different character.

For the Monte Carlo approach the assumption that $\underline{w}^k$ is a white Gaussian noise sequence is not restrictive. Indeed, in section 2.3.1 we discussed the problem of sampling from an arbitrary P.D.F. and showed that in this respect there is no fundamental difference between the normal and any arbitrary P.D.F. However, due to the computing facilities available it is convenient to draw samples from the normal P.D.F.

Our object in this chapter is to study those stochastic processes which arise as solutions of eqn.(3.1). The evolution in time of such a process is properly described, not by the system state vector $\underline{x}_k$, but by the time varying P.D.F. $p(\underline{x}_k, k)$. To economize on our notation we shall drop the time argument k in all future expressions involving $p(\underline{x}_k)$.

A solution of eqn.(3.1) can be found provided:

(1) the random initial condition is given in terms of a known P.D.F. $p(\underline{x}_1)$ (we assume the system starts at time $k = 1$).

(2) the P.D.F. of the disturbance $\underline{w}_k$, $p(\underline{w}_k)$, is known for all

times k.

All the states $\underline{x}_k$ of system (3.1) are assumed to be accessible. We call the estimation of some parameters of $p(\underline{x}_k)$ the state or trajectory prediction problem according to whether the time k is considered to be fixed or varying.

3.1.2 Markov sequences

In this and the following section we shall discuss some analytic aspects of the prediction problem in order to show the advantages of Monte Carlo methods. A detailed discussion of Markov sequences is given by Papoulis[35] and Doob[44].

Considering eqn.(3.1), it is clear that given a sample value of $\underline{x}_k$, the value taken by $\underline{x}_{k+1}$ depends only on the value taken by $\underline{w}_k$. But $\underline{w}_k$ is completely random, (it comes from a white sequence), and is, by assumption, independent of $\underline{x}_1$ and thus also independent of $\underline{x}_k,\ldots,\underline{x}_2$. It therefore follows that

$$p(\underline{x}_{k+1}|\underline{x}_k,\ldots,\underline{x}_1) = p(\underline{x}_{k+1}|\underline{x}_k), \text{ for any } k. \tag{3.6}$$

That is, the sequence $\underline{x}^k$ generated by eqn.(3.1) is a Markov sequence. The transitional P.D.F. $p(\underline{x}_{k+1}|\underline{x}_k)$ of eqn.(3.6) is an important expression for the prediction problem. The P.D.F. $p(\underline{x}_{k+1}|\underline{x}_k)$ can be expressed in terms of $p_{\underline{w}_k}(\underline{w}_k)$ using the fundamental transformation theorem for P.D.F.[35]. Suppose that eqn.(3.1) can be solved for $\underline{w}_k$. That is, for fixed $\underline{x}_k$ and k, $\underline{f}(\underline{x}_k, \cdot, k)$ has an inverse, say $\underline{f}^*$ such

that

$$\underline{w}_k = \underline{f}^* (\underline{x}_k, \underline{x}_{k+1}, k). \tag{3.7}$$

Then, we have from the theory of derived P.D.F.

$$p(\underline{x}_{k+1} | \underline{x}_k) = p_{\underline{w}_k} (\underline{f}^*(\underline{x}_k, \underline{x}_{k+1}, k)) \left| \frac{\partial \underline{f}^* (\underline{x}_k, \underline{x}_{k+1}, k)}{\partial \underline{x}_{k+1}} \right| . \tag{3.8}$$

For eqn.(3.8) to hold, we must at least have n=p, though this by itself is not sufficient.

Before proceeding with the prediction problem, we will make a small digression and prove the following result:

> If the dimension p < n or if p=n but $\underline{f}$ does not have an inverse $\underline{f}^*$, a suitable partitioning# of $\underline{f}$ (perhaps with a change of variables) such that the dimension of $\underline{f}^{(1)}$ is equal or less than p yields an expression for the transitional P.D.F. $p(\underline{x}_{k+1} | \underline{x}_k)$ which requires the inverse of only a lower dimensional function $\underline{f}^{(1)}$, say $\underline{f}^{*(1)}$.

Proof: We rewrite eqn.(3.1) in the form

$$\begin{aligned} \underline{x}_{k+1}^{(1)} &= \underline{f}^{(1)}(\underline{x}_k^{(1)}, \underline{x}_k^{(2)}, \underline{w}_k, k) \\ \underline{x}_{k+1}^{(2)} &= \underline{f}^{(2)}(\underline{x}_k^{(1)}, \underline{x}_k^{(2)}, k) \end{aligned} \tag{3.9}$$

such that $\underline{x}_k^{(1)}$ and $\underline{w}_k$ are of the same dimension $p' \leq p$ and $\underline{f}^{(1)}(\underline{x}_k^{(1)}, \underline{x}_k^{(2)}, \cdot, k)$ has an inverse, say $\underline{f}^{*(1)}$. Then the transitional P.D.F. $p(\underline{x}_{k+1} | \underline{x}_k)$ takes the form

---

Assuming such a partitioning exists.

$$p(\underline{x}_{k+1}|\underline{x}_k) \triangleq p(\underline{x}^{(1)}_{k+1}, \underline{x}^{(2)}_{k+1} | \underline{x}_k)$$

$$= p(\underline{x}^{(2)}_{k+1} | \underline{x}^{(1)}_{k+1}, \underline{x}_k)\, p(\underline{x}^{(1)}_{k+1}|\underline{x}_k) \qquad (3.10)$$

where the last equality follows from the usual rules of manipulating conditional P.D.F. Using eqn.(3.9) we now have as in eqn.(3.8)

$$p(\underline{x}^{(1)}_{k+1}|\underline{x}_k) = p_{\underline{w}_k}(\underline{f}^{*(1)}(\underline{x}^{(1)}_k, \underline{x}^{(2)}_k, \underline{x}^{(1)}_{k+1}, k)) \left| \frac{\partial \underline{f}^{*(1)}}{\partial \underline{x}^{(1)}_{k+1}} \right| \qquad (3.11)$$

and

$$p(\underline{x}^{(2)}_{k+1}|\underline{x}^{(1)}_{k+1}, \underline{x}_k) = \delta(\underline{x}^{(2)}_{k+1} - \underline{f}^{(2)}(\underline{x}^{(1)}_k, \underline{x}^{(2)}_k, k)) \qquad (3.12)$$

where $\delta(.)$ denotes the dirac delta function, defined by $\int \delta(x)\, dx = 1$ and for $x \neq 0$: $\delta(x) = 0$.

This completes the proof of the above result. We have determined the transitional P.D.F. $p(\underline{x}_{k+1}|\underline{x}_k)$ of the Markov sequence (3.1) for a special case in eqn. (3.8) and under more general conditions in eqns.(3.11) and (3.12).

We now develop a difference equation for the transition P.D.F. $p(\underline{x}_k|\underline{x}_i)$ for $k > i+1$. The marginal P.D.F. $p(\underline{x}_k|\underline{x}_{k-2})$ is obtained in the usual way as

$$p(\underline{x}_k|\underline{x}_{k-2}) = \int p(\underline{x}_k|\underline{x}_{k-1}, \underline{x}_{k-2})\, p(\underline{x}_{k-1}|\underline{x}_{k-2})\, d\underline{x}_{k-1}. \qquad (3.13)$$

With eqn.(3.6) we obtain from eqn.(3.13) the Chapman-Kolmogorov equation as

$$p(\underline{x}_k|\underline{x}_{k-2}) = \int p(\underline{x}_k|\underline{x}_{k-1})\, p(\underline{x}_{k-1}|\underline{x}_{k-2})\, d\underline{x}_{k-1}. \qquad (3.14)$$

We may replace $\underline{x}_{k-2}$ in eqn.(3.14) by $\underline{x}_i$, where $i \leq k-2$ and write

$$p(\underline{x}_k|\underline{x}_i) = \int p(\underline{x}_k|\underline{x}_{k-1})\, p(\underline{x}_{k-1}|\underline{x}_i)\, d\underline{x}_{k-1} \,. \tag{3.15}$$

Via eqn.(3.15), we can determine all other transition P.D.F. of the Markov sequence (3.1).

Finally, it is easy to see that the P.D.F. $p(\underline{x}_k)$, given only $p(\underline{x}_1)$, also satisfies eqn.(3.15); that is,

$$p(\underline{x}_k) = \int p(\underline{x}_k|\underline{x}_{k-1})\; p(\underline{x}_{k-1})\, d\underline{x}_{k-1} \tag{3.16}$$

is valid for k=2,3,... The integrals appearing above are shorthand for the n-fold integral with respect to the vector $\underline{x}_{k-1}$. Equation (3.16) represents a recursive solution of the nonlinear prediction problem and is directly derived from the Chapman-Kolmogorov equation (3.14). In actual practice the evaluation of eqn.(3.16) may involve the following difficulties:

(1) As the discussion of eqn.(3.8) has shown, the evaluation of the conditional P.D.F. $p(\underline{x}_k|\underline{x}_{k-1})$ may cause considerable difficulties. In particular, the computation of the inverse function $\underline{f}^*$ and the Jabobian $|\partial \underline{f}^*/\partial \underline{x}_{k+1}|$ are serious obstacles which can prevent a practical solution.

(2) Even if the integrand in eqn.(3.16) can be found the evaluation of the n-dimensional integral by means of analytic methods may be prohibitive.

As already mentioned in chapter two, one of the main advantages of Monte Carlo techniques is their suitability to evaluate high

dimensional integrals. Before such techniques can be applied the recursion (3.16) has to be replaced by an equivalent expression which involves only prior information about the system. Indeed, due to the Markov property, eqn.(3.16) can be rewritten as

$$p(\underline{x}_k) = \int p(\underline{x}_k \mid \underline{x}_{k-1})\, p(\underline{x}_1, \ldots, \underline{x}_{k-1})\, d\underline{x}_1 \ldots d\underline{x}_{k-1} \qquad (3.17)$$

where the integral to be evaluated is $n(k-1)$-fold.

Although eqn.(3.17) is amenable to a Monte Carlo solution this problem formulation based on the Chapman-Kolmogorov equation (3.14) is not efficient because sampling techniques are not suited to evaluating the entire P.D.F. However, they are very useful to estimate some parameters of $p(\underline{x}_k)$. In section 3.2 the prediction problem will be re-examined and starting from first principles a solution is sought which

(1) is based on the specific properties of a Monte Carlo computation;

(2) avoids the difficulties encountered when the solution is derived via the Chapman-Kolmogorov equation.

3.1.3 Linear and nonlinear predictors

It is well known that for a linear system of the form

$$\underline{x}_{k+1} = A_k \underline{x}_k + \underline{w}_k \qquad (3.18)$$

with a Gaussian P.D.F. $p(\underline{x}_1)$ specifying the initial conditions and a white Gaussian sequence $\underline{w}^k$, the P.D.F. $p(\underline{x}_k)$ of the state $\underline{x}_k$ is

Gaussian at any time k. Indeed, the P.D.F.

$$p(\underline{x}_k) = n(\underline{x}_k; \underline{\mu}_k, \Sigma_k) \tag{3.19}$$

has the mean $\underline{\mu}_k$ and variance Σ_k obtainable from eqn.(3.18) as

$$\begin{aligned} \underline{\mu}_{k+1} &= A_k \underline{\mu}_k \\ \Sigma_{k+1} &= A_k \Sigma_k A_k^T + \Sigma_{w_k} . \end{aligned} \tag{3.20}$$

These recursive equations start at time k=1 with

$$\underline{\mu}_1 = \underline{m}_x \quad \text{and} \quad \Sigma_1 = \Sigma_x , \tag{3.21}$$

where $\underline{m}_x$ and Σ_x denote mean and covariance matrix of $p(\underline{x}_1)$ respectively. Although eqn.(3.20) can be computed directly from eqn.(3.18) it also follows from eqn.(3.16).

To emphasize the difficulties arising from nonlinear transformations $\underline{f}$ let us now consider a nonlinear difference equation of the following form

$$\underline{x}_{k+1} = \underline{f}(\underline{x}_k, k) + \underline{h}(\underline{x}_k, k)\, \underline{w}_k \tag{3.22}$$

with $\underline{h}(\underline{x}_k, k) \neq 0$ and $\underline{w}^k$ a white random vector sequence. Based on the results of section 3.1.2, the transition P.D.F. is found by means of eqn.(3.7). If $p(\underline{w}_k)$ is normal with zero mean and variance Σ_{w_k} we find the mean $\underline{\mu}_k$ and the covariance Σ_k of $p(\underline{x}_k)$ as

$$\begin{aligned} \underline{\mu}_{k+1} &= E\,[\underline{f}(\underline{x}_k, k)] \\ \Sigma_{k+1} &= E\,[\underline{f}(\underline{x}_k, k)\, \underline{f}^T(\underline{x}_k, k)] - E[\underline{f}(\underline{x}_k, k)]\, E^T[\underline{f}(\underline{x}_k, k)] \\ &\quad + E\,[\underline{h}(\underline{x}_k, k)\, \underline{w}_k\, \underline{w}_k^T\, \underline{h}^T(\underline{x}_k, k)] . \end{aligned} \tag{3.23}$$

Now, eqn.(3.23) is a nonlinear predictor which reduces to eqn.(3.20) if $\underline{f}$ is linear and $\underline{h}$ a constant. We notice that eqn.(3.23) is not a difference equation because the knowledge of $p(\underline{x}_k)$ is required for its evaluation. Approximate equations for $\underline{\mu}_k$ and Σ_k may be obtained if we make the assumptions that $p(\underline{x}_k)$ is symmetric and close to the mean; that is, if $\underline{\mu}_k$ and Σ_k describe $p(\underline{x}_k)$ with sufficient accuracy we might expand the functions appearing on the R.H.S. of eqn.(3.23) in Taylor series about $\underline{\mu}_k$. The disadvantage of such an approximation procedure is the absence of an error analysis.

In conclusion, the discussion in this section 3.1 has shown that the success of the analytic approach to linear systems with additive white Gaussian noise is due entirely to the fact that the mean and covariance matrix of eqn.(3.20) describe the P.D.F. $p(\underline{x}_k)$ completely. On the other hand, in the nonlinear case, eqn.(3.23) indicates that its exact solution leads to a set of infinitely many equations. We do not attempt to solve the Chapman-Kolmogorov equation (3.14) directly. In the remainder of this chapter we rather concentrate on the design of Monte Carlo procedures to obtain statistical solutions for predicting the states based on the nonlinear system itself where error bounds are directly obtained as part of the solution.

3.2 A crude Monte Carlo predictor

3.2.1 The method of moments

Let us consider again the stochastic difference eqn.(3.1) with an arbitrary but known P.D.F. $p(\underline{x}_1)$ and a white random sequence $\underline{w}^k$ whose P.D.F. $p(\underline{w}_k)$ is also known for all k. It is desired to find the P.D.F. $p(\underline{x}_k)$ over a time sequence or at some fixed time k. Since Monte Carlo methods are not suited to determining entire functions, we are interested in estimating some parameters describing $p(\underline{x}_k)$ rather than in evaluating $p(\underline{x}_k)$ itself. We have to resort to the method of moments, see Deutsch[45], because the nonlinearity of the relations between the observables and the system parameters are in general such that most methods of estimation become too complicated to manipulate. The r:th order moment of the system (3.1) is defined by

$$m_k^{(c_1, c_2, \ldots, c_n)} \triangleq E[G(\underline{x}_k)] = \int G(\underline{x}_k)\, p(\underline{x}_k)\, d\underline{x}_k, \qquad (3.24)$$

where

$$G(\underline{x}_k) \triangleq x_k^{c_1}(1)\, x_k^{c_2}(2) \ldots x_k^{c_n}(n). \qquad (3.25)$$

$x_k(i)$ denotes the i:th component of the state vector $\underline{x}_k$ and the exponents c_i in eqn.(3.25) satisfy the condition

$$\sum_{i}^{n} c_i = r \quad , \quad r = 1,2,\ldots \qquad (3.26)$$

In particular, the first order moments, where $c_i = 1$, define the mean

vector $\underline{\mu}_k$ by

$$\underline{\mu}_k = E[\underline{x}_k] \triangleq \int \underline{x}_k \, p(\underline{x}_k) \, d\underline{x}_k \quad . \tag{3.27}$$

The matrix M_k of second order moments, where $c_i = 2$ or $c_i = c_j = 1$, is given by

$$M_k = E[\underline{x}_k \, \underline{x}_k^T] \triangleq \int \underline{x}_k \, \underline{x}_k^T \, p(\underline{x}_k) \, d\underline{x}_k . \tag{3.28}$$

Finally, the covariance matrix Σ_k is given by the second order central moments and defined by

$$\Sigma_k = E[(\underline{x}_k - \underline{\mu}_k)(\underline{x}_k - \underline{\mu}_k)^T] . \tag{3.29}$$

While a particular moment or a few of the moments give some information about the P.D.F., only the infinite set of all moments (3.24) ... (3.26) will determine the P.D.F. exactly. Yet, in practical applications only the first two moments are of great importance because it often is unnecessary to know the actual shape of the P.D.F. On the other hand, it is usually necessary to know at least the location of the P.D.F. and to have some estimate of its dispersion. In the following section 3.2.2 it will be shown how these characteristics can be estimated by examining a random sample drawn from a set of values known to have the P.D.F. in question.

3.2.2 The sample moments

The Monte Carlo evaluation of the integrals defined in the previous section yields the sample moments. By the strong law of large numbers,

(see section 2.2.2), these sample moments converge with probability one to the true moments if the moments are finite and if the sample size $N \to \infty$. The sample mean $\hat{\underline{\mu}}_k$ is computed by

$$\hat{\underline{\mu}}_k = N^{-1} \sum_j^N (\underline{x}_k)_j \quad . \tag{3.30}$$

Similarly the matrix $\hat{M}_k$ of the sample second order moments is given by

$$\hat{M}_k = N^{-1} \sum_j^N (\underline{x}_k)_j \, (\underline{x}_k)_j^T \quad . \tag{3.31}$$

The sample covariance matrix $\hat{\Sigma}_k$ is obtained as

$$\hat{\Sigma}_k = N^{-1} \sum_j^N [(\underline{x}_k)_j - \hat{\underline{\mu}}_k][(\underline{x}_k)_j - \hat{\underline{\mu}}_k]^T \quad . \tag{3.32}$$

Except in eqn.(3.30) the sample moments will be biased estimates of the population moments. This bias is negligible, however, for large N.

The quantity $(\underline{x}_k)_j$ appearing in the above expressions denotes the random vector selected by the j:th drawing from the P.D.F. $p(\underline{x}_k)$. It is generated by a direct simulation of eqn.(3.1) and is a function of k random vectors; $(\underline{x}_1)_j$ is drawn from the P.D.F. $p(\underline{x}_1)$ and $(\underline{w}^{k-1})_j$ is drawn from $p(\underline{w}_k)$. The N-fold repetition of the direct simulation of eqn.(3.1) yields the random sample $\{\underline{x}_k\}_j$ required to evaluate eqns.(3.30)...(3.32). This procedure is a typical illustration of remark (5) in section 2.2.3. Although it is not possible to give an analytic solution for $p(\underline{x}_k)$, direct simulation provides a powerful method for sampling from $p(\underline{x}_k)$.

From the above discussion it is obvious that, unlike theoretical

statistical moments which are computed from given P.D.F. and are therefore known functions of the distribution parameters, sample moments will themselves be random variables and are associated with P.D.F. This point has already been emphasized in section 2.2.3. In our application of statistical methods we are interested in relatively simple statistics such as averages, sums of squares, ratios and covariances. Due to the central limit theorem we are able to give at least approximate confidence intervals for the sample moments.

First, let us consider the moments of the sampling P.D.F. of $\hat{\underline{\mu}}_k$ introduced in eqn.(3.30). The following useful theorem is proven by Wilks[34]:

Let $\{\underline{x}_k\}_j$ be a random sample of size N drawn from the population whose P.D.F. is $p(\underline{x}_k)$. Denoting the mean of the population by $\underline{\mu}_k$, then $E[\hat{\underline{\mu}}_k] = \underline{\mu}_k$. We also assume that $p(\underline{x}_k)$ has a finite covariance matrix Σ_k. The sampling variance of $\hat{\underline{\mu}}_k$, denoted as $\text{var}(\hat{\underline{\mu}}_k)$, is defined by

$$\text{var}(\hat{\underline{\mu}}_k) \triangleq E[(\hat{\underline{\mu}}_k - E[\hat{\underline{\mu}}_k])(\hat{\underline{\mu}}_k - E[\hat{\underline{\mu}}_k])^T] \tag{3.33}$$

where

$$\hat{\underline{\mu}}_k - E[\hat{\underline{\mu}}_k] = N^{-1} \sum_{j}^{N} [(\underline{x}_k)_j - \underline{\mu}_k] . \tag{3.34}$$

Hence, using eqn.(3.29), $\text{var}(\hat{\underline{\mu}}_k)$ is found to be

$$\text{var}(\hat{\underline{\mu}}_k) = N^{-1}\Sigma_k = N^{-1} \text{var}(\underline{x}_k). \tag{3.35}$$

Thus, the covariance matrix of the sample mean is equal to the population covariance matrix Σ_k divided by the sample size N. Since Σ_k

is unknown it has to be replaced by $\hat{\Sigma}_k$ of eqn.(3.32); that is, $\text{var}(\hat{\underline{\mu}}_k)$ has to be replaced by its estimate $\hat{\text{var}}(\hat{\underline{\mu}}_k)$.

While it is relatively easy to derive the mean and variance of the sampling distribution of $\hat{\underline{\mu}}_k$, the complexity of the multivariate statistical analysis for the second order sample moment is much greater. Anderson[46] gives a full account of the sampling distribution of the sample covariance matrix $\hat{\Sigma}_k$. As a simplification we shall use eqn.(2.10) to estimate the mean and sampling variance of the second order central moments where the scalar function f(x) is to be replaced by $(\underline{x}_k - \underline{\mu}_k)(\underline{x}_k - \underline{\mu}_k)^T$. This procedure yields exact values provided the random sample is drawn from a population with a normal P.D.F. $p(\underline{x}_k)$.

The exact derivation for mean and variance of the second order central moment is given by eqn.(2.11) for the scalar case. In a similar manner, means and variances of higher order sample moments could be carried out. But even in the scalar case the tediousness of the process increases rapidly. In agreement with our previous remarks in section 3.2.1 we do not pursue the topic of higher order moments. For further details concerning this aspect of sampling theory we refer to Kendall and Stuart[59].

Equations (3.30)...(3.32) define crude Monte Carlo predictors. The only way of improving the accuracy of these estimators is to increase the sample size N. As can be immediately seen in the case of the sample mean $\hat{\underline{\mu}}_k$ this is not a very rewarding procedure because the sampling error, defined as the square root of the sampling variance

eqn.(3.35), is only proportional to $N^{-1/2}$.

The significance of variance reduction techniques has already been shown in chapter two. In the present context we found that the efficiency of Monte Carlo predictors can be considerably improved if we use correlated sampling methods:

(1) Negative correlation between the samples leads to the antithetic variate method, as described in section 3.3.

(2) Positive correlation between the samples leads to the control variate method; see section 3.4.

3.3 The antithetic variate method

The concept and name of the antithetic variate method was introduced by Hammersley and Morton[47] for the evaluation of scalar and multidimensional integrals using samples drawn from the uniform P.D.F. The general theoretical structure of the antithetic variate method in connection with the uniform P.D.F. is contained in a paper by Hammersley and Mauldon[48]. Mayne[29] applied the concept of antithetic variates to normal P.D.F. for a gradient technique to find the optimal control for a stochastic nonlinear system.

Suppose that $\hat{\theta}^+$ is a Monte Carlo estimate of an unknown parameter θ. The basic idea of the antithetic variate method is to seek a statistic $\hat{\theta}^-$ which

(1) has a strong negative correlation with $\hat{\theta}^+$, and

(2) is an unbiased estimate of θ; i.e. $E[\hat{\theta}^-] = \theta$.

If both these conditions are satisfied then $\hat{\theta}^+$ and $\hat{\theta}^-$ will mutually compensate each other's variations. After sampling $\hat{\theta}^+$ and $\hat{\theta}^-$ simultaneously we employ

$$\hat{\theta} = \frac{1}{2} (\hat{\theta}^+ + \hat{\theta}^-) \tag{3.36}$$

as the estimator of θ.

We shall elaborate on this concept in section 3.3.1, where a nonlinear predictor for the mean $\underline{\mu}_k$ is presented. The discussion in section 3.3.2 of the special case of a linear Gaussian system leads to an interesting result. Finally, the topic of a generalized antithetic variate method is discussed in section 3.3.3.

3.3.1 A nonlinear predictor using 2N antithetic variates

Let us consider a nonlinear system of the form of eqn.(3.1). In addition to the assumption of independence of $\underline{x}_1$ and $\underline{w}_k$, the corresponding P.D.F. must be symmetric and unimodal. In order to carry the development of nonlinear predictors beyond general statements the special form of normal multivariate P.D.F. will be assumed for $p(\underline{x}_1)$ and $p(\underline{w}_k)$; that is

$$p(\underline{x}_1) = n(\underline{x}_1; \underline{m}_x, \Sigma_x) \tag{3.37}$$

$$p(\underline{w}_k) = n(\underline{w}_k; \underline{0}, \Sigma_{w_k}) . \tag{3.38}$$

Let us denote by $\{\underline{x}_k\}_j^+$ the original sample as it has been used in connection with eqn.(3.30). In addition to this sample we generate

a negatively correlated sample $\{\underline{x}_k\}_j^-$ as N solutions of eqn.(3.1) using the antithetic initial condition variates

$$((\underline{x}_1)_j^- - \underline{m}_x) = -((\underline{x}_1)_j^+ - \underline{m}_x) \tag{3.39}$$

and the antithetic noise sequences

$$(\underline{w}_k)_j^- = -(\underline{w}_k)_j^+ , \quad k = 1,2,\ldots, \tag{3.40}$$

for $j = 1,2,\ldots,N$.

Because of eqn.(3.37)...(3.40) we have $p(\underline{x}_1^+) = p(\underline{x}_1^-)$ and $p(\underline{w}_k^+) = p(\underline{w}_k^-)$. These two conditions are essential for the antithetic variate method since they ensure that the new estimator for the mean $\underline{\mu}_k$, based on eqn.(3.36) and given by

$$\hat{\underline{\mu}}_k = \frac{1}{2N} \sum_j^N [(\underline{x}_k)_j^+ + (\underline{x}_k)_j^-] = \frac{1}{2} [\hat{\underline{\mu}}_k^+ + \hat{\underline{\mu}}_k^-], \tag{3.41}$$

where

$$\hat{\underline{\mu}}_k^{\pm} \triangleq N^{-1} \sum_j^N (\underline{x}_k)_j^{\pm} , \tag{3.42}$$

is unbiased. As $\hat{\underline{\mu}}_k$ is the sum of two correlated estimates its sampling variance, $\mathrm{var}(\hat{\underline{\mu}}_k)$, is given by

$$\mathrm{var}(\hat{\underline{\mu}}_k) = \frac{1}{4} \mathrm{var}(\hat{\underline{\mu}}_k^+) + \frac{1}{4} \mathrm{var}(\hat{\underline{\mu}}_k^-) + \frac{1}{2} \mathrm{cov}(\hat{\underline{\mu}}_k^+, \hat{\underline{\mu}}_k^-). \tag{3.43}$$

It can be made smaller than $\mathrm{var}(\hat{\underline{\mu}}_k^+)$, the sampling covariance matrix (3.35) of the crude Monte Carlo estimator, provided

$$\mathrm{cov}(\hat{\underline{\mu}}_k^+, \hat{\underline{\mu}}_k^-) < 0. \tag{3.44}$$

In actual practice, none of the terms of eqn.(3.43) is known and all therefore have to be replaced by their estimates.

To illustrate the antithetic variate method let us consider the simplified problem of estimating the value $\hat{\theta}$ of the scalar integral

$$\theta = \int f(x)\, p(x)\, dx \tag{3.45}$$

as

$$\hat{\theta} = N^{-1} \sum_{j}^{N} f(x_j). \tag{3.46}$$

The random sample $\{x\}_j$ is drawn from $p(x)$, e.g. a scalar normal P.D.F. In fig. 3.1 we plot a few variates $f(x_j)$ as generated by a crude Monte Carlo procedure and the resulting sampling P.D.F. of $\hat{\theta}$.

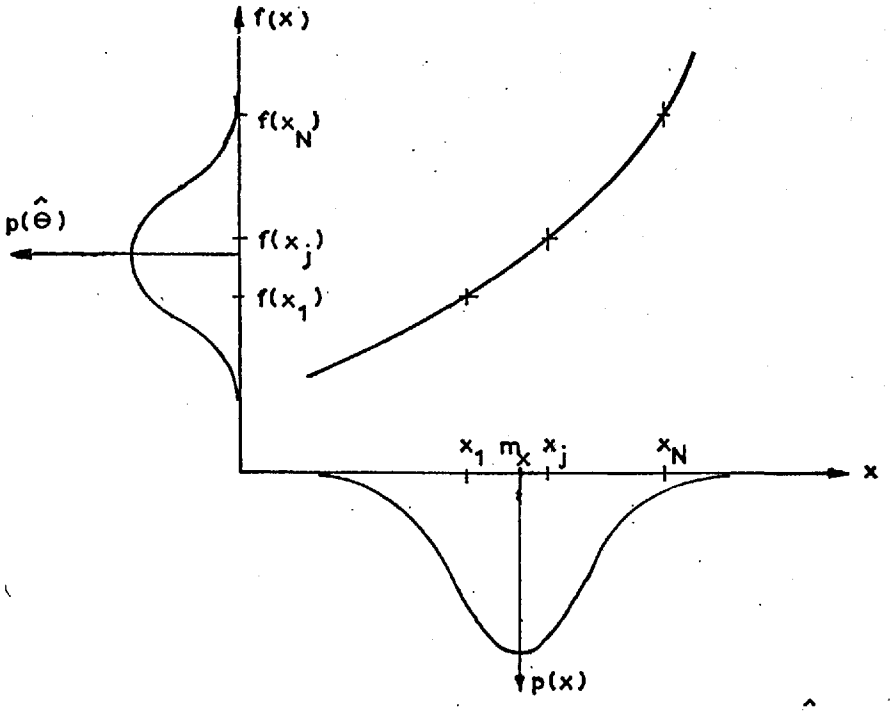

Fig. 3.1 Sampling distribution of the crude estimate $\hat{\theta}$ based on eqn. (3.46).

In fig. 3.2 we plot the sampling P.D.F. of $\hat{\theta} = N^{-1} \sum_{j}^{N} \xi_j$ using the antithetic variates ξ_j, defined by

$$\xi_j = \frac{1}{2} \left[f(x_j^+) + f(x_j^-) \right], \qquad j = 1,2,\ldots N. \tag{3.47}$$

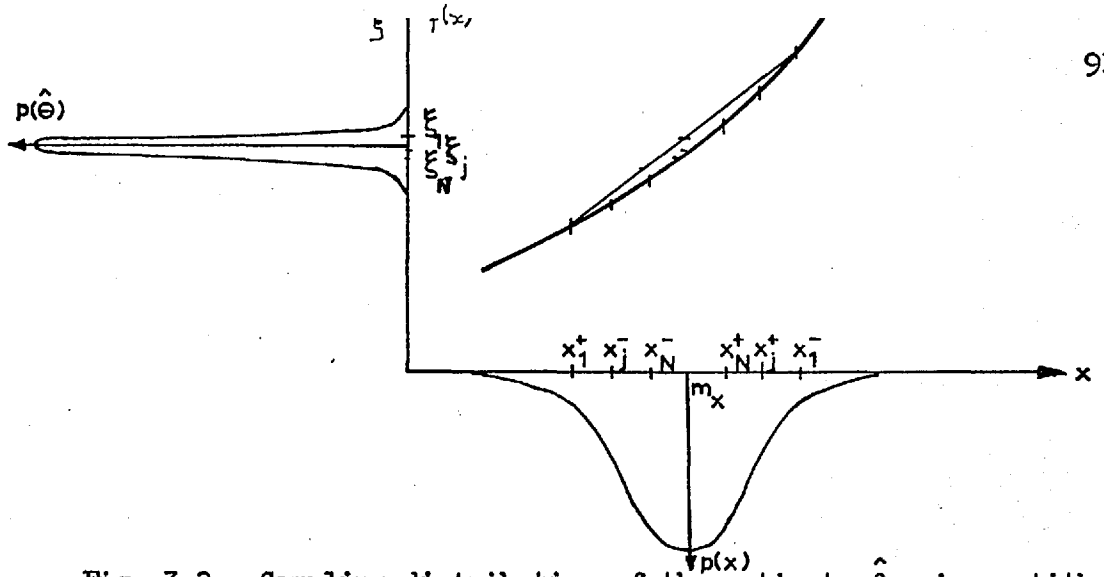

Fig. 3.2 Sampling distribution of the estimate $\hat{\theta}$ using antithetic variates ξ_j as defined by eqn.(3.47).

The dispersion of the sampling P.D.F. $p(\hat{\theta})$ of fig. 3.2 is much smaller than that of fig. 3.1 because the antithetic variates defined by eqn. (3.47), mutually compensate each others fluctuations around the mean m_x of $p(x)$.

3.3.2 A linear system with additive Gaussian noise

A useful property of a Monte Carlo method is to have zero sampling variance in those cases where analytic methods yield accurate results. Such a situation arises for linear systems of the form of eqn.(3.18). We shall prove the following result:

> The antithetic variate predictor of eqn.(3.41) yields an estimate $\hat{\mu}_k$ of the mean μ_k with zero sampling variance when applied to a linear Gaussian system (3.18).

For the proof of this result let us consider one antithetic pair

$(\underline{x}_1, \underline{w}^{k-1})_j^{\pm}$. Applied to eqn.(3.18) we find

$$(\underline{x}_k)_j^{\pm} = \prod_{i=1}^{k-1} A_i (\underline{x}_1)_j^{\pm} + \sum_{i}^{k-1} \prod_{r=i+1}^{k-1} A_r (\underline{w}_i)_j^{\pm} \quad . \tag{3.48}$$

Using eqn.(3.39) and (3.40) to substitute for $(\underline{x}_1)_j^-$ and $(\underline{w}^{k-1})_j^-$ in terms of $(\underline{x}_1)_j^+$ and $(\underline{w}^{k-1})_j^+$ yields for eqn.(3.41)

$$\underline{\mu}_k = \prod_{i=1}^{k-1} A_i \; \underline{m}_x . \tag{3.49}$$

This is an exact result and the sampling error, independent of the sample size N, is always equal to zero.

Of course, the importance of this result lies not in the fact that it provides a solution for the linear Gaussian system. Indeed, an analytic recursive solution for the mean $\underline{\mu}_k$ has been directly derived in eqn.(3.20). But due to our result it is plausible to expect a large variance reduction compared with the crude Monte Carlo predictor for a nearly linear system.

It follows from the proof of our result that the antithetic variate method applied to a linear Gaussian system does not give a zero sampling variance for the estimates of the second order moments and hence for the estimates of the covariance matrix Σ_k .

3.3.3 The antithetic variate method using several estimators

The symmetry between the two estimators $\hat{\underline{\mu}}_k^+$ and $\hat{\underline{\mu}}_k^-$ in eqn.(3.42) used to compute $\hat{\underline{\mu}}_k$ with eqn.(3.41), suggests the extension to use 2p estimators $\hat{\underline{\mu}}_{k,1}^{\pm}$, ..., $\hat{\underline{\mu}}_{k,p}^{\pm}$ and to form a set of p estimators $\hat{\underline{\mu}}_{k,p}$

such that

$$E[\hat{\underline{\mu}}_{k,p}] = F_p \underline{\mu}_k \quad \text{for } p=1,2,.. \tag{3.50}$$

where F_p is a suitable matrix with known elements, see eqn.(3.59).

This problem was originally posed by Hammersley and Morton[47] for the evaluation of integrals by drawing random samples from the uniform P.D.F. It is the object of this section to extend their solution to the case of a normal, n-variate P.D.F.

The symmetry and unimodality conditions are sufficient to ensure the existence of a pair of antithetic variates, arbitrarily denoted by $\underline{\xi}_j^{+-}$, belonging to $p(\underline{\xi})$, such that

$$p(\underline{\xi}^+) = p(\underline{\xi}^-) \ . \tag{3.51}$$

Hence, for $p(\xi)$ being a univariate normal P.D.F. there exists only one antithetic variate ξ_j^- for a given ξ_j^+ which satisfies eqn.(3.51).

In order to simplify our explanations we shall start with a bivariate normal P.D.F. $p(\underline{w}_k)$ for the plant noise $\underline{w}_k$ and show how negatively correlated samples can be found. From there it will be possible to derive the results pertinent to the n-variate P.D.F. As pointed out by Tukey[49] the final result is obtained by linear regression.

Let us now start with the discussion of antithetic variates in connection with the bivariate normal P.D.F. $p(\underline{w}_k)$. In fig. 3.3 we plot an elliptic contour of $p(\underline{w}_k)$ = const.

Let us assume the j:th realization of the sampling procedure

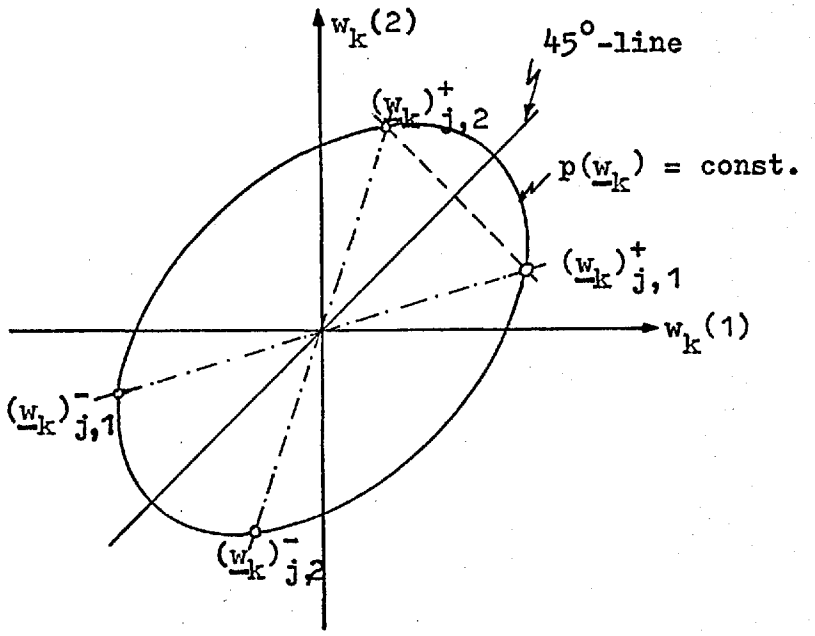

Fig. 3.3 Generation of antithetic variates for a given $(\underline{w}_k)^+_{j,1}$.

from $p(\underline{w}_k)$ yields $(\underline{w}_k)^+_{j,1}$. By means of eqn.(3.40) we determine $(\underline{w}_k)^-_{j,1}$. These two variates are symmetric to each other w.r.t. the origin. Symmetric w.r.t. the main axis of the ellipse in fig. 3.3 we find $(\underline{w}_k)^+_{j,2}$. Under the assumption that the elements along the main diagonal of the covariance matrix $\Sigma_{\underline{w}_k}$ are equal the main axis of the ellipse, given by $p(\underline{w}_k)$ = constant, coincides with the 45°-line. Thus, the components of $(\underline{w}_k)^+_{j,2}$ are obtainable from $(\underline{w}_k)^+_{j,1}$ by permuting the components of $(\underline{w}_k)^+_{j,1}$ as follows

$$\begin{aligned} (w_k(1))^+_{j,2} &= (w_k(2))^+_{j,1} \\ (w_k(2))^+_{j,2} &= (w_k(1))^+_{j,1} . \end{aligned} \tag{3.52}$$

As usual the arguments in brackets denote the coordinates of $\underline{w}_k$. Finally, $(\underline{w}_k)^-_{j,2}$ is found via eqn.(3.40) from $(\underline{w}_k)^+_{j,2}$. Denoting the correlation between $w_k(1)$ and $w_k(2)$ as ρ_{12} we find the following

correlation matrices for the above defined variates:

$$\mathrm{cor}\left((\underline{w}_k)_r^+, (\underline{w}_k)_r^-\right) = \begin{bmatrix} -1 & -\rho_{12} \\ -\rho_{12} & -1 \end{bmatrix} \quad \text{for } r=1,2 \qquad (3.53)$$

and

$$\mathrm{cor}\left((\underline{w}_k)_1^+, (\underline{w}_k)_2^-\right) = \begin{bmatrix} -\rho_{12} & -1 \\ -1 & -\rho_{12} \end{bmatrix} . \qquad (3.54)$$

The comparison between eqns.(3.53) and (3.54) shows the strict functional interdependence between the various $(\underline{w}_k)_{j,p}^{\pm}$, p=1,2. The derivation of additional variates with eqn.(3.52) effectively corresponds to increasing the random sample. It is therefore plausible to expect that the generalized antithetic variate method will lead to at least the same precision in the final estimate as if the sample size of the antithetic variate method, introduced in section 3.3.1 were increased to the equivalent size by conventional random number generation.

We make the conjecture that the negative correlation between the variates leads to a sampling variance reduction. We have not succeeded in proving or disproving this conjecture because in the prediction problem we have to deal with the variates $(\underline{x}_k)_{j,p}^{\pm}$ which are arbitrary nonlinear functions of $(\underline{x}_1)_{j,p}^{\pm}$ and $(\underline{w}^{k-1})_{j,p}^{\pm}$. However, we have already shown in section 3.3.2 that in the linear case the dependence between the variates $(\underline{x}_1)_j^{\pm}$ and $(\underline{w}^{k-1})_j^{\pm}$ leads to a zero sampling variance predictor. This result still holds if we use the

generalized random sample $\{\underline{x}_1, \underline{w}^{k-1}\}^{\pm}_{j,p}$, p=1,2 because two antithetic variates at a time compensate each others random error exactly.

We now proceed with the extension to the n-variate P.D.F. Retaining the above mentioned assumption for Σ_{w_k} we find n! variates by permuting the components of the original $(\underline{w}_k)^{+}_{j,1}$. Each of these n! variates has an antithetic variate obtainable from eqn.(3.40). With a similar procedure based on permuting the components of the random vector $(\underline{x}_1)^{+}_{j,1}$ which is obtained by sampling from $p(\underline{x}_1)$ it is possible to determine n! variates, and n! additional variates are subsequently obtained from eqn.(3.39).

The 2n! random samples $\{\underline{x}_k\}^{\pm}_{j,p}$, p=1,2,...n!, each of size N, are obtained by solving eqn.(3.1) 2n!N times. They are then used to form the following set of n! estimates $\hat{\underline{\mu}}_{k,p}$, (p=1,2,...n!), of $\underline{\mu}_k$:

$$\hat{\underline{\mu}}_{k,p} = (2N)^{-1} \sum_{j}^{N} \left[(\underline{x}_k)^{+}_{j,p} + (\underline{x}_k)^{-}_{j,p} \right]$$

$$= N^{-1} \sum_{j}^{N} (\underline{\xi}_k)_{j,p} \text{ , for p =1,2,...,n! ,} \qquad (3.55)$$

where we define the variate $(\underline{\xi}_k)_{j,p}$ by

$$(\underline{\xi}_k)_{j,p} \triangleq \tfrac{1}{2} \left[(\underline{x}_k)^{+}_{j,p} + (\underline{x}_k)^{-}_{j,p} \right] . \qquad (3.56)$$

We now define the (n·n!)-dimensional vector $\hat{\underline{\mu}}_k$ by

$$\hat{\underline{\mu}}^{T}_k \triangleq \left[\hat{\underline{\mu}}^{T}_{k,1} , \ldots , \hat{\underline{\mu}}^{T}_{k,n!} \right] \qquad (3.57)$$

whose elements are given by eqn.(3.55). Because all n! estimators

(3.55) are unbiased we have

$$E[\hat{\underline{\mu}}_k] = F\,\underline{\mu}_k \tag{3.58}$$

where F is an (nn!,n)-matrix of the following structure

$$F = \begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_p \\ \vdots \\ F_{n!} \end{bmatrix} . \tag{3.59}$$

Each matrix F_p, p=1,2,...,n!, is an (n,n)-dimensional identity matrix.

Finally, the minimum variance unbiased linear estimate $\hat{\hat{\underline{\mu}}}_k$ of $\underline{\mu}_k$ is found from eqn.(2.45) to be given by

$$\hat{\hat{\underline{\mu}}}_k = [F^T V_k^{-1} F]^{-1} F^T V_k^{-1} \hat{\underline{\mu}}_k . \tag{3.60}$$

Since the sampling covariance matrix $V_k = \mathrm{var}(\hat{\underline{\mu}}_k)$ is unknown it has to be replaced by its estimate $\hat{V}_k$. The matrix V_k can be partitioned into (n,n)-dimensional submatrices denoted by $V_{k,pq}$(p,q=1,2,...,n!). The estimate $\hat{V}_{k,pq}$ is computed by

$$\hat{V}_{k,pq} = N^{-2} \sum_{j}^{N} [(\underline{\xi}_k)_{j,p} - \hat{\mu}_{k,p}][(\underline{\xi}_k)_{j,q} - \hat{\mu}_{k,q}]^T \tag{3.61}$$

where $(\underline{\xi}_k)_{j,p}$ is given by eqn.(3.56) and $\hat{\mu}_{k,p}$ is given by eqn.(3.55). The substitution of $\hat{V}_k$ for V_k introduces little bias if any and, provided $\hat{V}_k$ is a good approximation to V_k, eqn.(3.60) yields an almost minimum variance estimator.

The sampling covariance matrix of $\hat{\hat{\underline{\mu}}}_k$, $\mathrm{var}(\hat{\hat{\underline{\mu}}}_k)$, is given by

$$\mathrm{var}(\hat{\mu}_k) = (F^T V_k^{-1} F)^{-1} \tag{3.62}$$

where the matrix F is defined by eqn.(3.59). If we set p=1, then all these results reduce to the case of two antithetic variates as discussed in section 3.3.1.

3.4 Nonlinear predictors using the control variate method

The fundamental concepts of the control variate method have been discussed in section 2.4.1. They will now be applied to the state variable prediction problem. In doing this we are required to solve the following two problems:

(1) find a linear model for the given nonlinear transformation (3.1) such that conditions (C1) and (C2) of section 2.4.1 are satisfied;

(2) use this model in a modified sampling procedure in order to improve the accuracy of the crude Monte Carlo estimator.

The new feature of the solutions presented is that a combination of analytic and statistical methods yields a predictor whose accuracy is determined by the sampling error. Unlike an entirely analytic approximation with its inherent approximation error the sampling error depends on the sample size of the Monte Carlo experiment and can sometimes be made smaller than the error belonging to the analytic solution.

We shall discuss several methods to establish a control variate

model. First, in section 3.4.1 a statistical linearization procedure is used to derive an analytic approximation similar to that one which we could derive from eqn.(3.23). In section 3.4.2 the linear regression technique is used to find a linear weighting sequence model. A two-stage control variate method using a gradient technique is contained in sections 3.4.3 and 3.4.4.

3.4.1 Statistical linearization

The method of statistical linearization to deal with nonlinear stochastic processes is similar to the describing function technique used for the analysis of deterministic nonlinear systems. Pervozvanskii[63] applies this method to nonlinear transformations of stochastic processes with stationary and non-stationary states. The underlying idea of this method is to assume that it is necessary to preserve only that term which is proportional to the covariance matrix Σ_{w_k} of the input signal $\underline{w}_k$ in the expression for the covariance matrix Σ_k of the output signal $\underline{x}_k$. This assumption considerably simplifies the analytic part and involves only the mean and the covariance matrix.

In the present context statistical linearization represents the first step toward a Monte Carlo solution. Indeed, the linearization procedure yields a model which makes an analytic approximation to the nonlinear prediction problem convenient. However, for our purposes this model only serves as a reference to the original nonlinear system and is used in the control variate method to improve the crude Monte

Carlo estimator of section 3.2.2 and possibly, as well, the analytic approximation based on statistical linearization alone.

Let us consider a nonlinear stochastic process of the form

$$\underline{x}_{k+1} = \underline{f}(\underline{x}_k, k) + \underline{w}_k \,. \tag{3.63}$$

The notation is the same as in eqn.(3.1) and we assume that $p(\underline{x}_1)$ and $p(\underline{w}_k)$ are specified by eqn.(3.37) and (3.38). This implies that the P.D.F. $p_a(\underline{x}_k^*)$ is Gaussian provided $\underline{x}_k^*$ denotes the state of a linear model with the same initial condition P.D.F. $p(\underline{x}_1)$ and noise P.D.F. $p(\underline{w}_k)$ as is used in eqn.(3.63). Let us specify this new P.D.F. $p_a(\underline{x}_k^*)$ as

$$p_a(\underline{x}_k^*) = n(\underline{x}_k^*; \underline{m}_k, P_k). \tag{3.64}$$

The problems to be solved are:

(1) to determine the parameters of the linear model whose states are $\underline{x}_k^*$;

(2) to find the mean $\underline{m}_k$ and the covariance matrix P_k of eqn.(3.64) in terms of the model parameters.

A linear model of eqn.(3.63) is obtained by expanding the nonlinear function $\underline{f}$ in a Taylor series around the mean $\underline{m}_k$:

$$\underline{f}(\underline{x}_k, k) = \underline{a}_k + B_k(\underline{x}_k - \underline{m}_k) + \underline{\varepsilon}_k \tag{3.65}$$

where $\underline{\varepsilon}_k$ denotes the collection of error terms. We shall follow Sunahara's procedure[16] and determine $\underline{a}_k$ and B_k in such a way that the expectation of the norm of $\underline{\varepsilon}_k$, defined by

$$\begin{aligned} E_{p_a}[\|\underline{\varepsilon}_k\|^2] &\triangleq E_{p_a}[\underline{\varepsilon}_k^T \underline{\varepsilon}_k] \\ &= E_{p_a}[\|\underline{f}(\underline{x}_k^*, k) - \underline{a}_k - B_k(\underline{x}_k^* - \underline{m}_k)\|^2] \end{aligned} \tag{3.66}$$

is minimal w.r.t. $\underline{a}_k$ and B_k. The expectation $E_{p_a}[\,.\,]$ is defined w.r.t. $p_a(\underline{x}_k^*)$ of eqn.(3.64). Hence, the necessary and sufficient conditions to minimize eqn.(3.66) are directly obtainable as

$$\underline{a}_k = E_{p_a}[\,\underline{f}(\underline{x}_k^*, k)] \tag{3.67}$$

$$B_k = E_{p_a}[(\underline{f}(\underline{x}_k^*, k) - \underline{a}_k)(\underline{x}_k^* - \underline{m}_k)^T]\, P_k^{-1} \tag{3.68}$$

where

$$P_k = E_{p_a}[(\underline{x}_k^* - \underline{m}_k)(\underline{x}_k^* - \underline{m}_k)^T] \tag{3.69}$$

is the covariance matrix of $p_a(\underline{x}_k^*)$ as mentioned before. It is worthwhile pointing out that both $\underline{a}_k$ and B_k depend on k, the mean $\underline{m}_k$ and the covariance matrix P_k.

It follows from the Gaussian property of $p_a(\underline{x}_k^*)$ that the elements of B_k are obtained as

$$B_k = \frac{\partial \underline{a}_k}{\partial \underline{m}_k} \quad . \tag{3.70}$$

Indeed, using eqn.(3.64) in eqn.(3.67) yields

$$\frac{\partial \underline{a}_k}{\partial \underline{m}_k} = \frac{\partial}{\partial \underline{m}_k} \int \underline{f}(\underline{x}_k^*, k) \frac{1}{\sqrt{(2\pi)^n\,|P_k|}} \exp\left[-\frac{1}{2}(\underline{x}_k^* - \underline{m}_k)^T P_k^{-1} (\underline{x}_k^* - \underline{m}_k)\right] d\underline{x}_k^*$$

$$= \int \underline{f}(\underline{x}_k^*, k)(\underline{x}_k^* - \underline{m}_k)^T P_k^{-1} p_a(\underline{x}_k^*)\, d\underline{x}_k^*$$

$$= B_k \,. \tag{3.71}$$

Thus, a linear model is obtained in the form

$$\underline{x}_{k+1}^* = \underline{a}_k + B_k(\underline{x}_k^* - \underline{m}_k) + \underline{w}_k \,. \tag{3.72}$$

Recalling the analytic solution (3.20) it follows from eqn.(3.72) that

the mean $\underline{m}_k$ and covariance matrix P_k are given by

$$\begin{aligned} \underline{m}_{k+1} &= \underline{a}_k \\ P_{k+1} &= B_k P_k B_k^T + \Sigma_{w_k} . \end{aligned} \tag{3.73}$$

These recursive equations start at time k=1 with

$$\begin{aligned} \underline{m}_1 &= \underline{m}_x \\ P_1 &= \Sigma_x \end{aligned} \tag{3.74}$$

that is, $p_a(\underline{x}_1^*) = p(\underline{x}_1)$.

Having completed the analytic approximation we now derive a Monte Carlo predictor for the mean $\underline{\mu}_k$ using control variates. To this end we break the integral in eqn.(3.27) into two parts such that

$$\underline{\mu}_k \triangleq E[\underline{x}_k] = \int \underline{x}_k \, p(\underline{x}_k) \, d\underline{x}_k - \int \underline{x}_k^* \, p_a(\underline{x}_k^*) \, d\underline{x}_k^* + \underline{m}_k . \tag{3.75}$$

where of course $\underline{m}_k = \int \underline{x}_k^* \, p_a(\underline{x}_k^*) \, d\underline{x}_k^*$.

The first two integrals are solved by crude Monte Carlo. Hence the new estimator for $\underline{\mu}_k$ takes the form

$$\hat{\underline{\mu}}_k = N^{-1} \sum_j^N \left[(\underline{x}_k)_j - (\underline{x}_k^*)_j \right] + \underline{m}_k , \tag{3.76}$$

where $(\underline{x}_k)_j$ denotes the state $\underline{x}_k$ obtained by simulation of eqn.(3.63); $(\underline{x}_1)_j$ is drawn from $p(\underline{x}_1)$, $(\underline{w}^{k-1})_j$ is drawn from $p(\underline{w}_k)$ and $(\underline{x}_k)_j$ is obtained as the solution of eqn.(3.63). In order to obtain the required positive correlation between $(\underline{x}_k)_j$ and $(\underline{x}_k^*)_j$ the same random sample $\{\underline{x}_1, \underline{w}^{k-1}\}_j$ of size N is used to generate $\{\underline{x}_k^*\}_j$ as the N solutions of the linear stochastic difference eqn.(3.72). This is equivalent to drawing $(\underline{x}_k^*)_j$ from $p_a(\underline{x}_k^*)$.

Equation (3.76) shows that $(\underline{x}_k^*)_j$ acts as a control variate of $(\underline{x}_k)_j$. A reduced sampling covariance matrix $\text{var}(\hat{\underline{\mu}}_k)$ is obtained provided that the linear model (3.72) giving rise to the control variate is a close approximation to the original nonlinear system and absorbs most of the variations in the sampling procedure. This, however, is ensured by the more elaborate statistical linearization procedure of eqns.(3.67) and (3.68) rather than, for example, linearization along a reference trajectory obtained by solving eqn.(3.63) with $\underline{w}_k = 0$, for all k.

The sampling covariance matrix of the new predictor (3.76) is given by

$$\text{var}(\hat{\underline{\mu}}_k) = N^{-1} \; E\left[(\underline{\xi}_k - \underline{\mu}_k)\,(\underline{\xi}_k - \underline{\mu}_k)^T\right] \tag{3.77}$$

where $\underline{\xi}_k$ is defined by

$$\underline{\xi}_k \triangleq \underline{x}_k - \underline{x}_k^* + \underline{m}_k \tag{3.78}$$

As $\underline{\mu}_k$ is unknown $\text{var}(\hat{\underline{\mu}}_k)$ has to be estimated by

$$\hat{\text{var}}(\hat{\underline{\mu}}_k) = N^{-2} \sum_j^N ((\underline{\xi}_k)_j - \hat{\underline{\mu}}_k)((\underline{\xi}_k)_j - \hat{\underline{\mu}}_k)^T. \tag{3.79}$$

In order to elaborate on the accuracy improvement on the crude Monte Carlo predictor (3.30) let us assume that all quantities in eqns.(3.77) and (3.78) are scalars and furthermore let us rewrite eqn. (3.78) with a scaling factor $\lambda \triangleq 1 - x_k^*/x_k$ as

$$\xi_k = \lambda x_k + m_k \,. \tag{3.80}$$

We make the assumption that λ is constant but in actual practice it depends on k and x_k. Using eqn.(3.80) in (3.77) yields

$$\begin{aligned} \mathrm{var}(\hat{\mu}_k) &= N^{-1} \ \mathrm{var}(\xi_k) \\ &= N^{-1} \{ E[(\lambda x_k + m_k)^2] - (E[\lambda x_k + m_k])^2 \} \\ &= N^{-1} \lambda^2 \ \mathrm{var}(x_k). \end{aligned} \tag{3.81}$$

Using eqn.(3.35), the variance reduction η_v, see eqn.(2.71), from crude Monte Carlo to the control variate method presented is

$$\eta_v = 1/\lambda^2 \ . \tag{3.82}$$

Fig. 3.4 exhibits the quadratic nature of η_v as a function of λ.

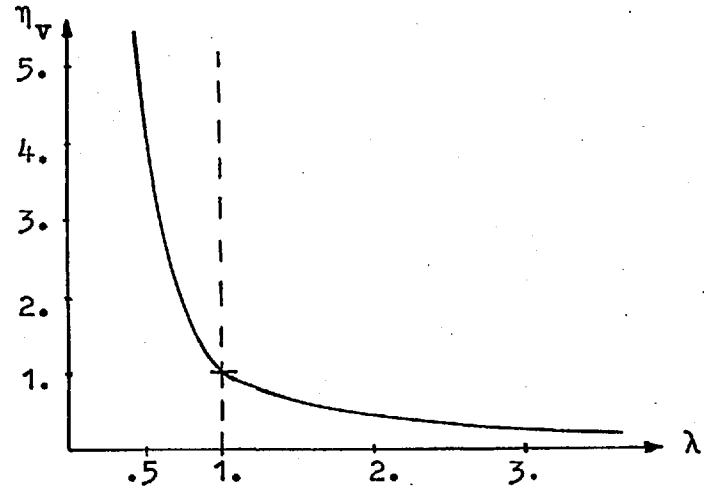

Fig. 3.4 Variance reduction η_v for the control variate method.

The case $\lambda = 1$ corresponds to $(x_k^*)_j = 0$, i.e. crude Monte Carlo. On the other hand, $\lambda = 0$ implies $(x_k^*)_j \equiv (x_k)_j$ which corresponds to an estimator with zero sampling variance. But for $\lambda > 1$, the control variate method does not yield any variance reduction.

Next, we consider the matrix M_k of second order moments which is defined in eqn.(3.28). Again we break the integral into two parts and

obtain the following estimator for $\hat{M}_k$

$$\hat{M}_k = N^{-1} \sum_{j}^{N} [(\underline{x}_k)_j (\underline{x}_k)_j^T - (\underline{x}_k^*)_j (\underline{x}_k^*)_j^T] + \theta_k^{(2)}. \qquad (3.83)$$

The matrix $\theta_k^{(2)}$ of second order moments defined for the linear model (3.72) is found from eqn.(3.73) to be

$$\theta_k^{(2)} \triangleq \int \underline{x}_k^* . \underline{x}_k^{*T} \, p_a(\underline{x}_k^*) d\underline{x}_k^* = P_k + \underline{m}_k \underline{m}_k^T . \qquad (3.84)$$

Finally, an estimate $\hat{\Sigma}_k$ of the population covariance matrix $\Sigma_k \triangleq \text{var}(\underline{x}_k)$ is obtained from eqn.(3.76) and (3.83) by

$$\hat{\Sigma}_k = \hat{M}_k - \hat{\underline{\mu}}_k \hat{\underline{\mu}}_k^T . \qquad (3.85)$$

As already mentioned in section 3.3.2 a useful requirement for a Monte Carlo procedure is that it gives zero sampling variance when applied to a linear Gaussian system. We have the following result:

> For a linear Gaussian system of the form (3.18) the control variate estimator (3.76) for the mean $\underline{\mu}_k$ and (3.83) for the second order moment M_k has sampling variances equal to zero.

To prove this result we first show that the statistical linearization procedure of eqns.(3.67) and (3.68) yields a model (3.72) which is identical to the original system (3.18). Indeed, it follows from eqn. (3.67) that $\underline{a}_k = A_k \underline{m}_k$ and using eqn.(3.70) we have $B_k = A_k$. Thus, eqn.(3.72) becomes

$$\underline{x}_{k+1}^* = A_k \underline{x}_k^* + \underline{w}_k , \qquad (3.86)$$

which is identical to eqn.(3.18). This, however, implies $p(\underline{x}_k) \equiv p_a(\underline{x}_k^*)$ and therefore $(\underline{x}_k)_j \equiv (\underline{x}_k^*)_j$ for any j provided we use the same random

sequence $(\underline{x}_1, \underline{w}^{k-1})_j$ both in eqn.(3.18) and in (3.86). In fig.3.4 this corresponds to the case $\lambda = 0$. Thus, there is no randomness in the estimators (3.76) and (3.83) and the proof of the above result is complete.

It is interesting to compare this result with that obtained in section 3.3.2. The control variate method yields zero sampling variance for both the first and second order moment estimates when applied to a linear system with Gaussian noise. By contrast, the antithetic variate method has this property only for the first order moment. While this comparison does not indicate which method is superior in a particular nonlinear system for predicting the mean $\hat{\underline{\mu}}_k$ it shows that the control variate method is to be preferred for estimating the second order moment $\hat{M}_k$.

A final remark is concerned with the evaluation of eqns.(3.67) and (3.68). On the assumption that $\underline{f}(\underline{x}_k, k)$ can be expressed as a polynomial in $\underline{x}_k$ (possibly with time-varying coefficients) the right hand side of eqns.(3.67) and (3.68) can be evaluated analytically. For higher order terms, Price's theorem, see Papoulis[35] page 226, is helpful to compute the r:th joint moment. For arbitrary functions $\underline{f}(\underline{x}_k, k)$ the integration can always be performed by Monte Carlo techniques with a random sample $\{\underline{x}_k^*\}_j$ drawn from $p_a(\underline{x}_k^*)$ if the dimensionality of $\underline{x}_k^*$ excludes standard numerical integration routines.

3.4.2 The control variate method using linear regression

Statistical linearization is by no means the only way to establish a linear model. In this section we shall be concerned with the linear regression technique[57] in order to find a control variate model. We confine ourselves to a scalar stochastic process of the form

$$x_{k+1} = f(x_k, w_k, k). \tag{3.87}$$

It exhibits all the features we are interested in without undue ramifications appropriate in the multivariate case.

No explicit use so far has been made of the specific property of Monte Carlo methods that we know the random sample $\{x_1, w^{k-1}\}_j$ used to simulate a process of the form of eqn.(3.87). Thus, an obvious proposition is to approximate x_k by x_k^* which itself is obtained from a linear scalar weighting sequence of the form

$$x_k^* = \sum_i^{k-1} \alpha_i w_i + \alpha_k x_1 . \tag{3.88}$$

Combining the parameters α_i in the vector $\underline{\alpha}_k$, defined by

$$\underline{\alpha}_k^T \triangleq [\alpha_1, \alpha_2, \ldots, \alpha_k] , \tag{3.89}$$

and the independent variates w^{k-1} and x_1 in the vector $\underline{\omega}_k$, defined by

$$\underline{\omega}_k^T \triangleq [w_1, w_2, \ldots, w_{k-1}, x_1], \tag{3.90}$$

eqn.(3.88) can be written as

$$x_k^* = \underline{\alpha}_k^T \underline{\omega}_k = \underline{\omega}_k^T \underline{\alpha}_k . \tag{3.91}$$

It is desired to estimate the vector $\underline{\alpha}_k$ such that the difference between the system state x_k and the model state x_k^* is small. To this end we have to simulate eqn.(3.87) with a random sample $\{\underline{\omega}_k\}_j$ of size N_1.

The components of the random vector $(\underline{\omega}_k)_j$ are defined by eqn.(3.90): $(w_i)_j$ is drawn from $p(w_i)$, $(i=1,2,\ldots,k-1)$, and $(x_1)_j$ is drawn from $p(x_1)$. Then $\{\underline{\omega}_k\}_j$ is stored in the (N_1,k)-dimensional matrix W_k, defined by

$$W_k \triangleq \begin{bmatrix} (\underline{\omega}_k)_1^T \\ \vdots \\ (\underline{\omega}_k)_{N_1}^T \end{bmatrix} . \tag{3.92}$$

The resulting N_1 solutions of eqn.(3.87) are stored in the N_1 - dimensional random vector $\underline{\xi}_k^T \triangleq [(x_k)_1, \ldots (x_k)_{N_1}]$. Since the linear model (3.91) holds for any random vector $(\underline{\omega}_k)_j$ we have

$$\underline{\xi}_k^* = W_k \underline{\alpha}_k \tag{3.93}$$

where $\underline{\xi}_k^{*T} \triangleq [(x_k^*)_1, \ldots, (x_k^*)_{N_1}]$. The least squares linear unbiased estimate of $\underline{\alpha}_k$ is defined as the vector $\hat{\underline{\alpha}}_k^o$ which minimizes the quadratic risk function $R(\underline{\alpha}_k)$ w.r.t. $\underline{\alpha}_k$; $R(\underline{\alpha}_k)$ is defined by

$$R(\underline{\alpha}_k) \triangleq (\underline{\xi}_k - W_k \underline{\alpha}_k)^T (\underline{\xi}_k - W_k \underline{\alpha}_k). \tag{3.94}$$

Differentiating eqn.(3.94) w.r.t. $\underline{\alpha}_k$ and equating the result to zero yields

$$\hat{\underline{\alpha}}_k^o = (W_k^T W_k)^{-1} W_k^T \underline{\xi}_k . \tag{3.95}$$

For $(W_k^T W_k)$ to be nonsingular it is necessary to have $N_1 \geq k$.

Returning to the control variate method we replace $\underline{\alpha}_k$ in eqn.(3.91) by $\hat{\underline{\alpha}}_k^o$ of eqn.(3.95). That is, the control variate model takes the form

$$x_k^* = \hat{\underline{\alpha}}_k^{oT} \underline{\omega}_k . \tag{3.96}$$

Referring to eqn.(3.75) the estimator for the sample mean $\hat{\mu}_k$ is now defined by

$$\hat{\mu}_k = N_2^{-1} \sum_{j}^{N_2} [(x_k)_j - (x_k^*)_j] + \hat{\alpha}_k^o \, m_x \, , \qquad (3.97)$$

where m_x denotes the mean of $p(x_1)$. The Monte Carlo procedure to evaluate eqn.(3.97) is based on the random sample $\{x_1, w^{k-1}\}_j$ of size N_2. Solving eqns.(3.87) and (3.96) with the j:th realization of this random sample yields $(x_k)_j$ and $(x_k^*)_j$ respectively. The last term in eqn.(3.97) is given by the analytic evaluation of the mean $E[x_k^*] = \hat{\alpha}_k^o \, m_x$ based on eqn.(3.96).

The sampling variance of $\hat{\mu}_k$ is directly obtainable from eqn.(3.97):

$$\mathrm{var}(\hat{\mu}_k) = N_2^{-1} \, [\, \mathrm{var}(x_k) + \mathrm{var}(x_k^*) - 2 \, \mathrm{cov}(x_k, x_k^*) \,] \quad , \qquad (3.98)$$

where in actual practice all terms on the R.H.S. have to be replaced by their estimates.

Referring to eqn.(3.83), the second order moment M_k is estimated by

$$\hat{M}_k = N_2^{-1} \sum_{j}^{N_2} [(x_k)_j^2 - (x_k^*)_j^2] + \hat{\alpha}_k^o (\Sigma_x + m_x^2) + \sum_{i}^{k-1} [\hat{\alpha}_i^o \, \Sigma_{w_i}] . \qquad (3.99)$$

The meaning of $(x_k)_j$ and $(x_k^*)_j$ is explained in the context of eqn.(3.97). The last two terms in eqn.(3.99) represent the analytic solution for the second order moment based on the control variate model (3.96). The parameters m_x, Σ_x and Σ_{w_i}, $(i=1,2,\ldots,k-1)$ are given by $p(x_1)$ and $p(w_i)$.

In the following section we shall develop a method similar to the one just presented. Indeed, at first sight the difference between the

two approaches may seem rather superficial. There is, however, a fundamental difference in estimating $\hat{\underline{a}}_k^o$.

3.4.3 A two-stage control variate predictor using a gradient technique

The concept of a multi-stage control variate method has been introduced in section 2.4. The object of this section is to derive a predictor for the mean μ_k using a two-stage control variate method. For the sake of clarity we restrict ourselves again to the scalar nonlinear system (3.87). The extension to the multidimensional case is contained in the section 3.4.4.

The first stage of the control variate estimator is concerned with the determination of the parameter $\underline{a}_k$ in the linear weighting sequence model (3.91) in order to make the model state x_k^* a close approximation to the nonlinear system state x_k. The random vector $\underline{\omega}_k$ is again defined by eqn.(3.90).

The assumptions

(1) x_1 is normally distributed and uncorrelated with w^k;

(2) w^k is a Gaussian white noise sequence imply that the P.D.F. $p(\underline{\omega}_k)$, defined by

$$p(\underline{\omega}_k) \triangleq \prod_{i=1}^{k-1} p(w_i)\, p(x_1), \tag{3.100}$$

is a k-variate normal P.D.F.

$$p(\underline{\omega}_k) = n(\underline{\omega}_k ; \underline{a}, \Sigma_\omega) . \tag{3.101}$$

The mean $\underline{a}$ and the covariance matrix Σ_ω are given by

$$\underline{a}^T = [\,0,\ 0,\ \ldots\ ,\ 0,\ m_x\,]$$

$$\Sigma_\omega = \begin{bmatrix} \Sigma_{w_1} & & & & 0 \\ & \ddots & & & \\ & & \Sigma_{w_{k-1}} & & \\ 0 & & & & \Sigma_x \end{bmatrix} \tag{3.102}$$

where m_x, Σ_x and Σ_{w_k} are scalars, see eqns.(3.37) and (3.38).

The parameter $\underline{a}_k$ is said to be optimal, denoted by $\underline{a}_k^o$, if the variance of the error e_k, defined as the difference between system and model state, i.e.

$$e_k \triangleq x_k - x_k^* \ , \tag{3.103}$$

is minimal. Thus, the functional to be minimized is

$$F(\underline{a}_k) = \mathrm{var}(e_k) = \mathrm{var}(x_k) + \mathrm{var}(x_k^*) - 2\ \mathrm{cov}(x_k,\ x_k^*). \tag{3.104}$$

Since $\mathrm{var}(x_k^*)$ is quadratic and $\mathrm{cov}(x_k,\ x_k^*)$ linear in $\underline{a}_k$ $F(\underline{a}_k)$ is a quadratic form in $\underline{a}_k$. Therefore the difference between $\underline{a}_k^o$ and an arbitrary value $\underline{a}_k$ is given by

$$\underline{a}_k^o - \underline{a}_k = -F_{a_k a_k}^{-1}\ \underline{F}_{a_k} \tag{3.105}$$

where $\underline{F}_{a_k}$ is the gradient and $F_{a_k a_k}$ is the matrix of second order derivatives of $F(\underline{a}_k)$ w.r.t. $\underline{a}_k$.

Using eqns.(3.91), (3.101) and (3.104) the gradient $\underline{F}_{a_k}$ is given by

$$\underline{F}_{\underline{\alpha}_k} = 2\int \underline{\omega}_k \underline{\omega}_k^T \underline{\alpha}_k p(\underline{\omega}_k) d\underline{\omega}_k - 2\int \underline{\alpha}_k^T \underline{\omega}_k p(\underline{\omega}_k) d\underline{\omega}_k \int \underline{\omega}_k p(\underline{\omega}_k) d\underline{\omega}_k$$

$$-2\int x_k \underline{\omega}_k p(x_k, \underline{\omega}_k) dx_k d\underline{\omega}_k + 2\int x_k p(x_k) dx_k \int \underline{\omega}_k p(\underline{\omega}_k) d\underline{\omega}_k$$

$$= 2 \Sigma_\omega \underline{\alpha}_k - 2E[x_k \underline{\omega}_k] + 2E[\underline{\omega}_k] E[x_k] . \qquad (3.106)$$

The matrix $F_{\underline{\alpha}_k \underline{\alpha}_k}$ is obtained by differentiating (3.106) w.r.t. $\underline{\alpha}_k$:

$$F_{\underline{\alpha}_k \underline{\alpha}_k} = 2 \Sigma_\omega . \qquad (3.107)$$

Since the P.D.F. $p(x_k)$ and $p(x_k, \underline{\omega}_k)$ are unknown in eqn.(3.106) the gradient $\underline{F}_{\underline{\alpha}_k}$ has to be found by Monte Carlo integration with a random sample $\{\underline{\omega}_k\}_j$ of size N_1 . The estimate $\hat{\underline{F}}_{\underline{\alpha}_k}$ is given by

$$\hat{\underline{F}}_{\underline{\alpha}_k} = 2 \Sigma_\omega \underline{\alpha}_k - 2N_1^{-1} \sum_j^{N_1} [(x_k)_j (\underline{\omega}_k)_j] + 2\underline{a}\, N_1^{-1} \sum_j^{N_1} (x_k)_j \qquad (3.108)$$

where $(x_k)_j$ denotes the solution of eqn.(3.87) obtained with the j:th random sequence realization $(\underline{\omega}_k)_j$.

As mentioned in section 2.4 the Monte Carlo integration of those terms in eqn.(3.108) which are predominantly odd in $\underline{\omega}_k$ may be improved by means of the antithetic variate method. Using a random sample of size N_1 we have

$$\hat{\underline{F}}_{\underline{\alpha}_k} = 2 \Sigma_\omega \underline{\alpha}_k - N_1^{-1} \sum_j^{N_1} [(x_k)_j^+ (\underline{\omega}_k)_j^+ + (x_k)_j^- (\underline{\omega}_k)_j^-] + \underline{a}\, N_1^{-1} \sum_j^{N_1} [(x_k)_j^+ + (x_k)_j^-] \qquad (3.109)$$

where $((\underline{\omega}_k)_j^- - \underline{a}) = -((\underline{\omega}_k)_j^+ - \underline{a})$ and $(x_k)_j^{\pm}$ denotes the j:th solution of eqn.(3.87) using $(\underline{\omega}_k)_j^{\pm}$.

As the gradient $\underline{F}_{\underline{\alpha}_k}$ has to be replaced by $\hat{\underline{F}}_{\underline{\alpha}_k}$, eqn.(3.105) has

to be replaced by a stochastic approximation procedure of the form

$$\hat{\underline{\alpha}}_{k,p+1} = \hat{\underline{\alpha}}_{k,p} - F^{-1}_{\alpha_k \alpha_k} \hat{\underline{F}}_{\alpha_k,p} \quad . \tag{3.110}$$

Here k is the time argument as in eqn.(3.87) and p denotes the iteration number of the algorithm. For recursion (3.110) to converge we compute the auxiliary constant $\hat{\underline{c}}_{k,p}$, defined by

$$\hat{\underline{c}}_{k,p} \triangleq F_{\alpha_k \alpha_k} \hat{\underline{\alpha}}_{k,p} - \hat{\underline{F}}_{\alpha_k,p} \quad . \tag{3.111}$$

Although $\hat{\underline{F}}_{\alpha_k,p}$ is estimated with a sample of fixed size N_1 the sampling covariance matrix of the combined estimate $\hat{\hat{\underline{c}}}_{k,p}$ decreases with increasing p. Feasible weighting procedures to compute $\hat{\hat{\underline{c}}}_{k,p}$ from $\hat{\underline{c}}_{k,1}$, $\hat{\underline{c}}_{k,2}, \ldots, \hat{\underline{c}}_{k,p}$ are discussed in section 2.4.4. The updating procedure (3.110) is replaced by

$$\hat{\underline{\alpha}}_{k,p+1} = F^{-1}_{\alpha_k \alpha_k} \hat{\hat{\underline{c}}}_{k,p} \quad . \tag{3.112}$$

Because the sequence $\hat{\hat{\underline{c}}}_{k,1}$, $\hat{\hat{\underline{c}}}_{k,2}$, ... converges with probability one to $\underline{c}_k \triangleq F_{\alpha_k \alpha_k} \underline{\alpha}^o_k$ and $F^{-1}_{\alpha_k \alpha_k}$ is a deterministic multiplier the sequence $\hat{\underline{\alpha}}_{k,1}$, $\hat{\underline{\alpha}}_{k,2}, \ldots$ based on eqn.(3.112) converges with probability one to $\underline{\alpha}^o_k$.

The comparison of eqns.(3.95) and (3.112) shows the main difference between our adaptive control variate method using a gradient technique and the linear regression method: the deterministic matrix $F_{\alpha_k \alpha_k}$ is computed analytically before the sampling procedure for estimating $\hat{\underline{\alpha}}^o_k$ starts whereas W_k of eqn.(3.95) is a matrix of random samples.

After a sufficient number of iterations of eqn.(3.112) we replace

$\underline{\alpha}_k$ in the control variate model (3.91) by $\hat{\underline{\alpha}}_{k,p}$. Drawing a random sample $\{x_1, w^{k-1}\}_j$ of size N_2 the mean μ_k of $p(x_k)$ is estimated, as in section 3.4.2, by

$$\hat{\mu}_k = N_2^{-1} \sum_j^{N_2} [(x_k)_j - (x_k^*)_j] + \hat{\underline{\alpha}}_{k,p}^T \underline{a} \,. \qquad (3.113)$$

The meaning of $(x_k)_j$ and $(x_k^*)_j$ is explained in the context of eqn. (3.97) and the vector $\underline{a}$ is defined by eqn.(3.102).

The computing routine of the adaptive control variate method is summarized as follows:

(1) At p=1, set the parameter $\underline{\alpha}_k$ to an arbitrary value $\underline{\alpha}_{k,p}$ and compute the matrix $F^{-1}_{\alpha_k \alpha_k}$ with eqns.(3.102) and (3.107).

(2) Draw a random sample $\{\underline{\omega}_k\}_j$ of fixed size N_1 and estimate the gradient $\hat{\underline{F}}_{\alpha_{k,p}}$ with eqns.(3.108) or (3.109).

(3) Estimate the vector $\hat{\underline{c}}_{k,p}$ with eqn.(3.111), determine the combined estimate $\hat{\hat{\underline{c}}}_{k,p}$ with a suitable weighting sequence and compute $\hat{\underline{\alpha}}_{k,p+1}$ with eqn.(3.112).

(4) Set p=p+1. If the sampling error of $\hat{\underline{\alpha}}_{k,p}$ is sufficiently small, go to (5); otherwise return to (2).

(5) Draw a random sample $\{\underline{\omega}_k\}_j$ of size N_2 and estimate the mean μ_k with eqn.(3.113). The sampling variance of $\hat{\mu}_k$ is given by eqn.(3.98).

3.4.4 Extension to the n-dimensional case

The two-stage control variate method introduced in the previous section is now extended to the n-dimensional case. Let $\underline{\mu}_k$ denote the mean of $p(\underline{x}_k)$ where $\underline{x}_k$ evolves according to the stochastic difference equation (3.1). It is desired to develop the two-stage control variate method to estimate the mean $\hat{\underline{\mu}}_k$. For the sake of simplicity we assume $\underline{x}_k$ and $\underline{w}_k$ to be of the same dimension.

In the linear control variate model of the form

$$\underline{x}_k^* = \sum_{i}^{k-1} A_i \underline{w}_i + A_k \underline{x}_1 \tag{3.114}$$

it is desired to determine the (nk,n)-dimensional matrix $\mathcal{A}_k$, defined by $\mathcal{A}_k^T \triangleq [A_1, A_2, \ldots, A_k]$, such that for a fixed time k the model state $\underline{x}_k^*$ is a close approximation to $\underline{x}_k$, the state of the nonlinear process (3.1). Under the assumption that

(1) $\underline{x}_1$ is normally distributed (see eqn.(3.37)) and is uncorrelated with $\underline{w}^k$,

(2) $\underline{w}^k$ is a Gaussian white noise sequence (see eqn.(3.38)),

the vector $\underline{\omega}_k^T \triangleq [\underline{w}_1^T, \underline{w}_2^T, \ldots, \underline{w}_{k-1}^T, \underline{x}_1^T]$ possesses a normal (nk)-variate P.D.F. $p(\underline{\omega}_k)$ given by

$$p(\underline{\omega}_k) = n(\underline{\omega}_k; \underline{a}, \Sigma_\omega) \tag{3.115}$$

where the mean $\underline{a}$ is given by

$$\underline{a}^T = [\underline{0}^T, \underline{0}^T, \ldots, \underline{0}^T, \underline{m}_x^T], \tag{3.116}$$

and the covariance matrix Σ_ω is given by

$$\Sigma_\omega = \begin{bmatrix} \Sigma_{w_1} & & 0 \\ & \ddots & \\ & & \Sigma_{w_{k-1}} \\ 0 & & & \Sigma_x \end{bmatrix} . \qquad (3.117)$$

The linear model (3.114) is a close approximation to the nonlinear process (3.1) if some measure of the covariance matrix $\text{var}(\underline{e}_k) \triangleq \text{var}(\underline{x}_k - \underline{x}_k^*)$ is minimized w.r.t. a_k. With eqn.(3.114) rewritten as

$$\underline{x}_k^* = a_k^T \underline{\omega}_k \qquad (3.118)$$

the covariance matrix $\text{var}(\underline{e}_k)$ is given by

$$\begin{aligned} \text{var}(\underline{e}_k) = {} & E[a_k^T \underline{\omega}_k \underline{\omega}_k^T a_k] - E[a_k^T \underline{\omega}_k] \, E[\underline{\omega}_k^T a_k] \\ & -2E[\underline{x}_k \underline{\omega}_k^T a_k] - 2E[\underline{x}_k] \, E[\underline{\omega}_k^T a_k] \\ & + \text{terms independent of } a_k . \end{aligned} \qquad (3.119)$$

Let us define the optimal value of a_k, denoted by a_k^o, such that a_k^o minimizes the trace of $\text{var}(\underline{e}_k)$. The i:th column of the matrix a_k is denoted by $\underline{a}_k^i$. Then, the functional $F(a_k)$ to be minimized can be written as

$$\begin{aligned} F(a_k) = \sum_i^n \{ & E[(\underline{a}_k^i)^T \underline{\omega}_k \underline{\omega}_k^T \underline{a}_k^i] - E[(\underline{a}_k^i)^T \underline{\omega}_k] \, E[\underline{\omega}_k^T \underline{a}_k^i] \\ & -2E[x_k(i) \underline{\omega}_k^T \underline{a}_k^i] - 2E[x_k(i)] \, E[\underline{\omega}_k^T \underline{a}_k^i] \} \\ = {} & \sum_i^n F_i(\underline{a}_k^i) . \end{aligned} \qquad (3.120)$$

As the n terms in eqn.(3.120)are uncoupled the optimal values of $\underline{\alpha}_k^i$, (i=1,2,...,n), denoted by $\underline{\alpha}_k^{oi}$, are given by

$$\underline{\alpha}_k^{oi} = \underline{\alpha}_k^i - F^{-1}_{\alpha_k^i \alpha_k^i} \underline{F}_{\alpha_k^i} \quad . \tag{3.121}$$

Using eqn.(3.120) the gradients $\underline{F}_{\alpha_k^i}$ are given by

$$\underline{F}_{\alpha_k^i} = 2 \Sigma_\omega \underline{\alpha}_k^i - 2E[x_k(i)\, \underline{\omega}_k] + 2E[x_k(i)]\, E[\underline{\omega}_k] \tag{3.122}$$

for i=1,2,...,n. The matrices $F_{\alpha_k^i \alpha_k^i}$ of second order derivatives are given by

$$F_{\alpha_k^i \alpha_k^i} = 2 \Sigma_\omega \quad , \text{ for } i=1,2,\ldots,n. \tag{3.123}$$

where Σ_ω is given by eqn.(3.117).

As the evaluation of $\underline{F}_{\alpha_k^i}$ involves the unknown P.D.F. $p(\underline{x}_k)$ and $p(\underline{x}_k, \underline{\omega}_k)$ we have to replace eqn.(3.121) by the stochastic approximation procedure

$$\hat{\underline{\alpha}}^i_{k,p+1} = \hat{\underline{\alpha}}^i_{k,p} - F^{-1}_{\alpha_k^i \alpha_k^i} \hat{\underline{F}}_{\alpha_k^i} \quad , \; i=1,2,\ldots,n. \tag{3.124}$$

Here k denotes the time argument as in eqn.(3.1), p denotes the iteration number of the algorithm and i denotes the column of the matrix α_k (that is, the component of the vector $\underline{\alpha}_k$).

In the p:th iteration the n gradients $\hat{\underline{F}}_{\alpha_k^i}$ are estimated by means of a random sample $\{\underline{\omega}_k\}_j$ of fixed size N_1. The simulation of eqn.(3.1) with $(\underline{\omega}_k)_j$ yields $(\underline{x}_k)_j$. The crude estimator of the gradient,

defined by

$$\hat{\underline{F}}_{\substack{i \\ \alpha_k,p}} = 2 \Sigma \omega \hat{\underline{\alpha}}^i_{k,p} - 2N_1^{-1} \sum_j^{N_1} [(x_k(i))_j (\underline{\omega}_k)_j] + 2\underline{a} N_1^{-1} \sum_j^{N_1} (x_k(i))_j \,, \tag{3.125}$$

may be improved by the antithetic variate estimator, defined by

$$\hat{\underline{F}}_{\substack{i \\ \alpha_k,p}} = 2 \Sigma \omega \hat{\underline{\alpha}}^i_{k,p} - N_1^{-1} \sum_j^{N_1} [(x_k(i))^+_j (\underline{\omega}_k)^+_j + (x_k(i))^-_j (\underline{\omega}_k)^-_j]$$

$$+ \underline{a} N_1^{-1} \sum_j^{N_1} [(x_k(i))^+_j + (x_k(i))^-_j] \,, \tag{3.126}$$

if the terms to be estimated are predominantly odd in $\underline{\omega}_k$. As before we have $((\underline{\omega}_k)^-_j - \underline{a}) = -((\underline{\omega}_k)^+_j - \underline{a})$ and $(x_k(i))^{+}_{-j}$ denotes the i:th component of the j:th solution $(\underline{x}_k)^{+}_{-j}$ of eqn.(3.1) using $(\underline{\omega}_k)^{+}_{-j}$.

In analogy to the scalar case we estimate the auxiliary vectors $\hat{\underline{c}}^i_{k,p}, (i=1,2,\ldots,n)$, in each iteration p by

$$\hat{\underline{c}}^i_{k,p} = F_{\substack{\alpha_k^i \alpha_k^i}} \hat{\underline{\alpha}}^i_{k,p} - \hat{\underline{F}}_{\substack{i \\ \alpha_k,p}} \,. \tag{3.127}$$

The combined estimates $\hat{\hat{\underline{c}}}^i_{k,p}$ are obtained by applying suitable weighting sequences to $\hat{\underline{c}}^i_{k,1}$, $\hat{\underline{c}}^i_{k,2}, \ldots$, $\hat{\underline{c}}^i_{k,p}$. Since $\hat{\hat{\underline{c}}}^i_{k,1}$, $\hat{\hat{\underline{c}}}^i_{k,2}$, ... converge with probability one to $\underline{c}^i_k \triangleq F_{\substack{\alpha_k^i \alpha_k^i}} \underline{\alpha}^{oi}_k$ and $F_{\substack{\alpha_k^i \alpha_k^i}}$ are deterministic multipliers the sequence of parameter estimates $\hat{\underline{\alpha}}^i_{k,1}$, $\hat{\underline{\alpha}}^i_{k,2}, \ldots$, obtained from

$$\hat{\underline{\alpha}}^i_{k,p+1} = F^{-1}_{\substack{\alpha_k^i \alpha_k^i}} \hat{\hat{\underline{c}}}^i_{k,p} \,, \tag{3.128}$$

converge with probability one to $\underline{\alpha}^{oi}_k$.

After the completion of a sufficient number of iterations p the

estimates $\hat{\underline{a}}^{i}_{k,p}$ form the matrix $\hat{a}_{k,p}$, a close approximation to a^{o}_{k}. Using $\hat{a}_{k,p}$ in eqn.(3.118) the control variate estimator for the mean $\underline{\mu}_k$ using a sample $\{\underline{\omega}_k\}_j$ of size N_2 becomes

$$\hat{\underline{\mu}}_k = N_2^{-1} \sum_{j}^{N_2} [(\underline{x}_k)_j - (\underline{x}^*_k)_j] + a^{T}_{k,p} \underline{a} \qquad (3.129)$$

where $\underline{a}$ is given by eqn.(3.116). The solutions of eqns.(3.1) and (3.118) based on $(\underline{\omega}_k)_j$ are denoted by $(\underline{x}_k)_j$ and $(\underline{x}^*_k)_j$ respectively.

Using the abbreviation $(\underline{\xi}_k)_j$, defined by

$$(\underline{\xi}_k)_j \triangleq (\underline{x}_k)_j - (\underline{x}^*_k)_j + \hat{a}^{T}_{k,p} \underline{a} \ , \qquad (3.130)$$

the sampling covariance matrix of the sample mean $\hat{\underline{\mu}}_k$, $\mathrm{var}(\hat{\underline{\mu}}_k)$, is estimated by

$$\hat{\mathrm{var}}(\hat{\underline{\mu}}_k) = N_2^{-2} \sum_{j}^{N_2} [(\underline{\xi}_k)_j - \hat{\underline{\mu}}_k] [(\underline{\xi}_k)_j - \hat{\underline{\mu}}_k]^{T}. \qquad (3.131)$$

3.5 Numerical examples

It is the object of this section to apply and compare several Monte Carlo methods to predict some parameters belonging to the P.D.F. $p(\underline{x}_k)$. This allows us to establish the efficiency gain of the proposed variance reduction techniques over crude Monte Carlo methods. Furthermore we show that sampling techniques may sometimes improve analytic approximations such as the statistical linearization procedure. To this end we consider a scalar example in section 3.5.1 and a three-dimensional stochastic process in section 3.5.2.

3.5.1 Example I

Let us consider the following discrete-time system specified by the scalar difference equation

$$x_{k+1} = x_k - 0.2\, x_k^3 + w_k , \tag{3.132}$$

together with the initial condition P.D.F. $p(x_1)$ specified by

$$p(x_1) = n(x_1;\ 1.0\ ,\ 10^{-3}). \tag{3.133}$$

The noise sequence w^k is assumed to possess the P.D.F. $p(w_k)$ given by

$$p(w_k) = n(w_k;\ 0.\ \ ,\ 5\cdot 10^{-2}). \tag{3.134}$$

The simulation part of the Monte Carlo solution requires the generation of the random sample $\{ x_1, w^{k-1} \}_j$ of size N in order to obtain the sample $\{x_k\}_j$ for k =1,2,... In our example we choose N = 500. The reported results are ensemble values; that is, the basic experiment is repeated several times, usually ten times, and the results are averaged over all ensemble runs. Thus, $\bar{\hat{\mu}}_k$ denotes the ensemble average of ten estimates $\hat{\mu}_k$. Each $\hat{\mu}_k$ is obtained from the appropriate predictor; e.g. from eqn. (3.30) for the crude Monte Carlo method, from (3.41) for the antithetic variate method and from (3.76) for the control variate method.

In order to estimate the sampling variance of $\hat{\mu}_k$ we may use either of the two alternatives:

(1) we perform a statistical analysis of the ten estimates $\hat{\mu}_k$. As the sample considered is small this analysis requires exact sampling theory in order to obtain meaningful results;

(2) we estimate the sampling variance $\mathrm{var}(\hat{\mu}_k)$ with the equations

derived in this chapter for the different Monte Carlo methods: eqn.(3.35) for crude Monte Carlo, (3.43) for the antithetic variate method and (3.77) for the control variate methods.

We adopt the second approach. In each Monte Carlo experiment we estimate $\hat{var}(\hat{\mu}_k)$ with a sample of size N = 500. The ensemble average of ten estimates is denoted by $\overline{\hat{var}(\hat{\mu}_k)}$.

Table 3.1 exhibits the trajectory prediction problem for the mean μ_k of the P.D.F. $p(x_k)$. The accuracy of the estimated mean $\hat{\mu}_k$ is given by the sampling variance $var(\hat{\mu}_k)$. We compare the following methods:

Method A : crude Monte Carlo; see section 3.2.2

Method B : antithetic variate method; see section 3.3.1

Method C : control variate method using statistical linearization; see section 3.4.1

Method D : control variate method with an arbitrary linear weighting model; that is α_i in eqn.(3.88) are arbitrarily chosen and set equal to 0.4 for all i.

The system (3.132) starts at time k=1 and the predicted trajectory extends to k=10.

In order to compare these methods we show in table 3.2 the variance ratios η_v and efficiency gains η of the improved sampling techniques compared with the crude Monte Carlo predictor of the mean μ_k. The labour ratio η_L is required to compute the efficiency gain η. The three variance reduction techniques involve twice as many simulations as the crude Monte Carlo method. This would indicate $\eta_L = 0.5$. However, as we mentioned in section 2.3 the generation

Time	Method A		Method B		Method C		Method D	
k	$\bar{\hat{\mu}}_k$	$\overline{\hat{var}(\hat{\mu}_k)}$	$\bar{\hat{\mu}}_k$	$\overline{\hat{var}(\hat{\mu}_k)}$	$\bar{\hat{\mu}}_k$	$\overline{\hat{var}(\hat{\mu}_k)}$	$\bar{\hat{\mu}}_k$	$\overline{\hat{var}(\hat{\mu}_k)}$
1	1.00037	$2.008.10^{-6}$	1.0	0.	1.0	0.	1.00022	$7.280.10^{-7}$
2	0.80216	$1.002.10^{-4}$	0.79937	$6.403.10^{-9}$	0.79937	$6.381.10^{-9}$	0.80096	$3.601.10^{-5}$
3	0.67844	$1.372.10^{-4}$	0.67344	$2.241.10^{-6}$	0.67336	$2.358.10^{-6}$	0.67585	$4.203.10^{-5}$
4	0.58903	$1.703.10^{-4}$	0.58555	$4.548.10^{-6}$	0.58557	$4.465.10^{-6}$	0.58646	$4.944.10^{-5}$
5	0.52576	$2.036.10^{-4}$	0.51667	$7.315.10^{-6}$	0.51659	$7.934.10^{-6}$	0.52055	$5.685.10^{-5}$
6	0.46640	$2.371.10^{-4}$	0.45962	$1.045.10^{-5}$	0.46019	$1.153.10^{-5}$	0.46158	$6.642.10^{-5}$
7	0.41602	$2.767.10^{-4}$	0.41044	$1.338.10^{-5}$	0.41139	$1.498.10^{-5}$	0.41136	$7.865.10^{-5}$
8	0.37223	$3.032.10^{-4}$	0.36674	$1.599.10^{-5}$	0.36813	$1.844.10^{-5}$	0.36745	$8.535.10^{-5}$
9	0.33350	$3.246.10^{-4}$	0.32870	$1.809.10^{-5}$	0.32978	$2.152.10^{-5}$	0.32861	$9.271.10^{-5}$
10	0.30046	$3.469.10^{-4}$	0.29530	$1.937.10^{-5}$	0.29655	$2.353.10^{-5}$	0.29528	$1.000.10^{-4}$

Table 3.1 Trajectory prediction of the mean μ_k of $p(x_k)$ describing the state of system (3.132).

of one random number is much more time consuming than additions and multiplications. Based on our computing experience with the present example we use $\eta_L = 0.8$ for all variance reduction techniques compared with the crude Monte Carlo method.

Time	Method B		Method C		Method D	
k	η_v	η	η_v	η	η_v	η
5	27.8	22.2	25.9	20.8	3.6	2.9
10	17.8	14.8	14.8	11.9	3.4	2.7

Table 3.2 Variance reductions η_v and efficiency gains η for predicting the mean $\hat{\mu}_k$ at time k=5 and k=10.

This table 3.2 shows that an arbitrary weighting sequence model (method D) gives little improvement whereas the antithetic variate method B gives a slightly better estimator than the control variate method C using statistical linearization. The variance reduction factors and hence the efficiency gains decrease as we increase the length of the prediction interval.

In fig. 3.5 we plot the ensemble values of the sampling errors, defined as $[\hat{var}(\hat{\mu}_k)]^{1/2}$, for the four listed Monte Carlo methods as function of the prediction time k. As expected, there is only a small difference between method B and method C. Method D decreases the sampling error compared with the crude Monte Carlo method only by a moderate amount.

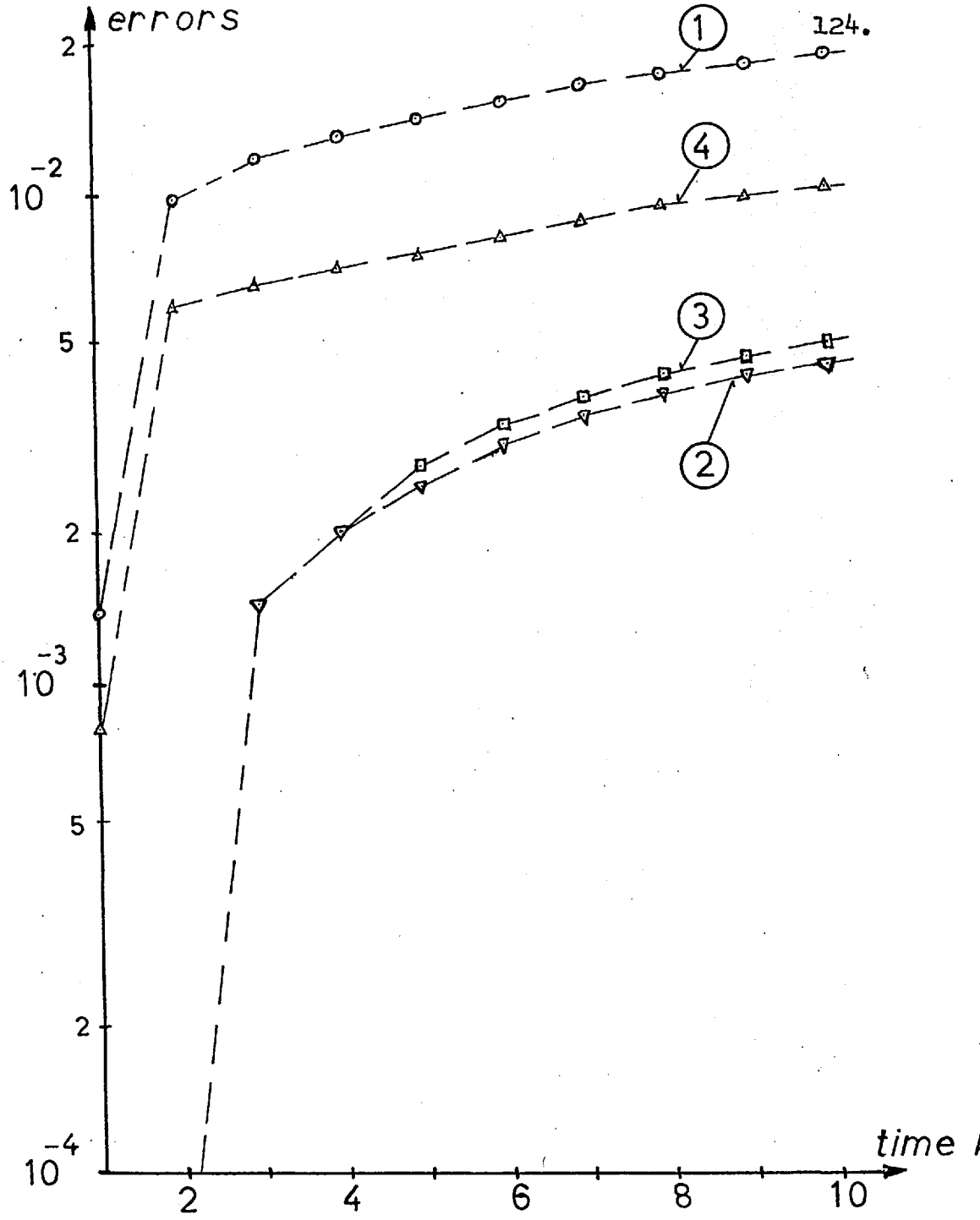

Fig. 3.5 Sampling errors $\overline{[\hat{var}(\hat{\mu}_k)]}^{\frac{1}{2}}$ of predicted mean $\hat{\mu}_k$ vs. time k.

1 crude Monte Carlo method A
2 antithetic variate method B
3 control variate method C
4 control variate method D

Table 3.3 exhibits the ensemble estimates $\overline{\hat{M}_k}$ of the second order moment M_k, defined by eqn.(3.28). The previously mentioned methods A, B and C are applied to find $\hat{M}_k$. Method D is not included because it gives unsatisfactory results. Indeed, the results of table 3.2 indicate that the use of an arbitrarily chosen control variate model does not necessarily improve the sampling accuracy. The estimated sampling variances of $\hat{M}_k$, $\hat{var}(\hat{M}_k)$, are only approximately valid and rely on N being large.

In table 3.4 we summarize some variance reduction factors and efficiency gains for methods B and C compared with A in connection with the second order moment M_k.

Time	Method B		Method C	
k	η_v	η	η_v	η
5	4.6	3.7	18.7	15.0
10	2.9	2.4	9.8	7.9

Table 3.4 Variance reduction η_v and efficiency gains η for predicting the second order moment $\hat{M}_k$ at k=5 and k=10.

In agreement with our previous remark in section 3.4.1, the antithetic variate method is not suitable for estimating the second order moments whereas the control variate method using statistical linearization yields significantly improved estimates for both the mean μ_k and the second order moment M_k over the prediction interval considered.

Time	Method A.		Method B		Method C	
k	$\bar{\hat{M}}_k$	$\overline{\hat{var}(\hat{M}_k)}$	$\bar{\hat{M}}_k$	$\overline{\hat{var}(\hat{M}_k)}$	$\bar{\hat{M}}_k$	$\overline{\hat{var}(\hat{M}_k)}$
1	1.00179	$8.323.10^{-6}$	1.00103	$1.323.10^{-8}$	1.00100	0.
2	0.69303	$1.715.10^{-4}$	0.68857	$9.558.10^{-6}$	0.68917	$8.921.10^{-9}$
3	0.52740	$1.738.10^{-4}$	0.52087	$1.889.10^{-5}$	0.52167	$3.455.10^{-6}$
4	0.43175	$1.903.10^{-4}$	0.42784	$3.473.10^{-5}$	0.42769	$5.518.10^{-6}$
5	0.37749	$1.936.10^{-4}$	0.36941	$4.208.10^{-5}$	0.36685	$8.034.10^{-5}$
6	0.33659	$2.018.10^{-4}$	0.33120	$5.081.10^{-5}$	0.32886	$1.015.10^{-5}$
7	0.31176	$2.163.10^{-4}$	0.30701	$6.373.10^{-5}$	0.30321	$1.510.10^{-5}$
8	0.29093	$2.438.10^{-4}$	0.28743	$7.295.10^{-5}$	0.28331	$1.830.10^{-5}$
9	0.27363	$2.539.10^{-4}$	0.27158	$7.767.10^{-5}$	0.26921	$2.167.10^{-5}$
10	0.26344	$2.665.10^{-4}$	0.26168	$9.055.10^{-5}$	0.26063	$2.409.10^{-5}$

Table 3.3 Trajectory prediction of the second order moment M_k of $p(x_k)$ describing the state of system (3.132).

To conclude the discussion of the trajectory prediction problem fig. 3.6 displays the variations with time of the ensemble estimates of the sampling errors of the second order moments, $\overline{[\hat{var}(\hat{M}_k)]}^{1/2}$ for methods A, B and C.

The control variate methods of sections 3.4.2 and 3 are particularly suitable for the state prediction problem. The control variate method using a linear weighting sequence model requires the estimation of the parameter vector $\underline{a}_k$ in eqn.(3.91). This can be done by a linear regression technique as described in section 3.4.2. An initial sample of size N_1 =200 is used to estimate $\hat{\underline{a}}_k^o$ with eqn.(3.95). This estimate is subsequently used in eqn.(3.97) to estimate $\hat{\mu}_k$ with a sample of size N_2=500. Table 3.5 contains the ensemble averages of this procedure for the mean $\overline{\hat{\mu}_k}$, predicted nine intervals ahead. The notation $\overline{\hat{\underline{a}}_k^o}$ indicates the ensemble averages over ten estimates $\hat{\underline{a}}_k^o$.

$\overline{\hat{\underline{a}}_k^o}$	0.176	0.263	0.365	0.436	0.500
	0.613	0.728	0.838	0.993	0.296
$\overline{\hat{\mu}_{10}} = 0.29560$			$\overline{\hat{var}(\hat{\mu}_{10})} = 2.321.10^{-5}$		$\eta_v = 10.6$

Table 3.5 Control variate method for state variable prediction using a linear regression technique.

The main advantage of the second order two-stage adaptive control variate method of section 3.4.3 has already been described as the fact that the matrix $F^{-1}_{a_k a_k}$, defined by eqn.(3.107), can be evaluated before

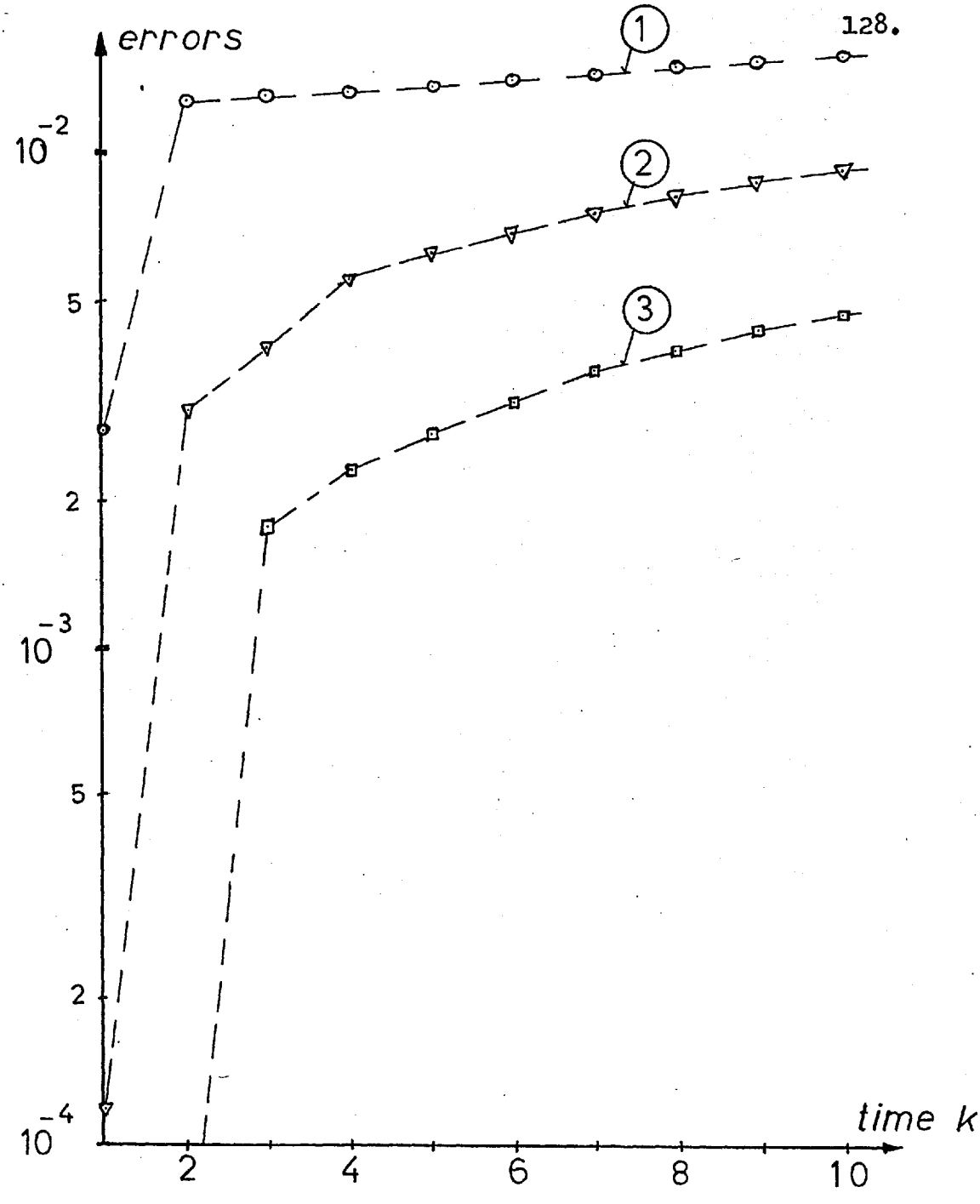

Fig. 3.6 Sampling errors $[\hat{\text{var}}(\hat{M}_k)]^{\frac{1}{2}}$ of predicted second order moment $\hat{M}_k$ vs. time k.

1 crude Monte Carlo method A

2 antithetic variate method B

3 control variate method C

the sampling procedure of stage one starts. The stochastic approximation algorithm (3.110) requires p iterations before the parameter estimate $\hat{\underline{\alpha}}_{k,p}$ can be assumed to be a good approximation to $\underline{\alpha}_k^o$. Table 3.6 shows the convergence of the first four components of the vector $\hat{\underline{\alpha}}_{k,p}$ for p=1,2,...,8. The time k is fixed as k=10. In each iteration p we estimate the gradient $\underline{F}_{\underline{\alpha}_{k,p}}$ by means of eqn.(3.109) using a sample of size N_1=50. Starting at p=1, the initial values of $\underline{\alpha}_k$ are chosen as $\underline{\alpha}_{k,1}$ = 8.0 and $\hat{\underline{\alpha}}_{k,p+1}$ is computed with eqn.(3.112).

Iterations p	$\hat{\alpha}_{k,p}(1)$	$\hat{\alpha}_{k,p}(2)$	$\hat{\alpha}_{k,p}(3)$	$\hat{\alpha}_{k,p}(4)$
1	8.0	8.0	8.0	8.0
2	0.076	0.232	0.419	0.407
3	0.119	0.203	0.350	0.409
4	0.143	0.218	0.364	0.397
5	0.139	0.225	0.346	0.426
6	0.145	0.215	0.350	0.423
7	0.140	0.218	0.355	0.424
8	0.143	0.221	0.365	0.427

Table 3.6 Stochastic approximation of the parameter $\hat{\underline{\alpha}}_{k,p}$.

The results of table 3.6 indicate that after three iterations (p=4) a sufficiently accurate estimate $\hat{\underline{\alpha}}_{k,p}$ of $\underline{\alpha}_k^o$ is obtained.

In the second stage of the control variate method, the parameter estimate $\hat{\underline{\alpha}}_{k,4}$ is kept constant to estimate the mean μ_{10}. With a random sample of size N_2=500 the evaluation of eqns.(3.113), (3.98)

and (2.71) yields the following ensemble averages for the mean $\hat{\mu}_{10}$ predicted nine intervals ahead:

$$\overline{\hat{\mu}_{10}} = 0.29585 \; ; \quad \overline{\hat{var}(\hat{\mu}_{10})} = 2.377.10^{-5} \; ; \quad \eta_v = 11.2 \qquad (3.136)$$

The comparison between the linear regression technique, see table 3.5, and the two-stage adaptive control variate method using a gradient technique shows that both methods yield nearly the same variance reduction η_v. Compared with the previously discussed antithetic variate method A and the control variate method B using statistical linearization the present variance reduction η_v of eqn. (3.136) is lower than in table 3.2. This may be explained by the lack of a zero sampling variance estimator for the mean $\hat{\mu}_k$ when either of the two methods discussed in sections 3.4.2 and 3 is applied to a linear Gaussian system.

3.5.2 Example II

Let us consider a three dimensional system. Using the notation of eqn.(3.63) we set

$$\underline{f}(\underline{x}_k,k) = \begin{bmatrix} x_k(1) - 0.5\, x_k(2)\, x_k(3) \\ x_k(2) + 0.3\, x_k(1)\, x_k(3) \\ x_k(3) - 0.1\, x_k(1)\, x_k(2) \end{bmatrix} . \qquad (3.137)$$

The arguments in brackets on the R.H.S. denote the components of the vector $\underline{x}_k$. The initial condition is assumed to be given as a constant

$$\underline{x}_1^T = [0.9 \quad 0.5 \quad 0.1] \quad . \tag{3.138}$$

The covariance matrix Σ_{w_k} of the zero mean additive noise vector $\underline{w}_k$ is

$$\Sigma_{w_k} = \begin{bmatrix} 0.05 & 0.005 & 0.0025 \\ 0.005 & 0.05 & 0.005 \\ 0.0025 & 0.005 & 0.05 \end{bmatrix} \quad . \tag{3.139}$$

In order to generate the three dimensional random sample $\{\underline{w}^{k-1}\}_j$ of size N with the specified covariance matrix Σ_{w_k}, see eqn.(3.139), the transformation matrix A of eqn.(2.21) is computed via eqn.(2.23) to be

$$A = \begin{bmatrix} 0.2236 & 0. & 0. \\ 0.0224 & 0.2225 & 0. \\ 0.0112 & 0.0213 & 0.2223 \end{bmatrix} \quad . \tag{3.140}$$

First, let us consider the antithetic variate method as described in section 3.3. In order to compare it with the crude Monte Carlo method we perform the following experiment. A random sample $\{\underline{w}^{k-1}\}_j^+$ of size N is used to generate $\{\underline{x}_k\}_j^+$ by simulating eqn.(3.63) and (3.137). In particular, we are interested in predicting the mean $\hat{\underline{\mu}}_k$ of the P.D.F. $p(\underline{x}_k)$ at times k=4 and k=6. In table 3.7 we compare the following methods:

Method A : Crude Monte Carlo technique. The sample mean $\hat{\underline{\mu}}_k$ is obtained from eqn.(3.30), its sampling variance is given by eqn.(3.35). Sample size N=100.

Method B : Antithetic variate method using 2N variates; see section 3.3.1. The antithetic random sample $\{\underline{x}_k\}_j^-$, is obtained by simulating eqn.(3.63) with $\{\underline{w}^{k-1}\}_j^-$, see eqn.(3.40). The sample mean $\hat{\mu}_k$ is obtained from eqn.(3.41); the sampling variance, $\text{var}(\hat{\mu}_k)$ is obtained from eqn.(3.43). Sample size N=100.

Method C : Generalized antithetic variate method; see section 3.3.3. Starting with a random sample $\{\underline{w}^{k-1}\}_{j,1}^+$ of size N=50 five additional samples $\{\underline{w}^{k-1}\}_{j,p}^+$, p=2, ...,6 are obtained by permuting the components of $(\underline{w}_k)_{j,1}^+$ as indicated by eqn.(3.52). For each sample $\{\underline{w}^{k-1}\}_{j,p}^+$ an antithetic sample $\{\underline{w}^{k-1}\}_{j,p}^-$ is obtained via eqn. (3.40). The sample mean $\hat{\hat{\mu}}_k$ is obtained from eqn.(3.60), the sampling variance, $\text{var}(\hat{\hat{\mu}}_k)$ from eqn.(3.62).

The results exhibited in table 3.7 are again averages over ten ensemble values.

The variance reduction η_v from method A to B can readily be computed from table 3.7 as the ratio of the two corresponding sampling variances. However, in order to obtain the correct variance reduction of method C we have to bear in mind that the generalized antithetic variate method C starts with a basic sample size N=50. Since it involves the computation of five additional samples the corresponding total size is N_c=300. Thus, the ratio of the sampling variances of method A and C has to be multiplied by 1/3. For the labour ratio, we recall that for the antithetic variate methods the system (3.63) has to be simulated twice as many times as in the crude Monte Carlo method.

Time		Method A		Method B		Method C	
k	i	$\bar{\hat{\mu}}_k(i)$	$\overline{\hat{var}(\hat{\mu}_k(i))}$	$\bar{\hat{\mu}}_k(i)$	$\overline{\hat{var}(\hat{\mu}_k(i))}$	$\bar{\hat{\mu}}_k(i)$	$\overline{\hat{var}(\hat{\mu}_k(i))}$
4	1	0.84953	$1.689 \cdot 10^{-3}$	0.84715	$4.641 \cdot 10^{-5}$	0.84919	$1.446 \cdot 10^{-5}$
	2	0.53618	$1.585 \cdot 10^{-3}$	0.54152	$1.618 \cdot 10^{-5}$	0.54058	$4.563 \cdot 10^{-6}$
	3	-0.03792	$1.443 \cdot 10^{-3}$	-0.03882	$1.972 \cdot 10^{-6}$	-0.03804	$5.740 \cdot 10^{-7}$
6	1	0.84410	$3.840 \cdot 10^{-3}$	0.82062	$5.449 \cdot 10^{-4}$	0.83744	$1.232 \cdot 10^{-4}$
	2	0.45902	$3.288 \cdot 10^{-3}$	0.47083	$1.818 \cdot 10^{-4}$	0.49413	$4.472 \cdot 10^{-4}$
	3	-0.12556	$2.544 \cdot 10^{-3}$	-0.12381	$2.937 \cdot 10^{-5}$	-0.12845	$5.072 \cdot 10^{-6}$

Table 3.7 State prediction of the mean $\underline{\mu}_k$ using a crude Monte Carlo method A and two antithetic variate methods B and C.

However, as the antithetic samples $\{\underline{w}^{k-1}\}_{j,p}^{-}$, for p=1 or p= 1,...,6, are generated by analytic methods it is realistic to work with a labour ratio $\eta_L = 0.8$.

In table 3.8 we compare the variance ratios η_v and efficiency gains η of methods B and C over A.

Next, we apply the control variate method using statistical linearization to solve the trajectory prediction problem. The mathematical relationships are derived in section 3.4.1. Due to the structure of $\underline{f}$ in eqn.(3.137) it is possible to compute the coefficients $\underline{a}_k$ and B_k of the linear model (3.72) analytically. They are given by eqn.

Time	Components	Method B		Method C	
k	i	η_v	η	η_v	η
4	1	36.4	29.1	39.0	31.2
	2	98.0	78.4	115.9	92.7
	3	730.2	586.1	837.1	669.2
6	1	7.0	5.6	10.4	8.3
	2	18.1	14.5	24.5	19.6
	3	84.1	67.3	168.1	133.9

Table 3.8 Variance ratios and efficiency gains of the antithetic variate methods.

(3.67) as

$$\begin{aligned} a_k(1) &= m_k(1) - 0.5\,[P_k(2,3) + m_k(2)\,m_k(3)] \\ a_k(2) &= m_k(2) + 0.3\,[P_k(1,3) + m_k(3)\,m_k(1)] \\ a_k(3) &= m_k(3) - 0.1\,[P_k(1,2) + m_k(1)\,m_k(2)] \end{aligned} \qquad (3.141)$$

where $\underline{m}_k$ and P_k are defined by eqns.(3.64) and (3.73).

Using eqn.(3.70) the matrix B_k is obtainable from eqn.(3.141) as

$$B_k = \begin{bmatrix} 1.0 & -0.5\,m_k(3) & -0.5\,m_k(2) \\ 0.3\,m_k(3) & 1.0 & 0.3\,m_k(1) \\ -0.1\,m_k(2) & -0.1\,m_k(1) & 1.0 \end{bmatrix}. \qquad (3.142)$$

The results shown in table 3.9 are also averages over ten ensemble values of $\hat{\mu}_k$ and $\hat{var}(\hat{\mu}_k)$. Each experiment is based on a random sample

Time	Compo-nents	Crude Monte Carlo Method		Control Variate Method Using Stat.Linearization	
k	i	$\hat{\mu}_k(i)$	$\widehat{var}(\hat{\mu}_k(i))$	$\hat{\mu}_k(i)$	$\widehat{var}(\hat{\mu}_k(i))$
2	1	0.87539	$5.017\ 10^{-5}$	0.87500	$1.967\ 10^{-9}$
	2	0.52081	$5.103\ 10^{-5}$	0.52699	$3.092\ 10^{-10}$
	3	0.05275	$4.831\ 10^{-5}$	0.05500	$7.334\ 10^{-11}$
3	1	0.85927	$1.037\ 10^{-4}$	0.85966	$6.223\ 10^{-7}$
	2	0.53388	$1.036\ 10^{-4}$	0.54152	$2.321\ 10^{-7}$
	3	0.01093	$9.594\ 10^{-5}$	0.00879	$2.518\ 10^{-8}$
4	1	0.84953	$1.689\ 10^{-4}$	0.85274	$4.243\ 10^{-6}$
	2	0.53618	$1.585\ 10^{-4}$	0.54009	$1.602\ 10^{-6}$
	3	-0.03792	$1.443\ 10^{-4}$	-0.03751	$1.888\ 10^{-7}$
5	1	0.84847	$2.594\ 10^{-4}$	0.85009	$1.619\ 10^{-5}$
	2	0.51035	$2.331\ 10^{-4}$	0.51804	$6.634\ 10^{-6}$
	3	-0.08179	$1.983\ 10^{-4}$	-0.08241	$8.249\ 10^{-7}$
6	1	0.84410	$3.840\ 10^{-4}$	0.84452	$4.772\ 10^{-5}$
	2	0.45902	$3.288\ 10^{-4}$	0.47093	$2.085\ 10^{-5}$
	3	-0.12556	$2.544\ 10^{-4}$	-0.12308	$2.971\ 10^{-6}$
7	1	0.82949	$5.362\ 10^{-4}$	0.82853	$8.325\ 10^{-5}$
	2	0.38449	$4.547\ 10^{-4}$	0.39627	$5.561\ 10^{-5}$
	3	-0.15966	$3.060\ 10^{-4}$	-0.15564	$8.918\ 10^{-6}$
8	1	0.79250	$7.282\ 10^{-4}$	0.79248	$1.205\ 10^{-4}$
	2	0.28324	$5.949\ 10^{-4}$	0.29452	$1.201\ 10^{-4}$
	3	-0.18027	$3.568\ 10^{-4}$	-0.17650	$2.285\ 10^{-5}$
9	1	0.73100	$9.691\ 10^{-4}$	0.73319	$1.692\ 10^{-4}$
	2	0.15309	$7.600\ 10^{-4}$	0.17050	$1.582\ 10^{-4}$
	3	-0.18509	$4.144\ 10^{-4}$	0.18462	$4.782\ 10^{-5}$
10	1	0.64686	$1.214\ 10^{-3}$	0.64842	$2.255\ 10^{-4}$
	2	0.01645	$9.485\ 10^{-4}$	0.03184	$1.926\ 10^{-4}$
	3	-0.17527	$4.816\ 10^{-4}$	-0.18129	$8.270\ 10^{-5}$

Table 3.9 Trajectory prediction of the mean $\underline{\mu}_k$ of $p(\underline{x}_k)$ describing the state of system (3.137).

$\{\underline{w}^{k-1}\}_j$ of size N = 1000. The same random sample $\{\underline{w}^{k-1}\}_j$ is used for the crude Monte Carlo method (eqns.(3.30) and (3.35)) and the control variate method (eqns.(3.76) and (3.77)). The mean μ_k is predicted up to time k=10.

The sampling variance ratios and the efficiency gains compared with the crude Monte Carlo predictor are shown in table 3.10 for k=4 and k=6. Again, a labour ratio $\eta_L = 0.8$ has been found adequate.

Time	Compo-nent	Control variate method using statistical linearization	
k	i	η_v	η
4	1	39.8	31.8
	2	99.1	79.4
	3	766.1	613.5
6	1	8.1	6.5
	2	15.7	12.5
	3	85.7	68.5

Table 3.10 Variance ratios η_v and efficiency gains η of the control variate method for estimating the mean $\hat{\mu}_k$.

It is interesting to compare these results with those of table 3.8. As in the scalar case there is little difference between the antithetic variate method and the control variate method but the generalized antithetic variate method gives small although consistently better estimates compared with the other two methods mentioned.

As the statistical linearization procedure provides an analytic approximation to the nonlinear prediction problem it is interesting to compare the approximation error $\underline{e}_k$, whose ensemble estimate is defined by

$$\bar{\hat{\underline{e}}}_k = \left| \bar{\hat{\underline{\mu}}}_k - \underline{m}_k \right| , \qquad (3.143)$$

with the sampling error $\left[\overline{\hat{var}(\hat{\mu}_k)}\right]^{1/2}$ of the Monte Carlo solutions. In fig. 3.7 we plot the error $\bar{\hat{e}}_k(1)$ together with the sampling errors $\left[\overline{\hat{var}(\hat{\mu}_k(1))}\right]^{1/2}$ of the crude estimator and the control variate method versus the time k. The sampling error of the control variate method is seen to be less than the approximation error estimate $\bar{\hat{e}}_k(1)$. But this of course is obtained only at the expense of generating the random sample of size N=1000 and simulating the nonlinear system together with the linear model.

Finally, we implement the two-stage control variate method discussed in section 3.4.4. Because the initial condition $\underline{x}_1$ of the system considered is a constant the last term in eqn.(3.114) is absent. In the first stage, starting at p=1, we set all elements of the (3,3)-dimensional matrices A_i equal to 0.4. The gradients $\underline{F}_{\underline{\alpha}_k i}$, (i=1,2,3), are estimated with eqn.(3.126) using a random sample of fixed size N_1=100. The auxiliary vectors $\hat{\underline{c}}^i_{k,p}$ are estimated with eqn.(3.127) and the combined estimates $\hat{\bar{\underline{c}}}^i_{k,p}$ are used to compute $\underline{\alpha}^i_{k,p+1}$ with eqn. (3.128). After three iterations (p=4) we terminate the first stage and use $\hat{\alpha}_{k,p}$ in the control variate model (3.118). The second stage of the control variate method is concerned with estimating $\hat{\underline{\mu}}_k$ using

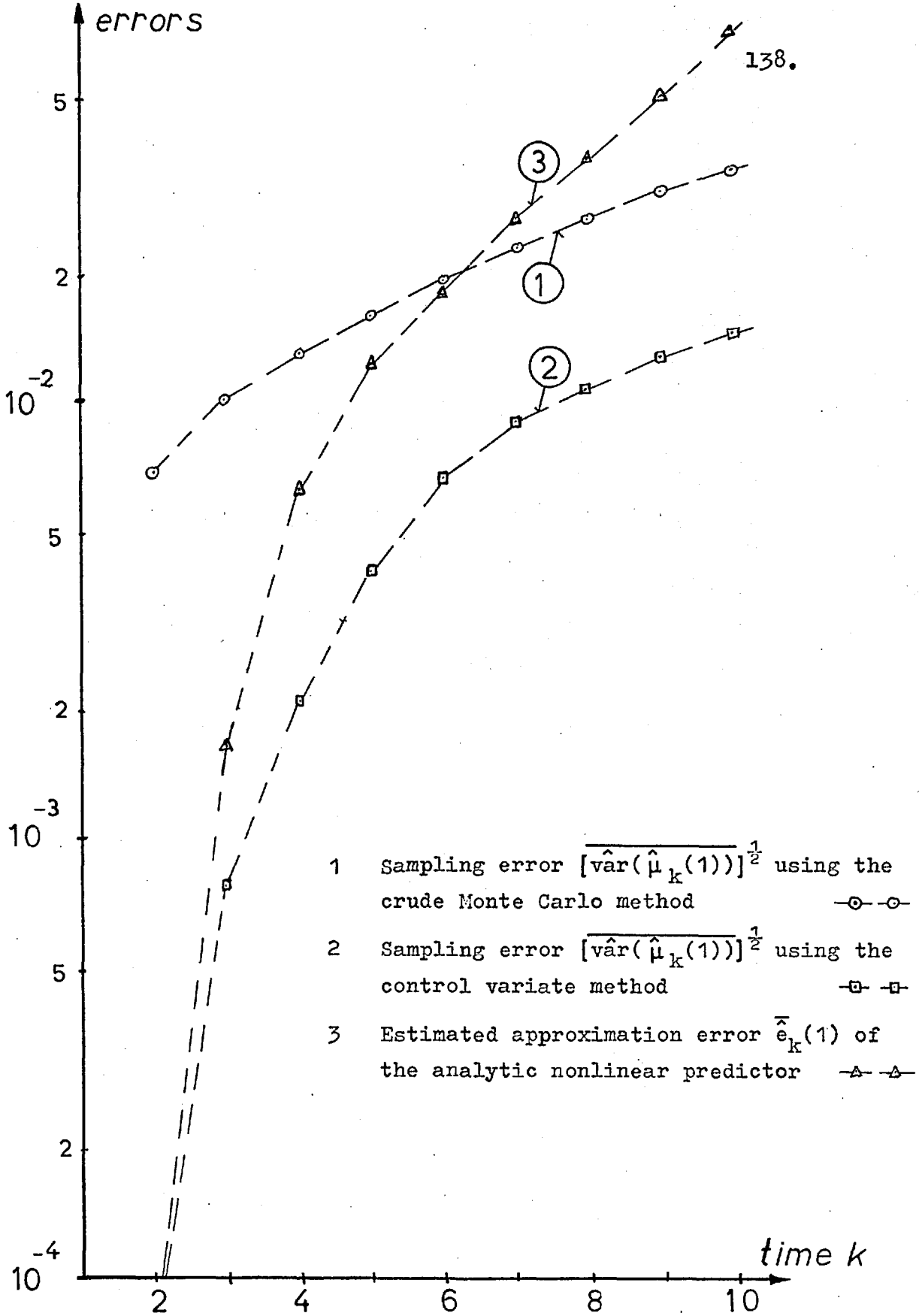

Fig.3.7 Prediction errors of the first component $\mu_k(1)$ of the mean vector $\underline{\mu}_k$ vs. time k.

eqn.(3.129) and $\hat{var}(\hat{\mu}_k)$ using eqn.(3.131) based on a sample of size N_2=1000. The results in tables 3.11 and 3.12 for the second stage are ensemble values averaged over ten realizations of the sampling experiment to estimate $\hat{\mu}_k$ and $\hat{var}(\hat{\mu}_k)$. Table 3.11 contains the estimates of μ_k and $var(\hat{\mu}_k)$ at time k=4 and table 3.12 those at k=6. In both tables we include a selection of parameter estimates of the matrix $\hat{a}_{k,p}$.

Stage	Iteration p	$\hat{A}_{1,p}(1,1)$	$\hat{A}_{2,p}(1,1)$	$\hat{A}_{3,p}(1,1)$
I	1	0.4	0.4	0.4
	2	1.0363	1.0132	0.9426
	3	1.0065	0.9995	1.0016
	4	1.0070	0.9999	0.9994
	Component i	$\bar{\hat{\mu}}_k(i)$	$\overline{\hat{var}(\hat{\mu}_k(i))}$	η_v
			10^{-6}	
II	1	0.85113	4.883	24.7
	2	0.54158	1.705	66.3
	3	-0.03872	0.431	238.1

Table 3.11 Two-stage predictor of μ_k at time k=4.

Stage	Iteration p	$\hat{A}_{1,p}(1,1)$	$\hat{A}_{2,p}(1,1)$	$\hat{A}_{3,p}(1,1)$	$\hat{A}_{4,p}(1,1)$	$\hat{A}_{5,p}(1,1)$
I	1	0.4	0.4	0.4	0.4	0.4
	2	1.0606	0.9925	0.9580	0.8144	0.9604
	3	1.0296	0.9855	0.9970	1.0210	0.9897
	4	1.0155	0.9943	0.9994	0.9921	0.9963

	Component i	$\overline{\hat{\mu}_k(i)}$	$\overline{\hat{var}(\hat{\mu}_k(i))}$	η_v
			10^{-5}	
II	1	0.84475	5.685	4.8
	2	0.47089	1.973	11.9
	3	-0.12451	0.417	43.7

Table 3.12 Two-stage predictor of μ_k at time k=6.

In order to compute the variance reductions η_v from the crude Monte Carlo method we have to take the different sample sizes into consideration. The present two-stage control variate method is based on a sample of total size $N_t=(4 N_1 + N_2) = 1400$ whereas the crude Monte Carlo results in table 3.9 are based on a sample of size N = 1000.

The present values of η_v are less favourable than the variance reductions shown in tables 3.8 and 3.10. However, the potentialities of the two-stage procedure cannot be entirely judged by the present example because its loss in accuracy has to be weighed against the more restrictive assumptions required for the antithetic variate method

and the control variate method using the statistical linearization procedure.

3.5.3 Summary of results

Several conclusions can be drawn from the numerical results of the preceding sections. Comparing the relative behaviour of the Monte Carlo procedures presented the following statements regarding the merits and defects of the different solutions can be made.

(1) The crude Monte Carlo predictor for an arbitrary moment of $p(\underline{x}_k)$ is very simple and not restricted by any further assumptions than those laid down in section 1.2. No difference arises whether the state or trajectory prediction problem is solved by the crude Monte Carlo method.

(2) The most serious deficiency of the crude Monte Carlo method is the lack of precision. Usually, the sample sizes required to reduce the sampling errors of the desired estimates are too high to provide conclusive results.

(3) A variety of variance reduction methods has been investigated. The antithetic variate method requires the additional condition that the P.D.F. from which samples are to be drawn must be symmetric ~~and unimodal~~. An interesting and new extension derived for the Gaussian multivariate case gives a variance reduction beyond that of the better known method, using only 2N variates.

(4) The combination of an analytic approximation with sampling techniques results in a new control variate method. The sampling

error can be reduced below the approximation error at the expense of generating a random sample and simulating both the original system and the linear model. The successful solution requires the special system formulation of eqn.(3.63) with additive Gaussian white noise.

(5) The adaptive two-stage control variate method is less restrictive as far as the basic assumptions are concerned. However, in the examples considered the sampling errors are always bigger than those of the antithetic variate method or the control variate method using the statistical linearization procedure.

CHAPTER FOUR

SINGLE-STAGE NONLINEAR FILTERING

4.1 Introduction

4.1.1 Problem statement

In this and the following chapter we consider the computation of conditional expectation parameter estimates for nonlinear stochastic processes. In this chapter attention is confined to the single-stage (memoryless) filtering problem. Since dynamic nonlinearities are not considered here, it will be sufficient to assume a static system. As the time is fixed we omit the index k in this chapter. Thus, the observation system (1.4) takes the form

$$\underline{y} = \underline{g}(\underline{x}, \underline{v}) \tag{4.1}$$

where in general the dimensionality of $\underline{y}$ will differ from that of $\underline{x}$. The problem to be solved is then to find the conditional mean $E[\underline{x}|\underline{y}]$ by evaluating the integral

$$E[\underline{x}|\underline{y}] = \int \underline{x}\, p(\underline{x}|\underline{y})\, d\underline{x} \tag{4.2}$$

for a given observation $\underline{y}$.

4.1.2 The posterior P.D.F. $p(\underline{x}|\underline{y})$

The evaluation of eqn.(4.2) requires the knowledge of the posterior

P.D.F. $p(\underline{x}|\underline{y})$. Assume the joint P.D.F. $p(\underline{x}, \underline{y})$ is given. This will be the case, for example, if $p(\underline{x})$ and $p(\underline{v})$ are given and $\underline{x}$ and $\underline{v}$ are independent. Then $p(\underline{y})$ and $p(\underline{y}|\underline{x})$ are obtainable from eqn. (4.1) and the posterior P.D.F. $p(\underline{x}|\underline{y})$ is given by Bayes' theorem as

$$p(\underline{x}|\underline{y}) = \frac{p(\underline{x})\ p(\underline{y}|\underline{x})}{p(\underline{y})} \quad . \tag{4.3}$$

The P.D.F. $p(\underline{y})$ in the denominator is a normalizing scalar constant defined by

$$p(\underline{y}) = \int p(\underline{x})\ p(\underline{y}|\underline{x})\ d\underline{x}. \tag{4.4}$$

The prior P.D.F. $p(\underline{x})$ is assumed to be known; e.g. Gaussian

$$p(\underline{x}) = n(\underline{x};\ \underline{m}_x,\ \Sigma_x). \tag{4.5}$$

The conditional P.D.F. $p(\underline{y}|\underline{x})$ is referred to as the likelihood function. If $p(\underline{v})$ is a Gaussian P.D.F. and, for the sake of simplicity, eqn.(4.1) is replaced by

$$\underline{y} = \underline{g}(\underline{x}) + \underline{v}, \tag{4.6}$$

then $p(\underline{y}|\underline{x})$ is given by

$$p(\underline{y}|\underline{x}) = \text{const. } \exp\left[-\frac{1}{2}\left([\underline{y}-\underline{g}(\underline{x})]^T\ \Sigma_v^{-1} [\underline{y}-\underline{g}(\underline{x})] \right)\right] . \tag{4.7}$$

If in eqn.(4.6) $\underline{v}$ is not an additive term the P.D.F. $p(\underline{y}|\underline{x})$ is obtained with a procedure similar to that discussed in section 3.1.2 using the Jacobian of $\underline{g}$.

Using eqn.(4.3) the conditional mean $E[\underline{x}|\underline{y}]$ of eqn.(4.2) can be

expressed as

$$E[\underline{x}|\underline{y}] = \frac{\int \underline{x}\, p(\underline{y}|\underline{x})\, p(\underline{x})d\underline{x}}{\int p(\underline{y}|\underline{x})\, p(\underline{x})d\underline{x}} \triangleq \frac{\underline{\theta}_n}{\theta_d} \cdot \qquad (4.8)$$

Unless the dimensionality of $\underline{x}$ is not more than two, it probably is not practical to attempt a direct numerical evaluation of $\underline{\theta}_n$ and θ_d because the number of points in a net used for straightforward numerical integration must grow exponentially with the dimensionality of $\underline{x}$. A less direct approach to the calculation of $\underline{\theta}_n$ and θ_d is therefore generally required.

4.1.3 Sampling techniques

Conditional expectation parameter estimates, as defined by eqns. (4.2) and (4.8), are known to be statistically optimal when the loss incurred due to incorrect estimation is a quadratic function of the estimation error (Kalman[4]). However, except in the case of a linear system with additive Gaussian noise, i.e.

$$\underline{y} = C\,\underline{x} + \underline{v}, \qquad (4.9)$$

the difficulty of obtaining such estimates has to a large extent prevented their use in either theoretical or practical estimation problems (Cox[20]). Ho and Lee[19] pinpoint the difficulties associated with the evaluation of the posterior P.D.F. $p(\underline{x}|\underline{y})$. It is the object of this chapter to introduce Monte Carlo methods in order to overcome the computational problems inherent in the evaluation of eqn.

(4.8). This approach differs entirely from the usual linearized perturbation analysis based on results derived in linear estimation theory. The Monte Carlo solutions offer two distinct advantages:

(1) It is possible to work with the nonlinear system directly.

(2) The error analysis is part of the sampling experiment.

Although there is no basic difference between the single-stage and the multi-stage filtering problems, the latter has the additional complication that the state $\underline{x}$ is changing from stage to stage according to the dynamic relationship (1.1). Multi-stage nonlinear filtering is discussed in chapter five. In order to avoid duplication we include only those results in this chapter which are specifically suited for the single-stage case.

In section 4.2 a crude Monte Carlo procedure is introduced to estimate $\underline{\theta}_n$ and θ_d as defined by eqn.(4.8). The resulting ratio estimator for the conditional mean $E[\underline{x}|\underline{y}]$ is discussed in section 4.2.2 and reference will be made in chapter five to this part. The discussion in section 4.3 is concerned with a variance reduction technique known as importance sampling. Finally, in section 4.4 we comment briefly on the difference between our Monte Carlo approach to the computation of conditional expectation parameter estimates and the conditional Monte Carlo scheme originally proposed by Trotter and Tukey[40].

4.2 The crude Monte Carlo estimator

4.2.1 Sampling from the prior density $p(\underline{x})$

As the evaluation of $\underline{\theta}_n$ and θ_d defined by eqn.(4.8) requires that the prior P.D.F. $p(\underline{x})$ be given and that the effects of measurement noise on the observation $\underline{y}$ be known precisely, we may interpret the integrals appearing in eqn.(4.8) as expectation operations such that

$$\underline{\theta}_n \triangleq E_y [\, \underline{x}\, p(\underline{y}|\underline{x}) \,] \tag{4.10}$$

and

$$\theta_d \triangleq E_y [\, p(\underline{y}|\underline{x}) \,] . \tag{4.11}$$

The expectation operations are w.r.t. $p(\underline{x})$ and the notation $E_y[\,\cdot\,]$ indicates that the observation $\underline{y}$ is kept constant. Referring to the theory contained in chapter 2, crude Monte Carlo estimates for $\underline{\theta}_n$ and θ_d are given by

$$\hat{\underline{\theta}}_n = N^{-1} \sum_{j}^{N} [\, \underline{x}_j\, p(\underline{y}|\underline{x}_j)] \tag{4.12}$$

and

$$\hat{\theta}_d = N^{-1} \sum_{j}^{N} p(\underline{y}|\underline{x}_j) . \tag{4.13}$$

By the strong law of large numbers, these estimates $\hat{\underline{\theta}}_n$ and $\hat{\theta}_d$ converge with probability one to $\underline{\theta}_n$ and θ_d respectively, provided the integrals (4.10) and (4.11) exist in the ordinary sense and the size N of the random sample $\{\underline{x}\}_j$ drawn from $p(\underline{x})$ tends to infinity. Therefore the ratio $\hat{\underline{\theta}}_n / \hat{\theta}_d$ tends with probability one to $E[\underline{x}|\underline{y}]$; that is

$$E[\underline{x}|\underline{y}] = \lim_{N\to\infty} \frac{\hat{\theta}_n}{\hat{\theta}_d} \quad . \tag{4.14}$$

An estimate of $E[\underline{x}|\underline{y}]$, denoted by $\hat{\underline{x}}_y$, can be obtained by limiting N to an acceptable finite value and is then given by

$$\hat{\underline{x}}_y = \frac{\hat{\theta}_n}{\hat{\theta}_d} \quad . \tag{4.15}$$

This result for $\hat{\underline{x}}_y$ implies that in the Monte Carlo approach the evaluation of $p(\underline{y}|\underline{x})$ over a finite net in a space of n-dimensions is replaced by the evaluation of $p(\underline{y}|\underline{x})$ on randomly selected points in the same space.

4.2.2 The ratio estimator

Because the numerator $\underline{\theta}_d$ and denominator θ_n have to be replaced by estimates, expression (4.15) for the conditional mean is a ratio of two random variables. An important problem is concerned with the effect of finite N on the estimate $\hat{\underline{x}}_y$. In this section we show that eqn.(4.15) is a first order approximation and that the bias resulting from this approximation is negligible compared with the sampling error of $\hat{\underline{x}}_y$ for large N.

Let us denote the r:th component of the n-dimensional vector $\underline{\theta}_n$ by $\theta_n(r)$. Expanding the ratio function (4.15) into a two-dimensional Taylor series around the respective means $\theta_n(r) = E_y[\hat{\theta}_n(r)]$ and

$\theta_d = E_y[\hat{\theta}_d]$ and neglecting terms of order higher than two yields

$$\hat{x}_y(r) = \frac{\theta_n(r)}{\theta_d} + \frac{1}{\theta_d}\left[\hat{\theta}_n(r) - \theta_n(r)\right] - \frac{\theta_n(r)}{\theta_d^2}\left[\hat{\theta}_d - \theta_d\right]$$

$$- \frac{1}{\theta_d^2}\left[\hat{\theta}_n(r) - \theta_n(r)\right]\left[\hat{\theta}_d - \theta_d\right] + \frac{\theta_n(r)}{\theta_d^3}\left[\hat{\theta}_d - \theta_d\right]^2 . \tag{4.16}$$

The given observation $\underline{y}$ is regarded as a constant. Truncating eqn. (4.16) after the first term yields eqn.(4.15).

Including terms in eqn.(4.16) up to second order and taking expectations on both sides with respect to $p(\underline{x})$ while $\underline{y}$ is held constant yields an approximation $\underline{x}_{bias}$ to the true bias. The r:th component of $\underline{x}_{bias}$ is defined by

$$x_{bias}(r) \triangleq E_y\left[\hat{x}_y(r)\right] - \frac{\theta_n(r)}{\theta_d}$$

$$= \frac{\theta_n(r)}{\theta_d}\left[\frac{\mathrm{var}(\hat{\theta}_d)}{\theta_d^2} - \frac{\mathrm{cov}(\hat{\theta}_n(r),\ \hat{\theta}_d)}{\theta_n(r)\ \theta_d}\right]. \tag{4.17}$$

The estimate $\hat{x}_{bias}(r)$ is obtained after replacing the true but unknown values $\mathrm{var}(\hat{\theta}_d)$ and $\mathrm{cov}(\hat{\theta}_n(r),\ \hat{\theta}_d)$ by their respective estimates.

The covariance matrix V_o associated with the random vector $\hat{\underline{x}}_y$ is defined by

$$V_o \triangleq E_y[(\hat{\underline{x}}_y - E[\underline{x}|\underline{y}])\ (\hat{\underline{x}}_y - E[\underline{x}|\underline{y}])^T] . \tag{4.18}$$

Including terms in eqn.(4.16) up to second order and taking the variance on both sides with respect to $p(\underline{x})$, keeping $\underline{y}$ constant, yields an approximation V to the sampling covariance matrix V_o. The (r,p):th

element of V is

$$V(r,p) = \frac{\theta_n(r)\theta_n(p)}{\theta_d^2} \left[\frac{\mathrm{var}(\hat{\theta}_d)}{\theta_d^2} + \frac{\mathrm{cov}(\hat{\theta}_n(r), \hat{\theta}_n(p))}{\theta_n(r)\,\theta_n(p)} \right.$$

$$\left. - \frac{\mathrm{cov}(\hat{\theta}_n(r), \hat{\theta}_d)}{\theta_n(r)\,\theta_d} - \frac{\mathrm{cov}(\hat{\theta}_n(p), \hat{\theta}_d)}{\theta_n(p)\,\theta_d} \right] . \qquad (4.19)$$

Since all the terms in eqn.(4.19) are unknown, they are replaced by their estimates to yield an estimate $\hat{V}$ of V; e.g. we use in eqn.(4.19)

$$\hat{\mathrm{var}}(\hat{\theta}_d) = N^{-2} \sum_{j}^{N} [\, p(\underline{y} \mid \underline{x}_j) - \hat{\theta}_d]^2 \qquad (4.20)$$

where the random sample $\{\underline{x}\}_j$ of size N is drawn from $p(\underline{x})$.

The following conclusions can be drawn from the analysis of the ratio estimator:

(1) The elements of the standard error matrix are computed from $\hat{V}$ as $[\hat{V}]^{1/2}$, where V is defined by eqn.(4.19).

(2) Since this error matrix determines the confidence region of the statistical estimate $\hat{\underline{x}}_y$, the accuracy of the Monte Carlo estimator (4.15) is proportional to $N^{-1/2}$.

(3) Equation (4.17) indicates that the estimator (4.15) is biased. However, this bias is negligible for large N since it is proportional to N^{-1} and therefore much smaller than the sampling error, which is proportional to $N^{-1/2}$.

(4) The bias term defined by eqn.(4.17) disappears if a sampling procedure is found for which $\mathrm{var}(\hat{\theta}_d) = 0$. That is, a variance reduction technique with zero sampling variance for the denominator θ_d yields

$\underline{x}_{bias} = 0$.

(5) In order to improve the accuracy of the Monte Carlo estimator (4.15), variance reduction techniques have to be found for the estimation of the numerator $\underline{\theta}_n$ and the denominator θ_d.

Finally, an interesting question is whether one should use the same random sample $\{\underline{x}\}_j$ for estimating $\underline{\theta}_n$ and θ_d, or two independent samples. In the former case, the two estimates will be correlated so that the last two terms of eqn.(4.19) are nonzero. It seems likely (at least if the posterior P.D.F. $p(\underline{x}|\underline{y})$ differs substantially from zero only in the neighbourhood of the posterior mean $E[\underline{x}|\underline{y}]$)that this covariance between $\hat{\underline{\theta}}_n$ and $\hat{\theta}_d$ leads to an improved sampling covariance matrix V.

4.3 Importance sampling

4.3.1 New estimators for $\underline{\theta}_n$ and θ_d

Importance sampling is mentioned in section 2.4 as one of the standard variance reduction techniques. The underlying concepts of this method are described by Kahn[26]. In our discussion we only deal with those aspects relevant to the problem at hand.

The crude Monte Carlo estimators (4.12) for $\hat{\underline{\theta}}_n$ and (4.13) for $\hat{\theta}_d$ may suffer from the deficiency that only a fraction of the realizations $\underline{x}_j$ belonging to the random sample $\{\underline{x}\}_j$ lies inside the region where the

functions $\underline{\Psi}_1(\underline{x})$ and $\Psi_2(\underline{x})$, defined by

$$\underline{\Psi}_1(\underline{x}) \triangleq \underline{x}\, p(\underline{y}|\underline{x}) \tag{4.21}$$

and

$$\Psi_2(\underline{x}) \triangleq p(\underline{y}|\underline{x}) , \tag{4.22}$$

are effectively different from zero. In other words, the terms that make large contributions to $\hat{\underline{\theta}}_n$ and $\hat{\theta}_d$ are encountered only rarely. This situation can be improved if we are able to choose judiciously a new P.D.F. from which the sample $\{\underline{x}\}_j$ is to be drawn. Indeed, if the numerator $\underline{\theta}_n$ of eqn.(4.8) is multiplied and divided by the function $h(\underline{x})$, $\underline{\theta}_n$ becomes

$$\underline{\theta}_n = \int \frac{\underline{x}\, p(\underline{y}|\underline{x})\, p(\underline{x})}{h(\underline{x})}\, h(\underline{x})\, d\underline{x} . \tag{4.23}$$

Similarly the denominator θ_d of eqn.(4.8) is rewritten with $h(\underline{x})$ as

$$\theta_d = \int \frac{p(\underline{y}|\underline{x})\ p(\underline{x})}{h(\underline{x})}\ h(\underline{x})\, d\underline{x} . \tag{4.24}$$

Thus, the integrands remain unchanged but the functions $\underline{\Psi}_1(\underline{x})$ and $\Psi_2(\underline{x})$, defined by eqns.(4.21) and (4.22), become

$$\underline{\Psi}_1'(\underline{x}) = \frac{\underline{x}\, p(\underline{y}|\underline{x})\, p(\underline{x})}{h(\underline{x})} \tag{4.25}$$

and

$$\Psi_2'(\underline{x}) = \frac{p(\underline{y}|\underline{x})\, p(\underline{x})}{h(\underline{x})} . \tag{4.26}$$

Thus, the crude Monte Carlo estimators (4.12) and (4.13) may be replaced by

$$\hat{\underline{\theta}}_n = N^{-1} \sum_j^N \frac{\underline{x}_j \; p(\underline{y}|\underline{x}_j) \; p(\underline{x}_j)}{h(\underline{x}_j)} \tag{4.27}$$

and

$$\hat{\theta}_d = N^{-1} \sum_j^N \frac{p(\underline{y}|\underline{x}_j) \; p(\underline{x}_j)}{h(\underline{x}_j)} \tag{4.28}$$

where the random sample $\{\underline{x}\}_j$ of size N is drawn from $h(\underline{x})$. Finally, the Monte Carlo estimate $\hat{\underline{x}}_y$ of the conditional mean $E[\underline{x}|\underline{y}]$ is obtained from eqns.(4.15), (4.27) and (4.28).

As $h(\underline{x})$ is assumed to be a P.D.F. it has to satisfy the condition

$$\int_{-\infty}^{+\infty} h(\underline{x}) \; d\underline{x} = 1 \; . \tag{4.29}$$

Moreover, $h(\underline{x})$ is subject to the constraints that it may not be equal to zero unless

$$\underline{x} = \underline{0} \quad \text{or} \quad p(\underline{y}|\underline{x}) = 0 \text{ or } p(\underline{x}) = 0. \tag{4.30}$$

In order to reduce the sampling variances $\text{var}(\hat{\underline{\theta}}_n)$ and $\text{var}(\hat{\theta}_d)$ the P.D.F. $h(\underline{x})$ has to be chosen such that the functions $\underline{\Psi}'_1(\underline{x})$ and $\Psi'_2(\underline{x})$ of eqns.(4.25) and (4.26) respectively stay almost constant for arbitrary values of $\underline{x}$.

It can be shown that there always exist an optimal function $h(\underline{x})$, denoted by $h^o(\underline{x})$, such that the sampling variance of the scalar estimate $\hat{\theta}_d$ is equal to zero. Indeed, it follows from the definition of $\text{var}(\hat{\theta}_d)$ as

$$\text{var}(\hat{\theta}_d) = N^{-1} \int [\frac{p(\underline{y}|\underline{x}) \; p(\underline{x})}{h(\underline{x})} - \theta_d]^2 \; h(\underline{x}) \; d\underline{x} \tag{4.31}$$

that $h^o(\underline{x})$ is given by

$$h^o(\underline{x}) = \alpha \; p(\underline{y}|\underline{x}) \; p(\underline{x}) \; . \tag{4.32}$$

As $h^o(\underline{x})$ has to satisfy eqn.(4.29) the constant parameter α is found from the definition of θ_d in eqn.(4.8) to be

$$\alpha = 1 / \theta_d \tag{4.33}$$

and hence

$$h^o(\underline{x}) = \frac{p(\underline{y}|\underline{x})\ p(\underline{x})}{\int p(\underline{y}|\underline{x})\ p(\underline{x})\ d\underline{x}} \quad . \tag{4.34}$$

The substitution of eqn.(4.34) in (4.31) shows the zero sampling variance property.

However, eqn.(4.33) indicates that for the determination of $h^o(\underline{x})$ one has to know θ_d, the quantity to be estimated. Hence, eqn.(4.34) merely proves the existence of $h^o(\underline{x})$. The significance of this result lies not in the possibility of actually constructing the best function $h^o(\underline{x})$ but in demonstrating that there is a very efficient sampling scheme within the infinite number of available techniques. A logical consequence of the above result for $h^o(\underline{x})$ is the multi-stage sampling procedure originally proposed by Kahn[37] in the context of importance sampling.

In section 4.3.2 we determine $h^o(\underline{x})$ for a linear Gaussian system and in section 4.3.3 we discuss a zero sampling variance estimator for $\hat{\theta}_n$ using $h^o(\underline{x})$. The derivation of an approximate sampling function $\hat{h}^o(\underline{x})$ for a nonlinear system of the form (4.6) is contained in section 4.3.4.

4.3.2 Determination of $h^o(\underline{x})$ for the linear case

Let us consider a linear observation system of the form

$$\underline{y} = \underline{c} + D\underline{x} + \underline{v} \tag{4.35}$$

where $\underline{c}$ denotes a constant, $\underline{x}$ the state and $\underline{v}$ the measurement noise. It follows from the relationship

$$p(\underline{y}|\underline{x})\ p(\underline{x}) = p(\underline{x}, \underline{y}) = p(\underline{x}|\underline{y})\ p(\underline{y}) \tag{4.36}$$

that the posterior P.D.F. $p(\underline{x}|\underline{y})$ is Gaussian if the observation system is linear, $p(\underline{x})$ being given by eqn.(4.5) and $p(\underline{y})$ by eqn.(1.6).

We have the following result:

> If the state $\underline{x}$ is observed through the linear system (4.35), the optimal P.D.F. $h^o(\underline{x})$ to estimate the denominator θ_d with zero sampling variance is given by
>
> $$h^o(\underline{x}) = n(\underline{x};\ \underline{\mu}\ ,\ \Sigma\ ) \tag{4.37}$$
>
> where
>
> $$\underline{\mu} = \Sigma[D^T \Sigma_v^{-1}\ (\underline{y} - \underline{c}) + \Sigma_x^{-1} \underline{m}_x] \tag{4.38}$$
>
> $$\Sigma = (D^T \Sigma_v^{-1}\ D + \Sigma_x^{-1})^{-1}\ . \tag{4.39}$$

The proof of this result starts from eqn.(4.36). As $p(\underline{y})$ is only a normalizing constant, the posterior P.D.F. is given by

$$p(\underline{x}|\underline{y}) = \text{const. } p(\underline{y}|\underline{x})\ p(\underline{x}) = \text{const. } \exp(-\frac{E}{2}) \tag{4.40}$$

where

$$E = (\underline{y}-\underline{c}-D\underline{x})^T\ \ \Sigma_v^{-1}\ (\underline{y}-\underline{c}-D\underline{x}) + (\underline{x} - \underline{m}_x)^T\ \ \Sigma_x^{-1}\ (\underline{x} - \underline{m}_x). \tag{4.41}$$

Completing the squares in $\underline{x}$ allows us to rewrite E as

$$E = (\underline{x} - K^{-1}\underline{a})^T K (\underline{x} - K^{-1}\underline{a}) \tag{4.42}$$

where

$$\underline{a} = D^T \Sigma_v^{-1} (\underline{y} - \underline{c}) + \Sigma_x^{-1} \underline{m}_x \tag{4.43}$$

$$K = D^T \Sigma_v^{-1} D + \Sigma_x^{-1} \quad . \tag{4.44}$$

Setting $h^o(\underline{x}) = p(\underline{x}|\underline{y})$, that is $\underline{\mu} = K^{-1}\underline{a}$ and $\Sigma = K^{-1}$, we obtain eqns.(4.37) to (4.39). Recalling that the estimator for θ_d of eqn. (4.24) is

$$\theta_d = E_y \left[\frac{p(\underline{y}|\underline{x})\ p(\underline{x})}{h(\underline{x})} \right] \tag{4.45}$$

where $E_y[\cdot]$ is w.r.t. $h(\underline{x})$ we see that $\hat{\theta}_d$ is constant if we replace $h(\underline{x})$ in eqn.(4.45) by $h^o(\underline{x})$ of eqn.(4.37) and hence $\text{var}(\hat{\theta}_d) = 0$.

4.3.3 Zero sampling variance estimator for $\underline{\theta}_n$

Maintaining the linearity and normality conditions of the previous section, we are now concerned with a zero sampling variance estimator for $\underline{\theta}_n$. We prove the following result:

> If the state $\underline{x}$ is observed through the linear system (4.35) the combination of importance sampling using $h^o(\underline{x})$ of eqn. (4.37) and the antithetic variate method using two random samples $\{\underline{x}\}_j^{\pm}$ of sizes N drawn from $h^o(\underline{x})$ yields a zero sampling variance estimator for $\underline{\theta}_n$.

To prove this result we rewrite estimator (4.23) for $\underline{\theta}_n$ as

$$\underline{\theta}_n = E_y\left[\underline{x}\,\frac{p(\underline{y}|\underline{x})\,p(\underline{x})}{h^o(\underline{x})}\right] \tag{4.46}$$

where $E_y[\cdot]$ is w.r.t. $h^o(\underline{x})$. It follows from the previous result that the function $\psi_2'(\underline{x})$ is constant if in eqn.(4.26) $h(\underline{x})$ is replaced by $h^o(\underline{x})$. Thus eqn.(4.46) becomes

$$\underline{\theta}_n = \text{const}\;\; E[\underline{x}] \tag{4.47}$$

where the expectation is w.r.t. $h^o(\underline{x})$. The sampling covariance matrix of the estimate $\hat{\underline{\theta}}_n$, when $\hat{\underline{\theta}}_n$ is given by

$$\hat{\underline{\theta}}_n = \text{const}\;\; (2N)^{-1} \sum_{j}^{N} [\underline{x}_j^+ + \underline{x}_j^-], \tag{4.48}$$

is zero if $\underline{x}_j^{\pm}$ denotes a pair of antithetic variates. The variate $\underline{x}_j^+$ is drawn from $h^o(\underline{x})$ and $\underline{x}_j^-$ is computed as

$$(\underline{x}_j^- - \underline{\mu}) = -(\underline{x}_j^+ - \underline{\mu}) \tag{4.49}$$

where $\underline{\mu}$ is given by eqn.(4.38). Using eqn.(4.49) in (4.48) yields $\underline{\theta}_n = \text{const}\;\; \underline{\mu} = \theta_d\, \underline{\mu}$.

The combination of this result with that of the previous section shows that the Monte Carlo estimator using importance sampling and the antithetic variate method together yields a zero sampling variance estimate $\hat{\underline{x}}_y$ for the conditional mean $E[\underline{x}|\underline{y}]$, if the system is linear and the random signals involved possess Gaussian P.D.F.

This result on its own is not important as there are exact solutions which do not require sampling techniques. It does, however, render plausible the argument that in almost linear situations the new Monte Carlo procedure may considerably improve the accuracy of the crude estimator.

4.3.4 Approximation of $h^o(\underline{x})$ for the nonlinear case

Retaining the assumption of normality for $p(\underline{x})$ and $p(\underline{y})$ we return to the nonlinear observation system (4.6). In order to determine a suitable sampling P.D.F. $h(\underline{x})$ which approximates $h^o(\underline{x})$, we may pursue one of the following methods:

(1) A linear model of the form (4.35) is obtained by statistical linearization of eqn.(4.6) w.r.t. $p(\underline{x})$. Setting $k = 1$, the coefficients $\underline{c}$ and D are obtained from eqn.(5.33). They are then used to determine $\underline{\mu}$ and Σ in eqn.(4.37) to approximate $h^o(\underline{x})$ by

$$h^o(\underline{x}) \doteq n(\underline{x};\ \underline{\mu}\ ,\ \Sigma\ ). \tag{4.50}$$

See eqn.(4.38) and (4.39) for the evaluation of $\underline{\mu}$ and Σ . This approach will be further developed in chapter five in the context of multi-stage nonlinear filtering using the control variate method.

(2) If the likelihood function $p(\underline{y}|\underline{x})$ does not differ too much from a Gaussian P.D.F., denoted by $p_a(\underline{y}|\underline{x})$, the covariance matrix of the latter, can be determined such that it minimizes the deviation of $p_a(\underline{y}|\underline{x})$ from $p(\underline{y}|\underline{x})$ in a least-squares sense. This approach is due to McGhee and Walford[50].

(3) Finally, the expansion of the exponent of $p(\underline{y}|\underline{x})$ up to second order terms yields a Gaussian approximation $p_a(\underline{y}|\underline{x})$ which is used to compute $p_a(\underline{x}|\underline{y})$. This approximate posterior P.D.F. $p_a(\underline{x}|\underline{y})$ is set equal to $h(\underline{x})$ and used to improve the estimators for $\underline{\theta}_n$ and θ_d. Here we follow this method for the case of a scalar nonlinear observation of the form (4.6). The multi-variable case is discussed in chapter five in

connection with the control variate method. It makes use of some results obtained by Sorenson[24].

The scalar likelihood function is obtained from eqn.(4.7) as

$$p(y|x) = \text{const.} \exp\left(-\frac{1}{2} \frac{[y - g(x)]^2}{\Sigma_v} \right). \qquad (4.51)$$

To approximate $p(x|y)$ by a Gaussian P.D.F. $p_a(x|y)$, we expand the exponent of eqn.(4.51) into a Taylor series around the mean m_x of $p(x)$. Neglecting terms of order higher than two yields:

$$[y-g(x)]^2 \doteq [y-g(m_x)]^2 - 2[y-g(m_x)]\, g'(m_x)\, (x-m_x) + [g'(m_x)^2 - [y-g(m_x)]\, g''(m_x)]\, (x-m_x)^2 \qquad (4.52)$$

where $g'(x)$ denotes the first and $g''(x)$ the second order derivative of $g(x)$ w.r.t. x. As the observation y is kept constant, an approximate $p_a(x|y)$ is obtained by using eqn.(4.52) in (4.51) and multiplying the resulting $p_a(y|x)$ by the Gaussian P.D.F. $p(x)$. This yields

$$p_a(x|y) = \text{const.} \exp(-\frac{E}{2}) \qquad (4.53)$$

where

$$E = a_1 - a_2 (x - m_x) + a_3 (x - m_x)^2 \qquad (4.54)$$

and

$$a_1 = [y - g(m_x)]^2 / \Sigma_v$$

$$a_2 = 2 [y-g(m_x)]\, g'(m_x) / \Sigma_v \qquad (4.55)$$

$$a_3 = [g'(m_x)^2 - [y - g(m_x)]\, g''(m_x)] / \Sigma_v + 1 / \Sigma_x.$$

Completing the squares in (4.54) yields

$$E = a_3 [x - (m_x + [a_2/2a_3])]^2 + \text{const.} \qquad (4.56)$$

Substituting eqn.(4.56) in (4.53) yields

$$p_a(x|y) = n(x;\ \mu\ ,\ \Sigma\ ) \tag{4.57}$$

where

$$\mu = m_x + \left(\frac{\Sigma}{\Sigma_v}\right)[y - g(m_x)]\ g'(m_x) \tag{4.58}$$

and

$$\Sigma^{-1} = 1/\Sigma_x + [\,g'(m_x)^2 - [y-g(m_x)]\ g''(m_x)]/\Sigma_v. \tag{4.59}$$

It is easily verified that eqns.(4.58) and (4.59) reduce to eqns.(4.38) and (4.39) if g(x) is a linear function of the form (4.35).

The approximate posterior P.D.F. $p_a(x|y)$ of eqn.(4.57) is used as sampling density h(x); that is

$$h(x) = n(x;\ \mu,\ \Sigma\ ) \tag{4.60}$$

where μ and Σ are given by eqns.(4.58) and (4.59). In order to obtain a feasible sampling procedure which makes use of eqn.(4.60) the expression (4.59) for the variance Σ must be positive and g(x) must be at least twice differentiable.

The sampling density h(x) given by eqn.(4.60) is used in eqn. (4.28) to compute $\hat{\theta}_d$. The antithetic variate method is used to estimate the numerator $\hat{\theta}_n$ as

$$\hat{\theta}_n = (2N)^{-1} \sum_j^N \left[\, x_j^+ \frac{p(y|x_j^+)p(x_j^+)}{h(x_j^+)} + x_j^- \frac{p(y|x_j^-)p(x_j^-)}{h(x_j^-)} \,\right] \tag{4.61}$$

where the random sample $\{x\}_j^+$ is drawn from h(x). The antithetic random variables x_j^- are found from x_j^+ as

$$(x_j^- - \mu) = -(x_j^+ - \mu) \tag{4.62}$$

where μ is given by eqn.(4.58).

4.3.5 Concluding remarks

The variance reduction technique known as importance sampling has been developed to improve the accuracy of the estimate $\hat{\underline{x}}_{\underline{y}}$ for the conditional mean $E[\underline{x}|\underline{y}]$. The central problem is to find a suitable sampling P.D.F. $h(\underline{x})$ as a close approximation to the optimal density $h^{o}(\underline{x})$ such that the sampling error of $\hat{\theta}_{d}$ based on eqn.(4.28) is less than the crude Monte Carlo sampling error. If this requirement is not met, importance sampling may deteriorate the performance of the estimator (4.28) compared with (4.13). Based on the results derived for the linear case, we propose that the use of antithetic variates together with $h(\underline{x})$ offers a feasible method for improving the accuracy of the estimator of $\hat{\underline{\theta}}_{n}$.

Although the importance sampling method discussed in this section seems to indicate a considerable variance reduction effect, there is one serious difficulty which usually limits the applicability of this concept to rather special cases. If the product $p(\underline{y}|\underline{x})\times p(\underline{x})$ does not tend to zero at the same rate as $h(\underline{x})$ the resulting ratio $\psi_{2}'(\underline{x})$ of eqn. (4.26) may exhibit unbounded variations outside a region of possibly good fit. These variations may reduce the variance reduction effect considerably. Indeed, it is possible that such variations can increase the sampling error. This point will be illustrated by means of some examples in chapter six.

This deficiency is overcome by the control variate method. As the major part of the next chapter will be devoted to this technique in connection with the multi-stage filtering problem, we shall leave till chapter five the derivation of the control variate estimator for the

single-stage case.

4.4 Conditional Monte Carlo

The object of this section is to comment on conditional Monte Carlo, originally invented by Trotter and Tukey[40], and contrast it with the Monte Carlo approach for the evaluation of conditional expectation parameter estimates. An analytical explanation of conditional Monte Carlo is given by Hammersley[51] whereas Wendel[52] discusses the same technique from a group theoretic aspect. Apart from two specific examples contained in these papers there are no other contributions in this field. It is worthwhile to compare our own work on conditional P.D.F. with those original techniques in order to obtain a better understanding for the similarities of the two approaches.

The basic concept of the original conditional Monte Carlo work has its roots in importance sampling. Indeed, the integral

$$\theta = \int f(\underline{x})\, p(\underline{x})\, d\underline{x} \tag{4.63}$$

can be rewritten as

$$\theta = E_p[f(\underline{x})] = E_q[f(\underline{x})\, w(\underline{x})] \quad . \tag{4.64}$$

The first expectation is with respect to $p(\underline{x})$ whereas the second is with respect to $q(\underline{x})$. This latter expression requires an additional weighting function $w(\underline{x}) = p(\underline{x})/q(\underline{x})$ in order to compensate for the different P.D.F. $q(\underline{x})$. The essential point of eqn.(4.64) is that any sample can

come from any P.D.F. if we use the correct weight $w(\underline{x})$.

Conditional Monte Carlo differs from importance sampling in the sense that a P.D.F. $q(\underline{z})$ is defined in the product space $Z = X \times Y$ and not (as in importance sampling) in the same space X in which $p(\underline{x})$ is defined. The auxiliary space Y is to be chosen such that $p(\underline{x})$ is the conditional P.D.F. of the joint density $q(\underline{x}, \underline{y})$. If there exist a (1,1)-transformation between Z and X,Y of the form

$$\underline{x} = t(\underline{z}) \quad \text{and} \quad \underline{y} = u(\underline{z}) , \tag{4.65}$$

we have

$$p(\underline{x}) = p(\underline{x}|\underline{y}) = \text{const.}\ q(\underline{x},\underline{y}). \tag{4.66}$$

As it may be difficult to draw samples from $p(\underline{x})$, eqn.(4.66) allows us to replace $p(\underline{x})$ by $q(\underline{x},\underline{y})$. The samples are thus drawn from the Z space.

The main difference between state variable estimation using Bayes' theorem and conditional Monte Carlo is the fact that the observation $\underline{y}$, is part of the problem formulation itself and is not introduced to improve the sampling procedure.

We may cast the two approaches in a unified problem formulation as follows. It is desired to estimate the conditional expectation $E[\underline{f}(\underline{x})|\underline{y}]$. In particular, we set $\underline{f}(\underline{x}) = \underline{x}$ if we want to estimate the conditional mean $E[\underline{x}|\underline{y}]$. Using Bayes' theorem (4.3) we can avoid sampling from the posterior P.D.F. $p(\underline{x}|\underline{y})$ required to find the Monte Carlo estimate $\hat{\underline{x}}_y$ of $E[\underline{x}|\underline{y}]$. After introducing a suitably modified weighting function $w(\underline{x})$, it is possible to draw the random sample $\{\underline{x}\}_j$ from the unconditional P.D.F. $p(\underline{x})$. That is, the L.H.S. in eqn.(4.67)

is replaced by the R.H.S. for evaluating the conditional expectation

$$E_q [\underline{f}(\underline{x})|\underline{y}] = E_p [\underline{f}(\underline{x})\, w(\underline{x})] ; \qquad (4.67)$$

while the P.D.F. $q(\underline{x}, \underline{y})$ is defined in the Z space the R.H.S. of eqn. (4.72) involves the X space only.

On the other hand, the conditional Monte Carlo proposed by Trotter and Tukey[40] starts with an unconditional expectation operation which is replaced by a conditional expectation. In other words, the L.H.S. of eqn.(4.68) is sought to be replaced by the R.H.S. as

$$E_p[\underline{f}(\underline{x})] = E_q[\underline{f}(\underline{x})\, w(\underline{x})|\underline{y}] . \qquad (4.68)$$

The main difficulty for the implementation of conditional Monte Carlo is the definition of a suitable observation $\underline{y}$ in order to define the new space Z. It is not apparent whether this is in general possible. It is felt that for this reason the work of conditional Monte Carlo has not been pursued very much since its invention in 1956.

In conclusion, our approach of estimation of parameters belonging to conditional P.D.F. could be termed as the inverse problem of conditional Monte Carlo. As the observation $\underline{y}$ is part of the problem formulation, a Monte Carlo procedure can always be found such that sampling from conditional P.D.F. can be replaced by sampling from unconditional P.D.F.

CHAPTER FIVE

MULTI-STAGE NONLINEAR FILTERING

5.1 Introduction

In this chapter we are concerned with the statistical estimation of the state variables of a noisy nonlinear dynamic system. The mathematical description of the process to be considered is explained in section 1.2. For ease of reference we recall that the dynamics are given by the vector difference equation

$$\underline{x}_{k+1} = \underline{f}(\underline{x}_k, \underline{w}_k, k). \tag{5.1}$$

The states $\underline{x}_k$ are observed through the nonlinear transformation

$$\underline{y}_k = \underline{g}(\underline{x}_k, \underline{v}_k, k). \tag{5.2}$$

The statistical properties of the initial condition $\underline{x}_1$ and the random signals $\underline{w}_k$ and $\underline{v}_k$ are assumed to be given. Having observed a sequence $\underline{y}^k$, defined by

$$\underline{y}^k \triangleq \underline{y}_1, \underline{y}_2, \cdots, \underline{y}_k, \tag{5.3}$$

one may seek an estimate of an entire sequence of states $\underline{x}^{k+m}$. This formulation includes as special cases the filtering problem where an estimate of the current state $\underline{x}_k$ is sought, the smoothing problem where an estimate of the sequence $\underline{x}^k$ is required and the prediction problem where an estimate of a future state $\underline{x}_{k+m}$ is desired.

In this chapter our interest is confined to the filtering problem .

The smoothing problem is not discussed in this thesis. The prediction problem has been discussed in chapter three under the assumption that all states $\underline{x}_k$ are accessible. If previous observations $\underline{y}^k$ have to be taken into consideration, the initial condition P.D.F. $p(\underline{x}_1)$ which is used in chapter three has to be interpreted as $p(\underline{x}_k|\underline{y}^k)$. The solutions developed in chapter three are still valid but the initial random sample $\{\underline{x}_1\}_j$, originally drawn from $p(\underline{x}_1)$, has now to be drawn from $p(\underline{x}_k|\underline{y}^k)$.

In the present formulation one can easily include as state variables unknown parameters, either constant or randomly varying. The estimation of such parameters constitutes a form of the identification problem. In one of the examples presented in chapter six this aspect will be further discussed.

Bayes' theorem gives a recursive equation for the posterior P.D.F. $p(\underline{x}_k|\underline{y}^k)$ which plays a central role for the solution of the minimum mean-square estimation problem (section 5.2). For a successful application of a Monte Carlo method the recursive structure of Bayes' theorem has to be replaced by an expression which only contains prior information about the system. The crude Monte Carlo filtering procedure is contained in section 5.3 and makes extensive use of the results obtained in section 4.2.2. Variance reduction techniques are introduced in sections 5.4 and 5. In the former a statistical linearization procedure is found to be very suitable for the control variate method. Approximate nonlinear filter equations for the conditional mean and covariance matrix are used in section 5.5 to improve the crude

Monte Carlo procedure as well as to improve the accuracy of the analytic approximation.

5.2 The Bayesian approach

5.2.1 The minimum mean-square estimation problem

In the Bayesian approach to the estimation problem for stochastic systems, one is concerned first of all with the posterior P.D.F. $p(\underline{x}_k | \underline{y}^k)$. The mathematical foundation for this approach is given by the following lemma, due to Doob[44]:

Suppose that a random variable $\underline{x}_k$ is to be estimated from the known variables $\underline{y}^k$ when $\underline{x}_k$ and $\underline{y}^k$ have the joint P.D.F. $p(\underline{x}_k, \underline{y}^k)$. Suppose the estimate $\underline{x}_{k|k}$ is to be chosen as a function of $\underline{y}^k$ so that the criterion C, defined by

$$C \triangleq E\left[(\underline{x}_{k|k} - \underline{x}_k)^T (\underline{x}_{k|k} - \underline{x}_k)\right] \tag{5.4}$$

$$= \int (\underline{x}_{k|k} - \underline{x}_k)^T (\underline{x}_{k|k} - \underline{x}_k) p(\underline{x}_k, \underline{y}^k) d\underline{x}_k \, d\underline{y}^k$$

is minimum. Then the mean square estimate of $\underline{x}_{k|k}$ is

$$\underline{x}_{k|k} = E\left[\underline{x}_k | \underline{y}^k\right] ; \tag{5.5}$$

that is, the mean of the posterior P.D.F. $p(\underline{x}_k | \underline{y}^k)$. Other cost functions than (5.4) have been suggested by Ho and Lee[19]. These lead to estimates not the same as eqn.(5.5) but the posterior P.D.F. $p(\underline{x}_k | \underline{y}^k)$ still provides all the information required to solve the filtering problem.

5.2.2 The posterior conditional P.D.F. $p(\underline{x}_k|\underline{y}^k)$

For the system (5.1) and (5.2), the posterior P.D.F. $p(\underline{x}_k|\underline{y}^k)$ evolves according to Bayes' theorem[19,23] as

$$p(\underline{x}_k|\underline{y}^k) = \frac{\int p(\underline{x}_{k-1}|\underline{y}^{k-1})p(\underline{x}_k|\underline{x}_{k-1})p(\underline{y}_k|\underline{x}_k)\, d\underline{x}_{k-1}}{p(\underline{y}_k|\underline{y}^{k-1})} \qquad (5.6)$$

where the denominator $p(\underline{y}_k|\underline{y}^{k-1})$ is a scaling factor defined by

$$p(\underline{y}_k|\underline{y}^{k-1}) = \int\int p(\underline{x}_{k-1}|\underline{y}^{k-1})p(\underline{x}_k|\underline{x}_{k-1})p(\underline{y}_k|\underline{x}_k)d\underline{x}_k d\underline{x}_{k-1}\,.$$

Equation (5.6) is only true if $\underline{w}_k$ and $\underline{v}_k$ are statistically independent. The initial condition of this integral recurrence relation is given by eqn.(4.3) when $\underline{x}$ and $\underline{y}$ are interpreted as $\underline{x}_1$ and $\underline{y}_1$. Because of the nonlinearities involved, the task of determining the mean of eqn. (5.6) is overwhelming. We therefore propose a statistical solution to estimate the conditional mean rather than approximate it with an analytic nonlinear filter equation.

To this end, the recursive structure of eqn.(5.6) has to be replaced such that a new expression for $p(\underline{x}_k|\underline{y}^k)$ is obtained which only makes use of prior information about the system. For the sake of simplicity, we assume additive Gaussian noise in eqn.(5.1) and (5.2). That is, we consider the system

$$\underline{x}_{k+1} = \underline{f}(\underline{x}_k, k) + \underline{w}_k \qquad (5.7)$$

$$\underline{y}_k = \underline{g}(\underline{x}_k, k) + \underline{v}_k \qquad (5.8)$$

in order to simplify the P.D.F. $p(\underline{x}_k|\underline{x}_{k-1})$ and $p(\underline{y}_k|\underline{x}_k)$. If non-additive noise were assumed it would be necessary to introduce the

Jacobians of $\underline{f}$ and $\underline{g}$ with the concomitant complications mentioned in section 3.1.2.

The repeated substitution of $p(\underline{x}_{k-i}|\underline{y}^{k-i})$ in eqn.(5.6) for $i=1,2,\ldots k-2$ and the use of eqn.(4.3) for $k=1$, together with the Markov property

$$\prod_{i=0}^{k-2} p(\underline{x}_{k-i}|\underline{x}_{k-1-i})p(\underline{x}_1) = p(\underline{x}_1, \ldots, \underline{x}_k), \tag{5.9}$$

allows us to rewrite eqn.(5.6) as

$$p(\underline{x}_k|\underline{y}^k) = \frac{\int\cdots\int \prod_{i=1}^{k} p(\underline{y}_i|\underline{x}_i)p(\underline{x}^k)d\underline{x}^{k-1}}{\int\cdots\int \prod_{i=1}^{k} p(\underline{y}_i|\underline{x}_i)p(\underline{x}^k)d\underline{x}^k}. \tag{5.10}$$

The parameter we are interested in is, from eqn.(5.5), the conditional mean $E[\underline{x}_k|\underline{y}^k]$. It is defined by

$$E[\underline{x}_k|\underline{y}^k] \triangleq \frac{\int\cdots\int \underline{x}_k \prod_{i=1}^{k} p(\underline{y}_i|\underline{x}_i)p(\underline{x}^k)d\underline{x}^k}{\int\cdots\int \prod_{i=1}^{k} p(\underline{y}_i|\underline{x}_i)p(\underline{x}^k)d\underline{x}^k} \triangleq \frac{\theta_{n,k}}{\theta_{d,k}}. \tag{5.11}$$

Compared with eqn.(5.6) the new relationship (5.11) has the advantage of being an expression which only contains prior information about the system. Since it no longer exhibits a recursive character, it is directly amenable to Monte Carlo integration.

5.3 A crude Monte Carlo filtering procedure

Equation (5.11) shows that the numerator $\underline{\theta}_{n,k}$ and the denominator $\theta_{d,k}$ are given as (nk)-fold integrals. Using the definition of the expectation operator, we can interpret $\underline{\theta}_{n,k}$ and $\theta_{d,k}$ as

$$\underline{\theta}_{n,k} = E_y \left[\underline{x}_k \prod_{i=1}^{k} p(\underline{y}_i | \underline{x}_i) \right] \tag{5.12}$$

and

$$\theta_{d,k} = E_y \left[\prod_{i=1}^{k} p(\underline{y}_i | \underline{x}_i) \right] . \tag{5.13}$$

The expectation $E_y[\cdot]$ is performed w.r.t. the joint P.D.F. $p(\underline{x}^k)$ and the observations $\underline{y}^k$ are assumed to be fixed.

Referring to the results obtained in section 2.2.2, the random vector

$$\hat{\underline{\theta}}_{n,k} = N^{-1} \sum_{j}^{N} \left[(\underline{x}_k)_j \prod_{i=1}^{k} p(\underline{y}_i | (\underline{x}_i)_j) \right] \tag{5.14}$$

converges, under fairly general conditions, with probability one to $\underline{\theta}_{n,k}$ as the sample size N tends to infinity. Similarly, the random variable

$$\hat{\theta}_{d,k} = N^{-1} \sum_{j}^{N} \left[\prod_{i=1}^{k} p(\underline{y}_i | (\underline{x}_i)_j) \right] \tag{5.15}$$

converges with probability one to $\theta_{d,k}$ as $N \to \infty$.

The evaluation of eqn.(5.14) and (5.15) involves the likelihood functions $p(\underline{y}_i | \underline{x}_i)$, i=1,2,...,k. Restricting our discussion to eqn. (5.8) and using the statistical description of eqn.(1.6) for the observation noise $\underline{v}_k$, we can rewrite the estimators (5.14) and (5.15) as

$$\hat{\underline{\theta}}_{n,k} = \text{const.}\ N^{-1} \sum_{j}^{N} \left[(\underline{x}_k)_j \exp\left(-\frac{1}{2} \sum_{i}^{k} \left[\underline{y}_i - \underline{g}((\underline{x}_i)_j, i) \right]^T \Sigma_{v_i}^{-1} \left[\underline{y}_i - \underline{g}((\underline{x}_i)_j, i) \right] \right) \right] \tag{5.16}$$

and

$$\hat{\theta}_{d,k} = \text{const.}\ N^{-1} \sum_{j}^{N} \left[\exp\left(-\frac{1}{2} \sum_{i}^{k} \left[\underline{y}_i - \underline{g}((\underline{x}_i)_j, i) \right]^T \Sigma_{v_i}^{-1} \left[\underline{y}_i - \underline{g}((\underline{x}_i)_j, i) \right] \right) \right] \tag{5.17}$$

The random sequence $(\underline{x}^k)_j$ denotes the values $\underline{x}^k$ obtained by simulation of the nonlinear system (5.1) or (5.7). The initial condition $(\underline{x}_1)_j$ is drawn from the known $p(\underline{x}_1)$. The random noise sequence $(\underline{w}^{k-1})_j$ is drawn from $p(\underline{w}_k)$, see eqn.(1.3). Finally, the trajectory $(\underline{x}^k)_j$ is obtained as the solution of eqn.(5.1) or (5.7). This procedure is repeated N times to generate the random sample $\{\underline{x}^k\}_j$ of size N required to evaluate either eqns.(5.14), (5.15) or eqns.(5.16), (5.17). It corresponds to drawing the random sample $\{\underline{x}^k\}_j$ from $p(\underline{x}^k)$.

The Monte Carlo estimate of the conditional mean $E[\underline{x}_k | \underline{y}^k]$, denoted $\hat{\underline{x}}_{k|k}$, that will be used is

$$\hat{\underline{x}}_{k|k} = \frac{\hat{\underline{\theta}}_{n,k}}{\hat{\theta}_{d,k}} . \tag{5.18}$$

This ratio estimator has been discussed in section 4.2.2. The main result was that eqn.(5.18) is a first order approximation. The resulting bias is negligible compared with the sampling error as $N \to \infty$.

The covariance matrix $V_{o,k}$, associated with the random vector

$\hat{\underline{x}}_{k|k}$, is defined by

$$V_{o,k} \triangleq E\left[(\hat{\underline{x}}_{k|k} - E[\underline{x}_k|\underline{y}^k])\,(\hat{\underline{x}}_{k|k} - E[\underline{x}_k|\underline{y}^k])^T\right]. \qquad (5.19)$$

An approximation V_k to the sampling covariance matrix (5.19) is found by expanding eqn.(5.18) into a Taylor series up to second order terms and taking the variance on both sides with respect to $p(\underline{x}^k)$, keeping $\underline{y}^k$ constant. This yields for the (r,p):th element of V_k

$$V_k(r,p) = \frac{\theta_{n,k}(r)\theta_{n,k}(p)}{\theta^2_{d,k}}\left[\frac{\mathrm{var}(\hat{\theta}_{d,k})}{\theta^2_{d,k}} + \frac{\mathrm{cov}(\hat{\theta}_{n,k}(r),\hat{\theta}_{n,k}(p))}{\theta_{n,k}(r)\,\theta_{n,k}(p)} - \frac{\mathrm{cov}(\hat{\theta}_{n,k}(r),\hat{\theta}_{d,k})}{\theta_{n,k}(r)\theta_{d,k}} - \frac{\mathrm{cov}(\hat{\theta}_{n,k}(p),\hat{\theta}_{d,k})}{\theta_{n,k}(p)\theta_{d,k}}\right]. \qquad (5.20)$$

Since all the terms in eqn.(5.20) are unknown they are replaced by their estimates to yield an estimate $\hat{V}_k$ of V_k; e.g. we use in eqn.(5.20)

$$\hat{\mathrm{var}}(\hat{\theta}_{d,k}) = N^{-2}\sum_j^N\left[\left\{\prod_{i=1}^k p(\underline{y}_i|(\underline{x}_i)_j)\right\} - \hat{\theta}_{d,k}\right]^2. \qquad (5.21)$$

The method of sample moments, introduced in section 3.2.1, can be applied by extending the result (5.18) for the first order conditional moment $E[\underline{x}_k|\underline{y}^k]$. Indeed, the matrix of conditional second order sample moments $\hat{M}_{k|k}$ is obtained again as a ratio estimator of the form (5.18) where the numerator $\theta_{n,k}$ of eqn.(5.11) has to be replaced by the matrix $\theta^{(2)}_{n,k} \triangleq E_y\left[\underline{x}_k\underline{x}_k^T\prod_{i=1}^k p(\underline{y}_i|\underline{x}_i)\right]$. It is estimated by

$$\hat{\theta}^{(2)}_{n,k} = N^{-1}\sum_j^N\left[(\underline{x}_k)_j\,(\underline{x}_k)_j^T\prod_{i=1}^k p(\underline{y}_i|(\underline{x}_i)_j)\right]. \qquad (5.22)$$

This result follows directly from the definition of the second order conditional moment of the posterior P.D.F. $p(\underline{x}_k|\underline{Y}^k)$ as $M_{k|k} \triangleq E[\underline{x}_k \underline{x}_k^T|\underline{Y}^k]$. The normalizing denominator $\theta_{d,k}$ remains unchanged and the Monte Carlo estimator for $M_{k|k}$ is $\hat{M}_{k|k} = \hat{\theta}^{(2)}_{n,k}/\hat{\theta}_{d,k}$.

5.4 Variance reduction techniques

In order to improve the accuracy of the crude Monte Carlo estimator derived in section 5.3, it is desirable to replace sampling techniques as much as possible by analytic methods. A linear model is amenable to deterministic analysis and can be used as a reference to the non-linear plant provided it approximates the original system closely enough. This is the central idea of the control variate method which has been introduced in section 2.4 and used in section 3.4 for the trajectory prediction problem.

Instead of estimating the numerator $\underline{\theta}_{n,k}$ and denominator $\theta_{d,k}$ of eqn.(5.11) directly, we first establish a linear model and estimate then the errors between the model and system numerators and denominators respectively. Finally, these estimates are added to the analytically obtained values of the model numerators and denominators respectively. Thus, the linear model controls the fluctuations of the nonlinear system and thereby reduces the sampling error of the estimate $\hat{\underline{x}}_{k|k}$.

As the derivation of a suitable linear model yields a set of approximate nonlinear filter equations, the proposed Monte Carlo

procedure may improve the results of approximate nonlinear filtering solutions.

5.4.1 Statistical linearization

Sunahara[16] describes one feasible method for nonlinear continuous-time filtering which yields simultaneously a linear model. This model depends of course on the particular sequence of observations $\underline{y}^k$. We assume the initial condition of the nonlinear system (5.7) is given as an n-variate normal P.D.F. $p(\underline{x}_1)$; that is

$$p(\underline{x}_1) = n(\underline{x}_1; \underline{m}_x, \Sigma_x). \tag{5.23}$$

The P.D.F. $p(\underline{w}_k)$ and $p(\underline{v}_k)$ are assumed to be Gaussian and specified by eqns.(1.5) and (1.6).

Now, let

$$p_a(\underline{x}_k^* | \underline{y}^k) = n(\underline{x}_k^*; \underline{\mu}_{k|k}, \Sigma_{k|k}) \tag{5.24}$$

denote the n-variate Gaussian posterior P.D.F. of the state $\underline{x}_k^*$ of a linear model which is intended to approximate the original system (5.7) and which is defined by

$$\underline{x}_{k+1}^* = \underline{a}_k + B_k(\underline{x}_k^* - \underline{\mu}_{k|k}) + \underline{w}_k . \tag{5.25}$$

Similarly a model of the nonlinear observation system (5.8) is assumed to be given by

$$\underline{y}_k = \underline{c}_k + D_k(\underline{x}_k^* - \underline{\mu}_{k|k-1}) + \underline{v}_k \tag{5.26}$$

where $\underline{\mu}_{k|k-1}$ is the mean of the conditional P.D.F. of $\underline{x}_k^*$ prior to the

occurrence of observation $\underline{y}_k$; that is

$$p_a(\underline{x}_k^* | \underline{Y}^{k-1}) = n(\underline{x}_k^*;\ \underline{\mu}_{k|k-1},\ \Sigma_{k|k-1}). \tag{5.27}$$

The parameters $\underline{a}_k$, B_k, $\underline{c}_k$ and D_k are found by statistical linearization of the nonlinear transformations $\underline{f}$ and $\underline{g}$ such that in the expansions

$$\underline{f}(\underline{x}_k, k) = \underline{a}_k + B_k(\underline{x}_k - \underline{\mu}_{k|k}) + \underline{\varepsilon}_k \tag{5.28}$$

and

$$\underline{g}(\underline{x}_k, k) = \underline{c}_k + D_k(\underline{x}_k - \underline{\mu}_{k|k-1}) + \underline{\delta}_k, \tag{5.29}$$

the conditional expectations

$$E[\underline{\varepsilon}_k^T\ \underline{\varepsilon}_k | \underline{Y}^k] = E[\|\underline{f}(\underline{x}_k^*,k) - \underline{a}_k - B_k(\underline{x}_k^* - \underline{\mu}_{k|k})\|^2 | \underline{Y}^k] \tag{5.30}$$

w.r.t. $p_a(\underline{x}_k^* | \underline{Y}^k)$ and

$$E[\underline{\delta}_k^T \underline{\delta}_k | \underline{Y}^{k-1}] = E[\|\underline{g}(\underline{x}_k^*,k) - \underline{c}_k - D_k(\underline{x}_k^* - \underline{\mu}_{k|k-1})\|^2 | \underline{Y}^{k-1}] \tag{5.31}$$

w.r.t. $p_a(\underline{x}_k^* | \underline{Y}^{k-1})$ are minimal w.r.t. the model parameters $\underline{a}_k$, B_k, $\underline{c}_k$ and D_k. The necessary and sufficient conditions for minimizing eqn. (5.30) w.r.t. $\underline{a}_k$ and B_k are that

$$\begin{aligned} \underline{a}_k &= E[\underline{f}(\underline{x}_k^*, k) | \underline{Y}^k] \\ B_k &= E[(\underline{f}(\underline{x}_k^*, k) - \underline{a}_k)(\underline{x}_k^* - \underline{\mu}_{k|k})^T | \underline{Y}^k]\ \Sigma_{k|k}^{-1}\ . \end{aligned} \tag{5.32}$$

The expectation $E[\,\cdot\,]$ is w.r.t. $p_a(\underline{x}_k^* | \underline{Y}^k)$ of eqn.(5.24). Similarly, the conditions for minimizing eqn.(5.31) w.r.t. $\underline{c}_k$ and D_k are that

$$\underline{c}_k = E\left[\underline{g}(\underline{x}_k^*, k) \mid \underline{Y}^{k-1}\right]$$
$$D_k = E\left[(\underline{g}(\underline{x}_k^*, k) - \underline{c}_k)(\underline{x}_k^* - \underline{\mu}_{k|k-1})^T \mid \underline{Y}^{k-1}\right] \Sigma_{k|k-1}^{-1}. \tag{5.33}$$

The expectation $E[\cdot]$ is w.r.t. $p_a(\underline{x}_k^* | \underline{Y}^{k-1})$ of eqn.(5.27). As we assumed the Gaussian property in (5.24) and (5.27), we have

$$B_k = \frac{\partial \underline{a}_k}{\partial \underline{\mu}_{k|k}} \tag{5.34}$$

and

$$D_k = \frac{\partial \underline{c}_k}{\partial \underline{\mu}_{k|k-1}} . \tag{5.35}$$

The proof of these two results has been given in section 3.4.1. in connection with eqn.(3.70). They are useful for the analytic evaluation of B_k and D_k when $\underline{f}(\underline{x}_k, k)$ and $\underline{g}(\underline{x}_k, k)$ are expressed as polynomials in $\underline{x}_k$.

The parameters $\underline{\mu}_{k|k}$ and $\Sigma_{k|k}$ of eqn.(5.24) and $\underline{\mu}_{k|k-1}$ and $\Sigma_{k|k-1}$ of eqn.(5.27) are found by applying linear filtering and prediction theory[4,5] to the model (5.25) and (5.26). The parameters of $p_a(\underline{x}_k^* | \underline{Y}^{k-1})$ are found from eqn.(5.25) for $k > 1$ as

$$\underline{\mu}_{k|k-1} = \underline{a}_{k-1}$$
$$\Sigma_{k|k-1} = B_{k-1} \Sigma_{k-1|k-1} B_{k-1}^T + \Sigma_{w_k} . \tag{5.36}$$

Due to eqn.(5.36), it is possible to determine $\underline{c}_k$ and D_k from eqn. (5.33) which then specify $\underline{\mu}_{k|k}$ and $\Sigma_{k|k}$ of eqn.(5.24) for $k \geq 1$ as

$$\underline{\mu}_{k|k} = \underline{\mu}_{k|k-1} + K_k(\underline{Y}_k - \underline{c}_k)$$
$$\Sigma_{k|k} = \Sigma_{k|k-1} - K_k D_k \Sigma_{k|k-1} \tag{5.37}$$

where

$$K_k = \Sigma_{k|k-1} D_k^T (D_k \Sigma_{k|k-1} D_k^T + \Sigma_{v_k})^{-1} . \qquad (5.38)$$

The recursion (5.36) starts at time k = 1 with

$$\begin{aligned} \underline{\mu}_{1|0} &= \underline{m}_x \\ \Sigma_{1|0} &= \Sigma_x , \end{aligned} \qquad (5.39)$$

i.e. $p_a(\underline{x}^*_1 | \underline{y}^0) \equiv p(\underline{x}_1)$ is used to determine $\underline{c}_1$ and D_1. This completes the linearization procedure since $\underline{a}_k$ and B_k are obtainable from eqn. (5.32) using eqns.(5.37) and (5.38).

5.4.2 The multi-stage filtering problem using control variates

Applying Bayes' theorem of eqn.(5.6) to the linear model (5.25) and (5.26), a similar expression to eqn.(5.11) is obtained. We now combine these two relationships in order to rewrite $\underline{\theta}_{n,k}$ of eqn.(5.11) as

$$\begin{aligned} \underline{\theta}_{n,k} = &\int \cdots \int \underline{x}_k \prod_{i=1}^{k} p(\underline{y}_i | \underline{x}_i)\, p(\underline{x}^k)\, d\underline{x}^k \\ &- \int \cdots \int \underline{x}^*_k \prod_{i=1}^{k} p_a(\underline{y}_i | \underline{x}^*_i) p_a(\underline{x}^{*k})\, d\underline{x}^{*k} + \underline{\theta}_{an,k} \end{aligned} \qquad (5.40)$$

and $\theta_{d,n}$ of eqn.(5.11) as

$$\begin{aligned} \theta_{d,k} = &\int \cdots \int \prod_{i=1}^{k} p(\underline{y}_i | \underline{x}_i)\, p(\underline{x}^k)\, d\underline{x}^k \\ &- \int \cdots \int \prod_{i=1}^{k} p_a(\underline{y}_i | \underline{x}^*_i)\, p_a(\underline{x}^{*k})\, d\underline{x}^{*k} + \theta_{ad,k} . \end{aligned} \qquad (5.41)$$

In order that $\underline{\theta}_{n,k}$ and $\theta_{d,k}$, defined by eqns.(5.40) and (5.41), are the same as those values defined in eqn.(5.11) we have to set

$$\underline{\theta}_{an,k} = \int\cdots\int \underline{x}^*_k \prod_{i=1}^{k} p_a(\underline{y}_i|\underline{x}^*_i) p_a(\underline{x}^{*k})\, d\underline{x}^{*k} \tag{5.42}$$

and

$$\theta_{ad,k} = \int\cdots\int \prod_{i=1}^{k} p_a(\underline{y}_i|\underline{x}^*_i) p_a(\underline{x}^{*k})\, d\underline{x}^{*k} . \tag{5.43}$$

We recall that $p_a(.)$ denotes P.D.F. of variables occurring in the linear model (5.25) and (5.26). Because of the linearity of the model and the Gaussian properties of the random disturbances involved, the (nk)-fold integrals of eqns.(5.42) and (5.43) can be evaluated analytically. The Monte Carlo estimators to be used for the evaluation of eqns.(5.40) and (5.41) are defined by

$$\hat{\underline{\theta}}_{n,k} = N^{-1} \sum_{j}^{N} \left[(\underline{x}_k)_j \prod_{i=1}^{k} p(\underline{y}_i|(\underline{x}_i)_j) - (\underline{x}^*_k)_j \prod_{i=1}^{k} p_a(\underline{y}_i|(\underline{x}^*_i)_j) \right] + \underline{\theta}_{an,k} \tag{5.44}$$

and

$$\hat{\theta}_{d,k} = N^{-1} \sum_{j}^{N} \left[\prod_{i=1}^{k} p(\underline{y}_i|(\underline{x}_i)_j) - \prod_{i=1}^{k} p_a(\underline{y}_i|(\underline{x}^*_i)_j) \right] + \theta_{ad,k} . \tag{5.45}$$

Again, as explained in connection with the crude Monte Carlo estimator (5.14) and (5.15), $(\underline{x}_k)_j$ denotes the value of $\underline{x}_k$ obtained as the solution of eqn.(5.7) using $(\underline{x}_1)_j$ drawn from $p(\underline{x}_1)$ and the random sequence $(\underline{w}^{k-1})_j$ drawn from $p(\underline{w}_k)$. Similarly $(\underline{x}^*_k)_j$ denotes the value of $\underline{x}^*_k$ obtained by simulation of the linear model (5.25). The same random sequence $(\underline{x}_1, \underline{w}^{k-1})_j$ previously used to obtain $(\underline{x}_k)_j$ is used again to obtain $(\underline{x}^*_k)_j$ as the solution of eqn.(5.25).

The middle terms in eqns.(5.44) and (5.45) are known as the control variates for the first terms. A reduction in $\mathrm{var}(\hat{\theta}_{n,k})$ and $\mathrm{var}(\hat{\theta}_{d,k})$ is obtained provided the linear model giving rise to the control variates is a close approximation to the original nonlinear problem and absorbs most of its variations in the sampling procedure; i.e. $(\underline{x}^*_k)_j$ must be a good approximation to $(\underline{x}_k)_j$.

In view of this remark, it does not seem feasible to use a more general linearization procedure which would be concerned with the dynamics and the observation system independently of the particular sequence $\underline{y}^k$. The advantage of such a model would be its applicability without referring to one specific sequence $\underline{y}^k$. Such a procedure, however, will be excluded because the approximation accuracy which can be achieved does not satisfy the above mentioned condition.

To complete the control variate estimators (5.44) and (5.45), we finally have to carry out the analytic evaluation of eqns.(5.42) and (5.43). The denominator of eqn.(5.6) defined for the linear model (5.25) and (5.26) is denoted by $\vartheta_{ad,k}$ and is given by

$$\vartheta_{ad,k} = \iint p_a(\underline{x}^*_{k-1}|\underline{y}^{k-1}) p_a(\underline{y}_k|\underline{x}^*_k) p_a(\underline{x}^*_k|\underline{x}^*_{k-1}) d\underline{x}^*_k \, d\underline{x}^*_{k-1} . \quad (5.46)$$

It is related to the denominator $\theta_{ad,k}$ of eqn.(5.43) by

$$\vartheta_{ad,k} = \frac{\int\cdots\int \prod_{i=1}^{k} p_a(\underline{y}_i|\underline{x}^*_i) p_a(\underline{x}^{*k}) \, d\underline{x}^{*k}}{\int\cdots\int \prod_{i=1}^{k-1} p_a(\underline{y}_i|\underline{x}^*_i) p_a(\underline{x}^{*k-1}) d\underline{x}^{*k-1}} = \frac{\theta_{ad,k}}{\theta_{ad,k-1}} . \quad (5.47)$$

This result follows directly from the substitution carried out in

section 5.2.2. The posterior P.D.F. $p_a(\underline{x}_k^* | \underline{Y}^k)$ of the linear model is given by eqn.(5.24). Therefore the numerator $\theta_{an,k}$ is given by

$$\underline{\theta}_{an,k} = \theta_{ad,k} \underline{\mu}_{k|k} \tag{5.48}$$

where, from eqn.(5.47),

$$\theta_{ad,k} = \prod_{i=1}^{k} \vartheta_{ad,i} \quad . \tag{5.49}$$

Thus, the analytic part of the control variate method is reduced to the evaluation of eqn.(5.46). We shall prove the following result:

For the linear model (5.25) and (5.26) with Gaussian independent random disturbances $\underline{x}_1$, $\underline{w}_k$ and $\underline{v}_k$

$$\vartheta_{ad,k} = s_k \exp(-\frac{E_k}{2}) \tag{5.50}$$

where for $k \geq 1$

$$s_k = [(2\pi)^m |R_k|]^{-1/2} \tag{5.51}$$

$$E_k = (\underline{y}_k - \underline{c}_k)^T R_k^{-1} (\underline{y}_k - \underline{c}_k). \tag{5.52}$$

The matrix R_k is given for $k > 1$ as

$$R_k = \Sigma_{v_k} + D_k \Sigma_{w_k} D_k^T + D_k B_{k-1} \Sigma_{k-1|k-1} B_{k-1}^T D_k^T \tag{5.53}$$

and for $k = 1$ as

$$R_1 = \Sigma_{v_1} + D_1 \Sigma_x D_1^T . \tag{5.54}$$

Proof: The likelihood function $p_a(\underline{y}_k | \underline{x}_k^*)$ is Gaussian as the observation model (5.26) is linear and $p(\underline{v}_k)$ is normal. Similarly, both $p_a(\underline{x}_k^* | \underline{x}_{k-1}^*)$ and $p_a(\underline{x}_{k-1}^* | \underline{Y}^{k-1})$ are Gaussian. Therefore the integration

of eqn.(5.46) yields, with eqn.(5.7) and (5.26),

$$\vartheta_{ad,k} = p_a(\underline{y}_k | \underline{Y}^{k-1}) = s_k \exp[-\frac{1}{2}(\underline{y}_k - \underline{m}_k)^T R_k^{-1} (\underline{y}_k - \underline{m}_k)]. \qquad (5.55)$$

The mean $\underline{m}_k$ of this normal P.D.F. is found from eqn.(5.26) and (5.27) to be

$$\underline{m}_k = \underline{c}_k \, . \qquad (5.56)$$

To calculate the covariance matrix R_k we rewrite eqn.(5.26) using (5.25) as

$$\underline{y}_k = \underline{c}_k + D_k(\underline{a}_{k-1} + B_{k-1}(\underline{x}_{k-1} - \underline{\mu}_{k-1|k-1}) + \underline{w}_k - \underline{\mu}_{k|k-1}) + \underline{v}_k. \qquad (5.57)$$

Therefore

$$R_k = \Sigma_{v_k} + D_k \Sigma_{w_k} D_k^T + D_k B_{k-1} \Sigma_{k-1|k-1} B_{k-1}^T D_k^T. \qquad (5.58)$$

From eqn.(5.39), recursion (5.58) starts at time k=1 with

$$R_1 = \Sigma_{v_1} + D_1 \Sigma_x D_1^T \, .$$

Finally, as $\underline{y}_k$ is an m-dimensional vector and $p_a(\underline{y}_k | \underline{Y}^{k-1})$ is a Gaussian P.D.F., the constant s_k is given by

$$s_k = [(2\pi)^m |R_k|]^{-1/2} \, . \qquad (5.59)$$

The Monte Carlo estimate $\hat{\underline{x}}_{k|k}$ of the conditional mean $E[\underline{x}_k | \underline{Y}^k]$ is given by the ratio (5.18). The numerator $\hat{\theta}_{n,k}$ is computed with eqn.(5.44) and the denominator $\hat{\theta}_{d,k}$ with eqn.(5.45).

A Monte Carlo estimate $\hat{M}_{k|k}$ of the matrix of second order moments is given by

$$\hat{M}_{k|k} = \frac{\hat{\theta}^{(2)}_{n,k}}{\hat{\theta}_{d,k}} \qquad (5.60)$$

where

$$\hat{\theta}^{(2)}_{n,k} = N^{-1} \sum_{j}^{N} \Big[(\underline{x}_k)_j (\underline{x}_k)_j^T \prod_{i=1}^{k} p(\underline{y}_i | (\underline{x}_i)_j)$$

$$- (\underline{x}^*_k)_j \, (\underline{x}^*_k)_j^T \prod_{i=1}^{k} p_a(\underline{y}_i | (\underline{x}^*_i)_j) \Big] + \theta^{(2)}_{an,k}$$

and

$$\theta^{(2)}_{an,k} = (\Sigma_{k|k} + \underline{\mu}_{k|k} \underline{\mu}^T_{k|k}) \theta_{ad,k}$$

The denominator $\hat{\theta}_{d,k}$ is given by eqn.(5.45). The meaning of $(\underline{x}_k)_j$ and $(\underline{x}^*_k)_j$ is explained in the context of eqns.(5.44) and (5.45).

5.4.3 The computational procedure

1. Basic assumptions

In order to apply the Monte Carlo solution for estimating the conditional mean the following assumptions must be satisfied:

(A1) The statistical properties of the random signals describing the initial condition $\underline{x}_1$, the input noise $\underline{w}_k$ and the observation noise $\underline{v}_k$ must be known. For unknown system parameters the prior P.D.F. must be known.

(A2) The plant and observation equations are assumed to be given in the form of (5.7) and (5.8). The structure of these equations must be known but they may contain unknown parameters.

The discussion of a control variate estimator for the conditional mean

leads to the following computational procedure.

2. Observation sequence

Generate an artificial observation sequence $\underline{y}^k$ by solving eqn. (5.7) for one particular random initial condition $(\underline{x}_1)_j$ and noise sequence $(\underline{w}^{k-1})_j$. Add the observation noise $\underline{v}_k$ to the ideal system response $\underline{g}(\underline{x}_k, k)$ as indicated by eqn.(5.8).

3. Analytic approximation

(1) Set the time argument k=1 and compute the constants $\underline{c}_1$ and D_1 of the linear observation model (5.26) by means of eqn.(5.33).

(2) At time k, compute the conditional mean $\underline{\mu}_{k|k}$ and the covariance matrix $\Sigma_{k|k}$ by means of eqn.(5.37) and the parameters $\underline{a}_k$ and B_k of the linear dynamic model (5.25) by means of eqn.(5.32).

(3) Set k=k+1. If the end of the observation sequence is reached - stop. Otherwise compute $\underline{\mu}_{k|k-1}$ and $\Sigma_{k|k-1}$ by means of eqn.(5.36). Compute $\underline{c}_k$ and D_k by means of eqn. (5.33) and return to step (2).

4. Sampling procedure

(1) Draw the j:th initial variate $(\underline{x}_1)_j$ from $p(\underline{x}_1)$ and the j:th random sequence $(\underline{w}^{k-1})_j$ from $p(\underline{w}_k)$. Using these random variables in eqn.(5.7) the resulting j:th solution is $(\underline{x}^k)_j$.

(2) Use the j:th realization $(\underline{x}_1, \underline{w}^{k-1})_j$ of step (1) to simulate the linear model (5.25) in order to obtain $(\underline{x}^{*k})_j$.

(3) Repeat steps (1) and (2) N times and estimate the numerator $\hat{\theta}_{n,k}$ by means of eqn.(5.44) and the denominator $\hat{\theta}_{d,k}$ by means of eqn.(5.45).

(4) Compute the conditional mean estimate $\hat{\underline{x}}_{k|k}$ using eqn.(5.18). Its sampling covariance matrix can be obtained from eqn.(5.20) when all quantities are replaced by their estimates, as indicated in eqn.(5.21).

5.4.4 Linear multi-stage filtering

As mentioned before, a useful requirement for a Monte Carlo procedure is that it should give zero sampling variance when applied to linear systems. Although this property is not important in view of the solution of the linear filtering problem, it makes it plausible to expect small sampling variances when such a procedure is applied to nearly linear systems. We have the following result:

For a linear system

$$\begin{aligned} \underline{x}_{k+1} &= A_k \underline{x}_k + \underline{w}_k \\ \underline{y}_k &= C_k \underline{x}_k + \underline{v}_k \end{aligned} \tag{5.61}$$

the control variate method based on eqn.(5.44) and (5.45) yields a zero sampling variance for the estimate $\hat{\underline{x}}_{k|k}$ of the conditional mean $E[\underline{x}_k | \underline{y}^k]$.

For the proof of this result we first apply the statistical linearization procedure of section 5.4.1 to eqn.(5.61). Using eqn.(5.32) where the expectation is w.r.t. $p_a(\underline{x}_k^* | \underline{y}^k)$ of eqn.(5.24) we find

$$\underline{a}_k = A_k \underline{\mu}_{k|k} , \tag{5.62}$$

and from eqn.(5.34),

$$B_k = A_k . \tag{5.63}$$

Similarly, from eqn.(5.33) where the expectation is w.r.t. $p_a(\underline{x}_k^* | \underline{y}^{k-1})$ of eqn.(5.27), we find

$$\underline{c}_k = C_k \underline{\mu}_{k|k-1} , \tag{5.64}$$

and from eqn.(5.35),

$$D_k = C_k . \tag{5.65}$$

The substitution of the values from eqns.(5.62) ... (5.65) in eqns. (5.25) and (5.26) yields a linear model which is identical to the original system (5.61). That is, $p(\underline{y}_i | \underline{x}_i) \equiv p_a(\underline{y}_i | \underline{x}_i^*)$ for $i=1,2,\ldots,k$. Also because of this property the simulation procedure yields two identical random samples $\{\underline{x}^k\}_j$ and $\{\underline{x}^{*k}\}_j$ if the same initial variate $(\underline{x}_1)_j$ and the same random sequence $(\underline{w}^{k-1})_j$ is used for the original system (5.61) and the model (5.25) and (5.26). Therefore, the differences in eqns.(5.44) and (5.45) are identically equal to zero. This implies that the random part of these estimators is entirely removed, yielding zero sampling variance for $\hat{\underline{\theta}}_{n,k}$ and $\hat{\theta}_{d,k}$ and therefore also for $\hat{\underline{x}}_{k|k}$.

5.5 Approximate nonlinear filtering equations

The statistical linearization procedure used in the previous section yields as part of the nonlinear filtering solution a linear model. This is a very attractive feature when we come to use the control variate method to estimate the conditional mean. There are, however, other approaches to approximate nonlinear filtering. They yield sets of equations for the mean and the covariance matrix of the conditional distribution for $\underline{x}_k$ - given the sequence of observations $\underline{y}^k$. It is the object of this section to establish a Monte Carlo sampling procedure which:

(1) makes use of such nonlinear filtering equations in order to obtain an estimator with a reduced sampling covariance matrix compared with crude Monte Carlo methods;

(2) can be used to establish the performance of the approximate nonlinear filtering equations e.g. with respect to accuracy without an inordinate large number of simulations.

5.5.1 Nonlinear filtering equations for the conditional mean and covariance matrix

As already mentioned in chapter one, the research into the nonlinear estimation problem has quite naturally taken two avenues of approach. Much research effort has been expended in the area of continuous-time systems[13]. Under the white noise assumption, a rigorous development of approximate filter equations is possible[11].

On the other hand, however, this assumption complicates the interpretation of the mathematical results in the light of reality. We therefore restrict our discussion to discrete-time systems.

In this field there are two distinct contributions. Cox[20] developes a dynamic programming formulation of the estimation problem using the assumption that only an estimate of the present state is of interest. Indeed, it should be noted that the mode of the joint P.D.F. for a sequence of states does not correspond to the mode of the marginal P.D.F. for the state at a particular time k unless the system (5.1) and (5.2) is linear and is disturbed by additive Gaussian white noise. The solution developed by Cox leads to a discrete-time, nonlinear two-point-boundary value problem which is solved by the technique of discrete invariant imbedding. A similar approach is proposed by Detchmendy and Sridhar[22] for the continuous-time case. Their results are extended to discrete-time systems by Sage and Masters[21]. Although the solution does not require detailed statistical information concerning system disturbances, it is very similar to the solution of Cox.

The more recent work by Aoki[23] and Sorenson[24] is a generalization of linear Kalman-Bucy filtering theory[4,5]. The central idea is to approximate the posterior P.D.F. $p(\underline{x}_k|\underline{Y}^k)$ by a Gaussian P.D.F. $p_a(\underline{x}_k^*|\underline{Y}^k)$. We shall follow this second approach in some detail as it offers a feasible alternative control variate method to that one derived in the previous section.

Throughout this section 5.5 the notation is to be interpreted in

the same way as that of section 5.4 but all the mathematical definitions are distinct.

5.5.2 An approximate Gaussian posterior P.D.F. $p_a(\underline{x}_k^* | \underline{Y}^k)$

Let us again consider a system of the form of eqns.(5.7) and (5.8) and assume that $\underline{f}$, the plant eqn.(5.7), has at least continuous first derivatives and $\underline{g}$, the observation eqn.(5.8), has first and second order derivatives w.r.t. $\underline{x}_k$. An approximate posterior P.D.F. $p_a(\underline{x}_k^* | \underline{Y}^k)$ is postulated to be Gaussian and defined by

$$p_a(\underline{x}_{k+1}^* | \underline{Y}^{k+1}) = n(\underline{x}_{k+1}^*; \underline{\mu}_{k+1|k+1}, \Sigma_{k+1|k+1}). \tag{5.66}$$

Based on eqns.(5.7) and (5.8), the mean $\underline{\mu}_{k+1|k+1}$ and the covariance matrix $\Sigma_{k+1|k+1}$ are given by the following result, obtained by Sorenson[24]:

$$\begin{aligned}
\underline{\mu}_{k+1|k+1} &= \underline{z}_{k+1} + \Sigma_{k+1|k+1}\, G_{k+1}\, \Sigma_{v_k}^{-1} [\underline{Y}_{k+1} - \underline{g}(\underline{z}_{k+1}, k+1)] \\
\underline{z}_{k+1} &= \underline{f}(\underline{\mu}_{k|k}, k) \qquad (5.67) \\
\Sigma_{k+1|k+1} &= [(\Sigma_{w_k} + F_k \Sigma_{k|k} F_k^T)^{-1} + G_{k+1}^T \Sigma_{v_k}^{-1} G_{k+1} \\
&\qquad - \sum_{i}^{m} J_{k+1}(i)\, u_{k+1}(i)]^{-1}
\end{aligned}$$

where $u_{k+1}(i)$ are the components of the vector $\underline{u}_{k+1}$, defined by

$$\underline{u}_{k+1} \triangleq \Sigma_{v_k}^{-1} [\underline{Y}_{k+1} - \underline{g}(\underline{z}_{k+1}, k+1)]. \tag{5.68}$$

In addition to the notation introduced in chapter one we define the

following abbreviations for the first partial derivatives

$$F_k = \left. \frac{\partial \underline{f}(\underline{x}_k, k)}{\partial \underline{x}_k} \right|_{\underline{x}_k = \underline{\mu}_{k|k}} \qquad G_{k+1} = \left. \frac{\partial \underline{g}(\underline{x}_{k+1}, k+1)}{\partial \underline{x}_{k+1}} \right|_{\underline{x}_{k+1} = \underline{f}(\underline{\mu}_{k|k}, k)}.$$

Let us denote the i:th component of the observation equation $\underline{g}$ at time $k+1$ by $g((\underline{x}_{k+1}, k+1), i)$. The second partials of the i:th component of $\underline{g}$ are denoted by

$$J_{k+1}(i) = \left. \frac{\partial^2 g((\underline{x}_{k+1}, k+1), i)}{\partial \underline{x}_{k+1} \, \partial \underline{x}_{k+1}} \right|_{\underline{x}_{k+1} = \underline{f}(\underline{\mu}_{k|k}, k)}$$

The recursive eqns.(5.67) start at time $k = 1$ with

$$\begin{aligned} \underline{\mu}_{1|1} &= \underline{m}_x + \Sigma_{1|1} G_1 \Sigma_{v_1}^{-1} (\underline{y}_1 - \underline{g}(\underline{m}_x, 1)) \\ \Sigma_{1|1} &= \left[\Sigma_x^{-1} + G_1^T \Sigma_{v_1}^{-1} G_1 - \sum_i^m J_1(i) \, u_1(i) \right]^{-1} \end{aligned} \tag{5.69}$$

where

$$\underline{u}_1 = \Sigma_{v_1}^{-1} (\underline{y}_1 - \underline{g}(\underline{m}_x, 1)). \tag{5.70}$$

The proof of this result is given by Sorenson[24]. It is based on the expansion of the exponents of $p(\underline{y}_k | \underline{x}_k)$ and $p(\underline{x}_k | \underline{x}_{k-1})$ into a second order multidimensional Taylor series. The function $\underline{f}$ is expanded around $\underline{\mu}_{k|k}$ whereas $\underline{g}$ is expanded around $\underline{f}(\underline{\mu}_{k|k}, k)$. The scalar version of this procedure has been used in section 4.3.4 in connection with importance sampling.

Commenting on eqns.(5.67), we see that, in contrast to linear

Kalman-Bucy filtering theory[4,5], the covariance matrix $\Sigma_{k|k}$ depends upon the measurement data. Because the plant equation $\underline{f}$ is expanded about the conditional mean $\underline{\mu}_{k|k}$, all second order terms from the plant equation are eliminated in the relation for $\underline{\mu}_{k|k}$.

The recursive eqns.(5.67) for the mean $\underline{\mu}_{k|k}$ and covariance matrix $\Sigma_{k|k}$ of the approximate posterior P.D.F. provide a first step towards a control variate method. We now have to find a linear control variate model which must be a close approximation to the nonlinear plant (5.7) and observation equation (5.8). Unfortunately it is not possible to proceed directly from the conditional mean $\underline{\mu}_{k|k}$ and covariance matrix $\Sigma_{k|k}$ of eqn.(5.67) to a corresponding linear representation of the form

$$\begin{aligned} \underline{x}^*_{k+1} &= \underline{a}_k + B_k(\underline{x}^*_k - \underline{\mu}_{k|k}) \\ \underline{y}_k &= \underline{c}_k + D_k(\underline{x}^*_k - \underline{\mu}_{k|k}) \quad . \end{aligned} \tag{5.71}$$

The coefficients $\underline{a}_k$, B_k, $\underline{c}_k$ and D_k have to be evaluated by means of eqn.(5.32) and (5.33). In all four expressions the expectation $E[\,\cdot\,]$ is now performed w.r.t. the approximate Gaussian posterior P.D.F., defined by eqns.(5.66) and (5.67). This yields

$$\underline{a}_k = E[\underline{f}(\underline{x}^*_k, k) \,|\, \underline{y}^k] \quad ; \quad B_k = \frac{\partial \underline{a}_k}{\partial \underline{\mu}_{k|k}} \tag{5.72}$$

and

$$\underline{c}_k = E[\underline{g}(\underline{x}^*_k, k) \,|\, \underline{y}^k] \quad ; \quad D_k = \frac{\partial \underline{c}_k}{\partial \underline{\mu}_{k|k}} \quad . \tag{5.73}$$

As a consequence of eqn.(5.72) and (5.73), the mean $\underline{m}_{k|k}$ and covariance matrix $P_{k|k}$ of the posterior P.D.F. belonging to eqn.(5.71)

will differ from (5.67). This point emphasizes the suitability of the statistical linearization procedure of section 5.4 where the coefficients $\underline{a}_k$, B_k, $\underline{c}_k$ and D_k are obtained as part of the approximate filtering equation. In the present approach, however, they are introduced only for the development of the control variate method and are not part of the nonlinear filter approximation.

5.5.3 An alternative control variate method

As the linear model (5.71) and (5.72) is obtained in a different way to that used in section 5.4, we have to modify the analytic solution for $\underline{\theta}_{an,k}$ and $\theta_{ad,k}$, defined by eqns.(5.42) and (5.43), to

$$\underline{\theta}_{an,k} = \int\cdots\int \underline{x}_k^* \prod_{i=1}^{k} p_a(\underline{y}_i|\underline{x}_i^*)\, p_a(\underline{x}^{*k})\, d\underline{x}^{*k} \qquad (5.74)$$

and

$$\theta_{ad,k} = \int\cdots\int \prod_{i=1}^{k} p_a(\underline{y}_i|\underline{x}_i^*)\, p_a(\underline{x}^{*k})\, d\underline{x}^{*k}. \qquad (5.75)$$

The evaluation of these integrals proceeds again in a similar way to that considered in section 5.4. The denominator $\vartheta_{ad,k}$ of Bayes' theorem (5.6) applied to the linear system (5.71) is an m-variate Gaussian P.D.F. of the form

$$\vartheta_{ad,k+1} = p_a(\underline{y}_{k+1}|\underline{y}^k) = s_{k+1} \exp(-\tfrac{1}{2} E_{k+1}) \qquad (5.76)$$

where the exponent E_{k+1} is no longer given by eqn.(5.52) but found from eqn.(5.71) as

$$E_{k+1} = (\underline{y}_{k+1} - \underline{m}_{k+1})^T R_{k+1}^{-1} (\underline{y}_{k+1} - \underline{m}_{k+1})$$

and

$$s_{k+1} = [\,(2\pi)^m |R_{k+1}|\,]^{-1/2} .$$

The mean $\underline{m}_{k+1}$ and the covariance matrix R_{k+1} are directly obtainable from eqn.(5.71) and (5.76) as

$$\begin{aligned} \underline{m}_{k+1} &= \underline{c}_{k+1} + D_{k+1}\,(\underline{a}_k + B_k(\underline{m}_{k|k} - \underline{\mu}_{k|k}) - \underline{\mu}_{k+1|k+1}) \\ R_{k+1} &= D_{k+1}\, B_k\, P_{k|k}\, B_k^T\, D_{k+1}^T + D_{k+1}\, \Sigma_{w_k}\, D_{k+1}^T + \Sigma_{v_k}. \end{aligned} \tag{5.77}$$

These recursive equations start at time k=1 with

$$\begin{aligned} \underline{m}_1 &= \underline{c}_1 + D_1(\underline{m}_x - \underline{\mu}_{1|1}) \\ P_1 &= D_1\, \Sigma_x\, D_1^T + \Sigma_{v_1}. \end{aligned} \tag{5.78}$$

As already mentioned, the mean $\underline{m}_{k|k}$ and the covariance matrix $P_{k|k}$ of the posterior P.D.F. of the model (5.71) differ from the non-linear filter approximations $\underline{\mu}_{k|k}$ and $\Sigma_{k|k}$. This is the main difference to the control variate method proposed in section 5.4. Applying linear filtering theory to eqn.(5.71) yields for $\underline{m}_{k|k}$ and $P_{k|k}$:

$$\begin{aligned} \underline{m}_{k+1|k+1} &= \underline{a}_k + B_k(\underline{m}_{k|k} - \underline{\mu}_{k|k}) + P_{k|k}\, D_k\, \Sigma_{v_k}^{-1}(\underline{y}_k - \underline{m}_k) \\ P_{k+1|k+1} &= [(\Sigma_{w_k} + B_k\, P_{k|k}\, B_k^T)^{-1} + D_k^T\, \Sigma_{v_k}^{-1}\, D_k]^{-1}. \end{aligned} \tag{5.79}$$

These recursive equations start in the usual way at time k=1 from

$$\begin{aligned} \underline{m}_{1|1} &= \underline{m}_x + P_{1|1}\, D_1\, \Sigma_{v_1}^{-1}(\underline{y}_1 - \underline{m}_1) \\ P_{1|1} &= [\Sigma_x^{-1} + D_1^T\, \Sigma_{v_1}^{-1}\, D_1]^{-1}. \end{aligned} \tag{5.80}$$

This excursion into linear filtering theory is entirely dictated by the requirement for the control variate method that eqns.(5.74) and

(5.75) have to be evaluated analytically. As eqns.(5.79) and (5.77) define the solution for $\vartheta_{ad,k}$, we have, as in section 5.4,

$$\theta_{ad,k} = \prod_{i=2}^{k} p_a(\underline{y}_i|\underline{y}^{i-1}) p_a(\underline{y}_1) = \prod_{i=1}^{k} \vartheta_{ad,i} \qquad (5.81)$$

for eqn.(5.75) and finally

$$\underline{\theta}_{an,k} = \underline{m}_{k|k}\, \theta_{ad,k} \qquad (5.82)$$

for eqn.(5.74).

The estimators for $\hat{\underline{\theta}}_{n,k}$ and $\hat{\theta}_{d,k}$ are again given by eqns.(5.44) and (5.45) but $\theta_{ad,k}$ and $\underline{\theta}_{an,k}$ are now given by eqns.(5.81) and (5.82) respectively.

5.5.4 The computational procedure

Compared with section 5.4.3, we have to change the basic assumptions and the analytic approximations as follows.

1. Basic assumptions

In addition to the requirements laid down in section 5.4.3 we need the conditions:

(1) The plant equation $\underline{f}$ must have at least continuous first derivatives and the observation equation $\underline{g}$ at least continuous first and second order derivatives.

(2) As the defining equation for the approximate conditional covariance matrix $\Sigma_{k|k}$ contains negative terms, the present control variate method is only feasible if this matrix is positive definite.

3. Analytic approximation

(1) Evaluate the approximate conditional mean $\underline{\mu}_{k|k}$ and covariance matrix $\Sigma_{k|k}$ using eqn.(5.67).

(2) Compute the parameters $\underline{a}_k$, B_k, $\underline{c}_k$ and D_k of the linear model (5.71) with eqns.(5.72) and (5.73).

(3) Compute the mean $\underline{m}_{k|k}$ and covariance matrix $P_{k|k}$, eqns.(5.79), specifying the posterior P.D.F. of the model (5.71) in order to compute $\theta_{ad,k}$ with eqn.(5.81) and $\theta_{an,k}$ with eqn.(5.82). This requires the evaluation of $\underline{m}_k$ and R_k of the m-variate Gaussian P.D.F. $\vartheta_{ad,k}$ of eqn.(5.76) using eqn.(5.77).

The remaining operations 'observation sequence' and 'sampling procedure' are the same as in section 5.4.3. 2 and 4.

In chapter six which follows the nonlinear filtering techniques of this chapter for the multi-stage case and of chapter four for the single-stage case are applied to a variety of examples. The numerical study produces a number of interesting results.

CHAPTER SIX

NUMERICAL EXAMPLES OF NONLINEAR FILTERING TECHNIQUES

6.1 Introduction

The purpose of this chapter is to demonstrate how the Monte Carlo techniques developed in the previous two chapters can be employed for practical solutions of nonlinear single-stage and multi-stage filtering problems.

Section 6.2 is devoted to the memoryless case where we are given only one observation y. In order to avoid duplication we discussed the control variate methods in chapter five only. But these techniques can of course also be applied to the single-stage case. We will do this in order to compare them with importance sampling.

In subsequent examples contained in sections 6.3 and 6.4, we consider the problem of scalar multi-stage filtering. There are two justifications for this study:

(1) It provides an illustration of how variance reduction techniques improve a crude Monte Carlo estimator.

(2) The Monte Carlo approach can also be viewed as a method for improving approximate nonlinear filtering equations.

Finally, the multi-stage filtering problem is solved for a multivariate system in section 6.5. This application shows how our methods can be used for the state variable estimation problem when some of the states are system parameters.

6.2 Single-stage nonlinear filtering

6.2.1 Purpose and procedure

In this section we consider several scalar observation systems of the form of eqn.(4.6). In order to compare various solution procedures, we make the following assumptions with respect to the statistical properties of the prior P.D.F. $p(x)$ and the observation noise P.D.F. $p(v)$:

$$p(x) = n(x;\ 1.0,\ 0.5) \text{ and } p(v) = n(v;\ 0.0,\ 0.2). \qquad (6.1)$$

The observation y is assumed throughout section 6.2 to be given as $y = 1.3$.

Our first aim is to obtain results from the crude Monte Carlo estimator and then, subsequently, to improve these estimates by means of variance reduction techniques. Although the two procedures, importance sampling and the control variate method, are based on a similar concept they yield considerably different results. The numerical investigation reveals the reasons why this discrepancy arises.

The notation used in this section is compatible with that of chapter four. The Monte Carlo estimate of the conditional mean $E[x|y]$, denoted by $\hat{x}_y$, is computed with a random sample $\{x\}_j$ of size $N = 100$. The sampling variance $\mathrm{var}(\hat{x}_y)$ has to be replaced by its estimate $\hat{\mathrm{var}}(\hat{x}_y)$ as discussed in the context of eqn.(4.19). The approximate bias, denoted by x_{bias}, is defined by eqn.(4.17). An estimate $\hat{x}_{bias}$ is obtained if $\mathrm{var}(\hat{\theta}_d)$ and $\mathrm{cov}(\hat{\theta}_n, \hat{\theta}_d)$ are replaced by their estimates.

All the reported results are ensemble values. That is, the

Monte Carlo procedures for estimating $\hat{x}_y$, $\hat{var}(\hat{x}_y)$ and $\hat{x}_{bias}$ are repeated twenty times and $\bar{\hat{x}}_y$, $\overline{\hat{var}(\hat{x}_y)}$ and $\bar{\hat{x}}_{bias}$ denote the ensemble avarages of the outcomes of the Monte Carlo computations.

The comparison with the crude Monte Carlo estimator is based on the variance ratio η_v, defined by eqn.(2.71). In addition, we introduce a bias ratio η_{bias} which is defined in a similar way by

$$\eta_{bias} = \frac{(Bias)_{method\ 1}}{(Bias)_{method\ 2}} \quad . \tag{6.2}$$

Furthermore, for the comparison with approximate analytic solutions, we define the approximation error e by

$$e = |E[x|y] - \mu| \tag{6.3}$$

where μ is the analytic approximation of the conditional mean given by eqn.(4.58) for importance sampling and by eqn.(5.37) for the control variate method with k=1. Since $E[x|y]$ is unknown we replace it by the ensemble estimate $\bar{\hat{x}}_y$ of the Monte Carlo solution. Thus, the ensemble value $\bar{\hat{e}}$ of the error e is given by

$$\bar{\hat{e}} = |\bar{\hat{x}}_y - \mu| . \tag{6.4}$$

In order that the error estimates $\bar{\hat{e}}$ based on eqn.(6.4) are meaningful the sample size N must be big enough such that $\bar{\hat{x}}_y$ is a close approximation to $E[x|y]$ where the bias term $\bar{\hat{x}}_{bias}$ can be neglected and the sampling error of $\hat{x}_y$ is small compared with e.

6.2.2 The linear case

We start with a linear observation system of the form

$$y = x + v. \tag{6.5}$$

The conditional mean and variance of the posterior P.D.F. can easily be derived analytically as mentioned in section 4.3.2. Indeed, the optimal sampling P.D.F. $h^o(x)$ is found to be

$$h^o(x) = p(x|y) = n(x;\ \mu, \Sigma) \tag{6.6}$$

where, from eqn.(4.38) and (4.39), the mean μ and the variance Σ are given by

$$\mu = 1.21429 \qquad \Sigma = 0.14286. \tag{6.7}$$

The crude Monte Carlo estimator (4.15) using (4.12) and (4.13) yields the estimate $\hat{x}_y$ for the conditional mean $E[x|y]$. The sampling variance $\mathrm{var}(\hat{x}_y)$ and the bias x_{bias} are computed by means of eqns. (4.19) and (4.17) respectively. Using a sample of size N=100 the repeated application of the Monte Carlo procedure yields the following ensemble estimates:

$$\bar{\hat{x}}_y = 1.21285;\ \overline{\hat{\mathrm{var}}(\hat{x}_y)} = 1.2909.10^{-3};\ \bar{\hat{x}}_{bias} = -1.068.10^{-3} \tag{6.8}$$

Thus, the difference between $\bar{\hat{x}}_y$ and μ is less than one standard deviation $\left[\overline{\hat{\mathrm{var}}(\hat{x}_y)}\right]^{1/2}$. Assuming $\bar{\hat{x}}_y$ is a close approximation to $E[x|y]$ we have from eqns.(4.17), (6.7) and (6.8) $x_{bias} = -1.44.10^{-3}$. Furthermore, the sampling variance of the conditional mean is given by $\mathrm{var}(\hat{x}_y) = N^{-1}\Sigma$ where Σ is the conditional variance of $p(x|y)$.

Then, from eqn.(6.7) we have $\mathrm{var}(\hat{x}_y) = 1.429.10^{-3}$ as compared with $\overline{\hat{\mathrm{var}}(\hat{x}_y)} = 1.291.10^{-3}$ of eqn.(6.8). This comparison shows that the evaluation of eqn.(4.17) for estimating the bias $\hat{x}_{bias}$ and the evaluation of eqn.(4.19) for estimating the sampling variance $\hat{\mathrm{var}}(\hat{x}_y)$ yield values which are consistent with the true values.

Using eqn.(4.28) together with the antithetic variate method for the numerator θ_n in eqn.(4.48) yields a zero sampling variance estimator for the conditional mean $\hat{x}_y$; see sections 4.3.2 and 3.

6.2.3 Nonlinear observation system I

Next, we consider the observation system

$$y = 2\ \mathrm{Tanh}(x) + v \tag{6.9}$$

and begin with the evaluation of the parameters μ and Σ for the approximate importance function h(x). Using the results of section 4.3.4 we have, from eqns.(4.58) to (4.60),

$$h(x) = n(x;\ \mu,\ \Sigma\ ) \tag{6.10}$$

where $\mu = 0.7714$

$\Sigma = 0.2439.$

Table 6.1, which follows, contains the ensemble estimates $\bar{\hat{x}}_y$ of the conditional mean $E[x|y]$, of eqn.(4.8), the ensemble estimates $\bar{\hat{x}}_{bias}$ of the bias x_{bias} of eqn.(4.17) and the ensemble estimates $\overline{\hat{\mathrm{var}}(\hat{x}_y)}$ of the sampling variance $\mathrm{var}(\hat{x}_y)$ of eqn.(4.19). We set N=100 and use Method A : crude Monte Carlo method based on eqns.(4.12) and (4.13).

Method B : importance sampling for the denominator θ_d using eqns.(4.60) in (4.28) together with the antithetic variate method for the numerator θ_n using eqn.(4.60) in (4.61).

Method C : control variate method based on eqns.(5.44) and (5.45) with k= 1. The exponent of the approximate likelihood function $p_a(y|x)$ is given by eqn.(4.52) and $p_a(x|y)$ is set equal to h(x) of eqn.(6.10).

Method	$\bar{\hat{x}}_y$	$\overline{\hat{var}(\hat{x}_y)}$	$\bar{\hat{x}}_{bias}$
A	1.0577	$1.941 \cdot 10^{-3}$	$1.976 \cdot 10^{-3}$
B	1.0243	$3.975 \cdot 10^{-3}$	$5.536 \cdot 10^{-3}$
C	1.0613	$2.049 \cdot 10^{-4}$	$3.834 \cdot 10^{-4}$

Table 6.1 Sampling estimates for the conditional mean E[x|y =1.3]

These results show that method B does not improve the accuracy of the estimates for the conditional mean as the sampling variance of $\hat{x}_y$ is increased - compared with method A. On the other hand, the control variate method yields for the variance ratio η_v and the bias ratio η_{bias}, respectively

$$\eta_v = 9.5 \qquad \eta_{bias} = 5.1 \ . \tag{6.11}$$

To explain this result we plot in fig.6.1 the nonlinear function g(x) = 2 Tanh(x) together with the difference d, defined by

$$d = p(y|x) - p_a(y|x) \tag{6.12}$$

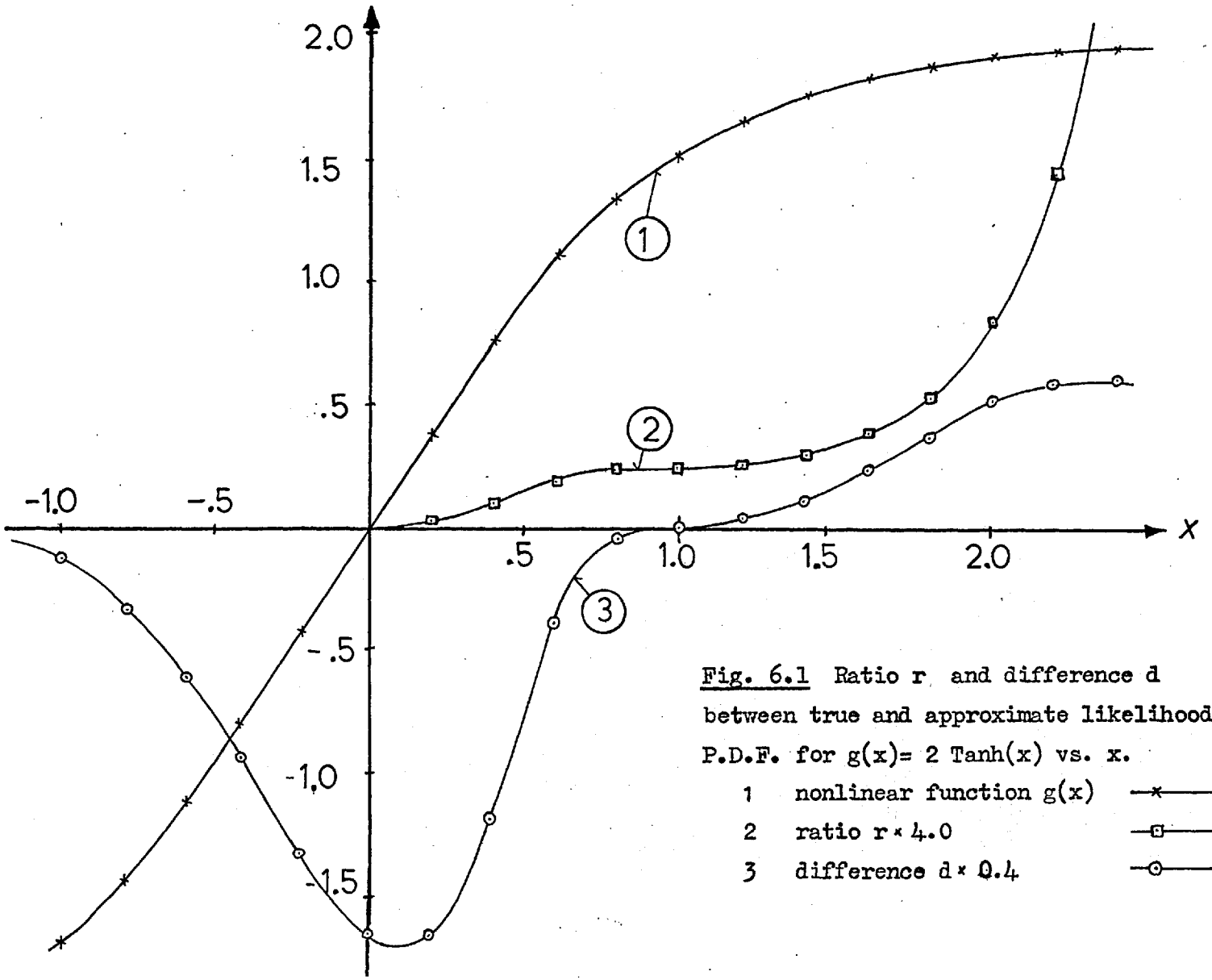

Fig. 6.1 Ratio r and difference d between true and approximate likelihood P.D.F. for g(x)= 2 Tanh(x) vs. x.

1 nonlinear function g(x)

2 ratio r × 4.0

3 difference d × 0.4

where

$$p(y|x) = \text{const.} \exp\left(-\frac{1}{2}\,\frac{[y - g(x)]^2}{\Sigma_v}\right) . \qquad (6.13)$$

For the P.D.F. $p_a(y|x)$ we use the second order expansion (4.52) as exponent in eqn.(4.51). Finally, we plot the ratio r defined by

$$r = \frac{p(y|x)\ p(x)}{h(x)} = \frac{p(y|x)\ p(x)}{p_a(y|x)p(x)} = \frac{p(y|x)}{p_a(y|x)} . \qquad (6.14)$$

We recall that the fixed observation is y=1.3.

The following observations and conclusions can be drawn from fig.6.1.

(1) Near the origin x=0, the function g(x) is almost linear but the majority of variates x_j drawn from p(x), eqn.(6.1), or h(x), eqn.(6.10), falls into the nonlinear part of the function g(x).

(2) The ratio r of eqn.(6.14) used for importance sampling method B stays close to unity around x=1.0 but differs considerably outside the interval $0.5 \leq x \leq 1.5$.

(3) The variations of the difference d used for the control variate method C remain close to zero and do not exhibit the unbounded variations of the ratio r.

(4) As a consequence of this behaviour, it is plausible to expect better estimates from the control variate method than from importance sampling (see table 6.1).

6.2.4 Nonlinear observation system II

The second nonlinear example is intended to show the difference between the statistical linearization procedure of section 5.4.1 (with k= 1) and the linearization of the exponent of the likelihood function p(y|x) of section 4.3.4 in order to design two control variate estimators. Let us assume that the nonlinear transformation g(x) of eqn.(4.6) is given by

$$y = 1.3\,(x - 1) - 0.2\,(x - 1)^3 + v$$

$$= 0.7x + 0.6x^2 - 0.2x^3 - 1.1 + v. \qquad (6.15)$$

The linearization of the exponent of p(y|x), described in section 4.3.4, to obtain an approximation h(x) to the optimal importance sampling P.D.F. $h^o(x)$ yields with eqns.(4.58)....(4.60)

$$h_2(x) = n(x;\, \mu,\, \Sigma) \qquad (6.16)$$

where

$$\mu = 1.8086 \quad \text{and} \quad \Sigma = 0.0957.$$

As mentioned in section 4.3.4, the function h(x) is set to be equal to the approximate posterior P.D.F. $p_a(x|y)$.

The statistical linearization procedure described in section 5.4.1 can of course also be applied to the system (6.15) if we set k=1 . The linear model (5.26) is found from eqn.(5.33) to be

$$y = 1.9(x^* - 1.) + v. \qquad (6.17)$$

Using this model to compute the approximate posterior P.D.F. $p_a(x^*|y)$ of eqn.(5.24) yields for the mean μ and variance Σ (obtainable from

eqn.(5.37) with k = 1) the values

$$\mu = 1.6159 \qquad \text{and} \qquad \Sigma = 0.0498 \; . \tag{6.18}$$

In order to establish the accuracy of the two approximations of eqns. (6.16) and (6.18) for μ, we apply the following Monte Carlo methods:

Method A : crude Monte Carlo, section 4.2.

Method B : importance sampling using eqn.(6.16) in eqn.(4.28) for the denominator $\hat{\theta}_d$ and using (6.16) in eqn.(4.61) for the numerator $\hat{\theta}_n$.

Method C : control variate method of section 5.4 for k=1 using eqn. (5.44) for $\hat{\theta}_n$ and (5.45) for $\hat{\theta}_d$. The model (6.17) is used for $p_a(y|x)$.

The results shown in table 6.2 are ensemble averages over twenty Monte Carlo experiments each based on a sample of size N = 100. The ensemble values $\overline{\hat{\text{var}}(\hat{x}_y)}$ and $\bar{\hat{x}}_{bias}$ are computed as discussed in section 6.2.3.

Method	$\bar{\hat{x}}_y$	$\overline{\hat{\text{var}}(\hat{x}_y)}$	$\bar{\hat{x}}_{bias}$	η_v	η_{bias}
A	1.9051	$5.331\ 10^{-3}$	$1.117\ 10^{-2}$	-	-
B	1.9032	$2.058\ 10^{-3}$	$5.017\ 10^{-3}$	2.6	2.2
C	1.8974	$1.903\ 10^{-4}$	$4.839\ 10^{-4}$	28.0	23.7

Table 6.2 Estimation of conditional mean for nonlinear system II using three different Monte Carlo methods.

In fig. 6.2 we plot the estimates of method A and C together with their 95% confidence intervals which are approximately given by two

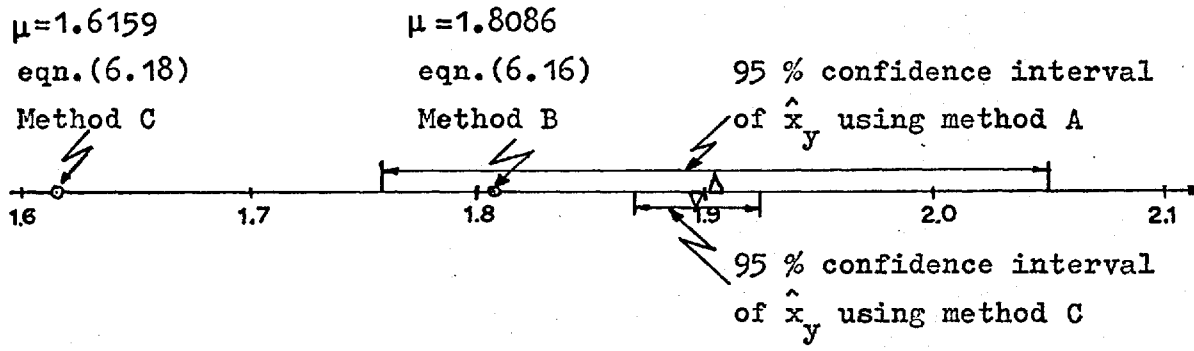

Fig. 6.2 Comparison of Monte Carlo estimates with analytic approximations.

standard deviations around the estimated mean $\bar{\hat{x}}_y$, i.e. $\left[\bar{\hat{x}}_y \pm 2[\overline{\hat{var}(\hat{x}_y)}]^{\frac{1}{2}}\right]$.

The conclusions to be drawn from this example are:

(1) The control variate method again gives results superior to those obtained from the importance sampling method B and the crude Monte Carlo method A. Indeed, the accuracy of method B compared with method A is only moderately improved because the ratio r, defined by eqn.(6.14), shows unbounded variations as already indicated in fig. 6.1 for system (6.9).

(2) The 95% confidence limit of the crude Monte Carlo estimate $\hat{x}_y$ excludes the approximate mean μ of eqn.(6.18) but it is not possible to estimate the approximation error $e = |E[x|y] - \mu|$ when μ = 1.8086 (see eqn.(6.16)) because $var(\hat{x}_y)$ of method A is too large.

(3) The sample size N of the crude Monte Carlo method A has to be increased considerably before we obtain a reliable estimate of the approximation error e, defined by eqn.(6.3). On the other hand, the control variate method C yields a sampling error which is significantly less than the estimated approximation error $\hat{e}$ of eqn.(6.4). As the 95%

confidence interval of $\hat{x}_y$, computed by means of method C, excludes both approximations of eqns.(6.16) and (6.18) we are able to decide which approximation is better suited for the problem at hand. Since the error estimate $\bar{\hat{e}}_C$ (using eqn.(6.18) for μ) is larger than the error $\bar{\hat{e}}_B$ (using eqn.(6.16) for μ), we conclude that eqn.(6.16) yields a better approximation to the conditional mean than eqn.(6.18).

(4) A comparison between the Monte Carlo estimates derived for the two observation systems (6.9) and (6.15) shows that the accuracy improvement of the proposed control variate method depends on the type of nonlinear function g(x). In the case of eqn.(6.15) all the derivatives of g(x) of order higher than three are equal to zero whereas in eqn.(6.9) this is not the case. As the proposed control variate method C is a second order approximation, (the mean μ depends on the variance Σ), it is plausible that the results for system (6.15) are more accurate than those for system (6.9).

6.3 Linear plant with a nonlinear observation system

6.3.1 Purpose and procedure

In order to proceed to the nonlinear multi-stage filtering problem, let us now consider the following scalar system

$$x_{k+1} = x_k + w_k. \tag{6.19}$$

The state is to be estimated from measurement data y^k that are related to the state by

$$y_k = x_k^2 + v_k. \tag{6.20}$$

The initial state x_1 and the plant and measurement noise sequences w_k and v_k are Gaussian variates possessing known P.D.F. specified by

$$\begin{aligned} p(x_1) &= n(x_1;\ 1.0\ ,\ 10^{-3}) \\ p(w_k) &= n(w_k;\ 0.\ ,\ 10^{-2}) \\ p(v_k) &= n(v_k;\ 0.\ ,\ 10^{-1}). \end{aligned} \tag{6.21}$$

The present example can be interpreted as a parameter estimation problem and has originally been studied by Denham and Pines[53] and later by Sorenson[24] to obtain approximate nonlinear filtering equations. They developed digital computer programs using crude Monte Carlo techniques to compare several different estimation policies. The main object of this section is to show that variance reduction techniques provide powerful simulation procedures for which it is no longer true that in order to compute meaningful sample means and variances the number of runs must be inordinately high.

The results summarized in tables 6.3 ... 6.5 are ensemble averages over ten realizations of the basic sampling experiment each of which is based on a sample of size N=500. Computing the estimates of the conditional mean $E[x_k|y^k]$, the sampling variance $\mathrm{var}(\hat{x}_{k|k})$ and the bias $x_{bias,k}$ we compare the following methods.

Method A : crude Monte Carlo method as described in section 5.3; see table 6.3.

Method B : control variate method using statistical linearization as described in section 5.4. The analytic approximation yielding $\mu_{k|k}$ and $\Sigma_{k|k}$ is based on eqns.(5.37) and (5.38); see table 6.4.

Method C : control variate method using the set of approximate nonlinear filtering equations (5.66)...(5.70); see table 6.5.

The observation sequence y^k is assumed to be given by $y_k=0.95$ for k =1,2....,10. The variance reduction factor η_v measures the accuracy improvement of methods B and C over A. The sample size N is chosen as N=500 such that the control variate methods B and C yield estimates $\hat{x}_{k|k}$ whose sampling error is less than the approximation error of the analytic solutions. Similar to eqn.(6.3) we define the approximation error e_k by

$$e_k = | E[x_k|y^k] - \mu_{k|k} | . \tag{6.22}$$

Assuming the ensemble value $\bar{\hat{x}}_{k|k}$ of the Monte Carlo estimates $\hat{x}_{k|k}$ is a close approximation to $E[x_k|y^k]$ the ensemble estimate $\bar{\hat{e}}_k$ of the error e_k is given by

$$\bar{\hat{e}}_k = | \bar{\hat{x}}_{k|k} - \mu_{k|k} | . \tag{6.23}$$

We use the notation $\bar{\hat{e}}_{k,B}$ and $\bar{\hat{e}}_{k,C}$ to indicate whether $\mu_{k|k}$ is the analytic approximation of method B or C.

In order to verify the expression for the sampling variance $\mathrm{var}(\hat{x}_{k|k})$ of eqn.(5.20) we compare the values based on the evaluation of eqn.(5.20) with those obtained from the statistical analysis of the Monte Carlo estimates $\hat{x}_{k|k}$. Since the Monte Carlo computation to evaluate $\hat{x}_{k|k}$ is repeated ten times in order to obtain the ensemble value $\bar{\hat{x}}_{k|k}$ we can estimate the range $\hat{s}$ between the largest and smallest value of $\hat{x}_{k|k}$. Assuming the estimates $\hat{x}_{k|k}$ are normally distributed, the

Time	Observation	Conditional mean	Bias	Sampling variance
k	y_k	$\bar{\hat{x}}_{k\|k}$	$\bar{\hat{x}}_{bias,k}$	$\overline{\hat{var}(\hat{x}_{k\|k})}$
			10^{-5}	10^{-5}
1	.95	.99914	0.188	0.182
2	.95	.98920	1.101	1.234
3	.95	.97610	1.029	1.729
4	.95	.97085	0.464	2.099
5	.95	.96805	-0.523	2.518
6	.95	.96638	-1.033	3.082
7	.95	.96111	-1.715	3.599
8	.95	.96007	-2.849	4.357
9	.95	.95874	-2.109	5.116
10	.95	.95818	-2.601	6.415

Table 6.3 Crude Monte Carlo method A to estimate the conditional mean $E[x_k | y^k]$.

Time	Observation	Control Variate Method B				Analytic Approximation B		
k	y_k	$\bar{\hat{x}}_{k\|k}$	$\bar{\hat{x}}_{bias,k}$	$\overline{\hat{var}(\hat{x}_{k\|k})}$	η_v	$\mu_{k\|k}$	$\Sigma_{k\|k}$	$\bar{\hat{e}}_{k,B}$
			10^{-7}	10^{-7}			10^{-2}	10^{-3}
1	.95	.99899	0.001	0.001	10000	.99902	0.0962	0.033
2	.95	.98834	0.627	0.512	2412	.99003	0.7625	1.692
3	.95	.97681	2.235	1.363	127	.98017	1.0423	3.361
4	.95	.96892	3.049	2.001	105	.97318	1.1442	4.251
5	.95	.96439	2.752	2.453	103	.96892	1.1832	4.530
6	.95	.96165	1.996	3.095	100	.96644	1.1997	4.793
7	.95	.96017	2.263	3.839	94	.96504	1.2074	4.875
8	.95	.95970	2.411	4.647	94	.96425	1.2113	4.554
9	.95	.95926	4.309	6.104	84	.96381	1.2134	4.546
10	.95	.95861	3.524	8.128	80	.96356	1.2145	4.946

Table 6.4 Control variate method B to estimate the conditional mean $E[x_k | y^k]$.

Time	Obser-vation	Control Variate Method C				Analytic Approximation C		
k	y_k	$\bar{\hat{x}}_{k\|k}$	$\bar{\hat{x}}_{bias,k}$	$\overline{\hat{var}(\hat{x}_{k\|k})}$	η_v	$\mu_{k\|k}$	$\Sigma_{k\|k}$	$\bar{\hat{e}}_{k,C}$
			10^{-7}	10^{-7}			10^{-2}	10^{-2}
1	.95	.99899	0.0003	0.002	10000	.99904	0.0960	0.005
2	.95	.98834	0.295	0.571	2161	.99177	0.7569	0.342
3	.95	.97682	2.314	1.556	111	.98489	1.0316	0.807
4	.95	.96889	3.780	2.405	87	.98043	1.1309	1.154
5	.95	.96456	5.061	3.036	83	.97786	1.1682	1.329
6	.95	.96173	5.815	4.190	74	.97642	1.1835	1.469
7	.95	.96041	6.770	5.199	69	.97563	1.1905	1.522
8	.95	.95972	7.561	6.767	64	.97519	1.1938	1.548
9	.95	.95914	8.382	9.065	56	.97496	1.1956	1.582
10	.95	.95865	9.770	9.997	55	.97483	1.1965	1.618

Table 6.5 Control variate method C to estimate the conditional mean $E[x_k|y^k]$.

range $\hat{s}$ multiplied by an appropriate factor λ (given in statistical tables) is a good measure of the standard error of $\hat{x}_{k|k}$.

In table 6.6 we compare the sampling variances of method A and B based on eqn.(5.20) with those values of the sampling variances obtained from the ten Monte Carlo estimates $\hat{x}_{k|k}$. Denote the range $\hat{s}$ by $\hat{s}_A$ for method A and $\hat{s}_B$ for method B. Then the alternative values of the sampling variances are given by $(\hat{s}_A\lambda)^2$ and $(\hat{s}_B\lambda)^2$ where $\lambda = 0.325$.

Time k	$(\hat{s}_A\lambda)^2$	$\overline{\hat{var}(\hat{x}_{k\|k})_A}$	$(\hat{s}_B\lambda)^2$	$\overline{\hat{var}(\hat{x}_{k\|k})_B}$
	10^{-5}	10^{-5}	10^{-7}	10^{-7}
1	0.24	0.182	0.0005	0.0002
2	0.582	1.234	0.923	0.512
3	1.378	1.729	0.942	1.363
4	1.244	2.099	1.801	2.001
5	0.950	2.518	3.762	2.453
6	1.141	3.082	2.450	3.095
7	1.757	3.599	4.928	3.839
8	3.526	4.357	3.610	4.647
9	5.984	5.116	5.251	6.104
10	5.512	6.415	6.914	8.128

Table 6.6 Sampling variances for $\hat{x}_{k|k}$ obtained from statistical analysis and from the evaluation of eqn.(5.20).

The results in table 6.6 show that the evaluation of eqn.(5.20) gives estimates of the sampling variance of $\hat{x}_{k|k}$ which are consistent with those based on the statistical analysis of the Monte Carlo estimates $\hat{x}_{k|k}$ and hence it is plausible to rely on eqn.(5.20).

In fig. 6.3 we plot the sampling errors $[\overline{\hat{var}(\hat{x}_{k|k})}]^{1/2}$ of methods A, B and C together with the approximation errors $\bar{\hat{e}}_{k,B}$ and $\bar{\hat{e}}_{k,C}$ defined by eqn.(6.23), over the time interval considered.

Finally, in table 6.7 we summarize the Monte Carlo estimates of the second order conditional moment $M_{k|k}$ using the control variate method B. The estimator (5.45) for the denominator $\hat{\theta}_{d,k}$ remains unchanged whereas the estimator for the numerator $\hat{\theta}^{(2)}_{n,k}$ is given by eqn. (5.60). The main purpose of estimating the second order moment is to

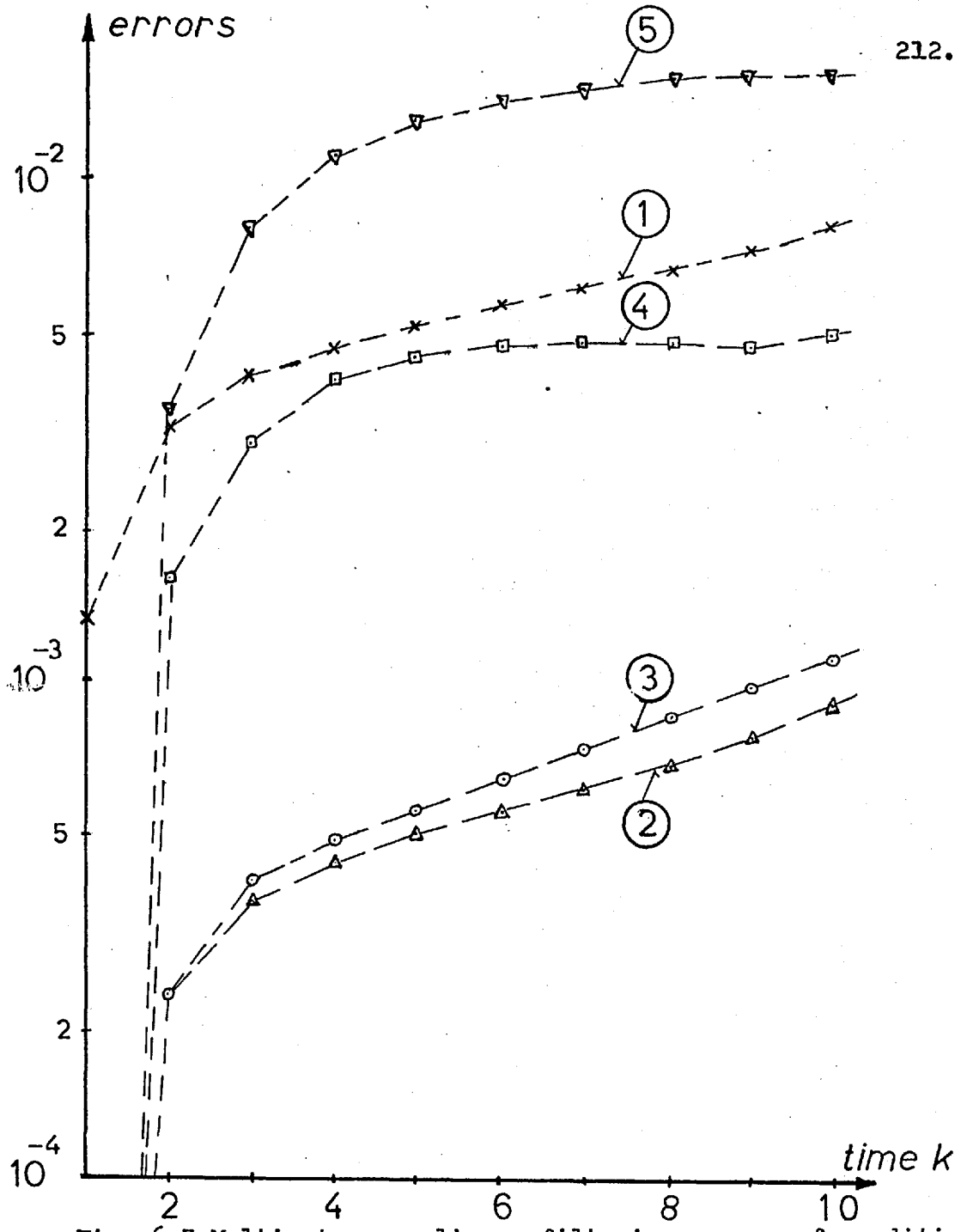

Fig. 6.3 Multi-stage nonlinear filtering errors of conditional mean $E[x_k | y^k]$ for system (6.19) and (6.20) vs. time k.

Sampling errors using:

1 crude Monte Carlo method A

2 control variate method B

3 control variate method C

Approximation errors of nonlinear filter equations using the analyti part of:

4 method B

5 method C

k	y_k	$\bar{\hat{M}}_{k\|k}$	$\overline{\widehat{var}(\hat{M}_{k\|k})}$	$M_{k\|k}$	$\bar{\hat{e}}_k^{(2)}$
			10^{-6}		10^{-2}
1	.95	.99893	.0001	.99708	0.185
2	.95	.98444	.1858	.97254	1.190
3	.95	.96469	.4571	.95031	1.438
4	.95	.95055	.6486	.93564	1.491
5	.95	.94223	.7956	.92696	1.527
6	.95	.93716	1.0188	.92202	1.514
7	.95	.93434	1.2953	.91923	1.511
8	.95	.93349	1.5754	.91767	1.582
9	.95	.93263	2.0696	.91679	1.584
10	.95	.93146	2.7986	.91630	1.516

Table 6.7 Control variate method B to estimate the conditional second order moment $E[x_k^2|y^k]$.

establish whether there exists an accuracy improvement for our Monte Carlo approach compared to the approximate analytic nonlinear filtering method. The descriptions of the P.D.F. involved and the sample size are the same as those previously used for estimating the conditional mean. The linear model allows us to compute the conditional second order moment, denoted by $M_{k|k}$, as $M_{k|k} = \Sigma_{k|k} + \mu_{k|k}^2$. The ensemble average of the approximation error $\bar{\hat{e}}_k^{(2)}$ shown in table 6.7 is defined as the difference between the Monte Carlo estimate $\bar{\hat{M}}_{k|k}$ of the second order moment and the analytic approximation $M_{k|k}$.

6.3.2 Observations and discussion

Several conclusions can be drawn from the numerical results of section 6.3.1.

(1) The crude Monte Carlo method provides a simple sampling technique for the nonlinear multi-stage filtering procedure. The result of eqn.(5.20) is verified in that the magnitude of the sampling variance based on eqn.(5.20) is consistent with the value we obtain from a statistical analysis of ten Monte Carlo estimates $\hat{x}_{k|k}$. The size of the approximate bias $x_{bias,k}$ obtained from evaluating eqn.(4.17) at each time k, is comparable with the sampling variance $var(\hat{x}_{k|k})$. We can therefore neglect the bias compared with the sampling error for the sample size chosen.

(2) Both control variate methods considered improve the crude sampling method significantly. Although the variance reduction factor η_v decreases as the time k increases, the rate of change in η_v is large only for the first two iterations.

(3) From fig. 6.3 it becomes apparent that method B yields a smaller sampling error than method C for estimating the conditional mean. This behaviour can be explained from the approximation error, also shown in fig. 6.3, this being considerably larger for method C than B. But the difference between the sampling errors of methods B and C is much smaller than the difference between the two approximation errors.

(4) Moreover, it can be seen from fig. 6.3 that the combination of sampling techniques and analytic approximations yields a procedure which improves the accuracy of the previously derived nonlinear filtering

solutions. Without an excessively large sample size N it is possible to decide that approximation B is better suited than method C for the problem at hand.

(5) Based on the numerical evidence of tables 6.3...6.5 and fig.6.3, it is clear that method B is well suited to solve the problem posed in section 6.3.1. This conclusion is supported by the results of table 6.7. Again the Monte Carlo solution for the estimation of the conditional second order moment is about an order of magnitude more accurate than the analytic approximation.

6.4 Nonlinear dynamics and nonlinear observation system

6.4.1 Purpose and procedure

While the example of the previous section 6.3 is a constant parameter estimation problem, the following scalar example represents a true multi-stage filtering problem. Let us consider the system

$$x_{k+1} = x_k - 0.2\, x_k^3 + w_k \tag{6.24}$$

where the states are observed by

$$y_k = \mathrm{Tanh}(x_k) + v_k. \tag{6.25}$$

The statistical information is assumed to be given by

$$p(x_1) = n(x_1;\ 1.0\ ,\ 10^{-2}) \tag{6.26}$$

$$p(w_k) = n(w_k;\ 0.\ ,\ 10^{-2}) \tag{6.27}$$

$$p(v_k) = n(v_k;\ 0.\ ,\ 10^{-1}) \tag{6.28}$$

In table 6.8 we compare the analytic approximations of methods B and C, specified in section 6.3.1. The objects of this example are to establish which set of nonlinear filtering equations yields smaller approximation errors and whether method B or C is to be preferred for the Monte Carlo solution.

Time	Observation	Method B		Method C	
k	y_k	$\mu_{k\|k}$	$\Sigma_{k\|k}$	$\mu_{k\|k}$	$\Sigma_{k\|k}$
			10^{-2}		10^{-2}
1	1.1	1.01410	0.9827	1.01367	0.9622
2	0.79	0.80757	1.1005	0.81281	1.0918
3	0.68	0.70390	1.3238	0.71140	1.3145
4	0.58	0.63162	1.5205	0.64100	1.5190
5	0.5	0.57394	1.6911	0.58480	1.7045
6	0.44	0.52497	1.8398	0.53681	1.8725
7	0.4	0.48300	1.9693	0.49533	2.0214
8	0.36	0.45072	2.0799	0.45728	2.1532
9	0.33	0.41874	2.1742	0.42271	2.2670
10	0.28	0.38501	2.2535	0.38613	2.3718

Table 6.8 Analytic approximation of nonlinear filtering equations.

In order to establish the accuracy of these results, we now perform a Monte Carlo experiment. We first apply the crude Monte Carlo method A which yields the ensemble estimates $\bar{\hat{x}}_{k|k}$ of the conditional mean $E[x_k \mid y^k]$ using eqn.(5.18). The control variate method makes either, use of statistical linearization of the plant and observation system (method B, section 5.4), or the linearization of the exponent of the likelihood function (method C, section 5.5). Table

6.9 contains the results of the three Monte Carlo experiments. The notation indicates that the results are ensemble averages over ten experiments each of size N=500.

In fig. 6.4 we plot the sampling errors of the crude estimator $\overline{[\hat{var}(\hat{x}_{k|k})_A]}^{1/2}$ and of the two control variate methods $\overline{[\hat{var}(\hat{x}_{k|k})_B]}^{1/2}$ and $\overline{[var(x_{k|k})_C]}^{1/2}$ respectively, and compare them with the approximation error $\bar{\hat{e}}_{k,B}$ defined by

$$\bar{\hat{e}}_{k,B} = |\bar{\hat{x}}_{k|k} - \mu_{k|k}| \tag{6.29}$$

where $\hat{x}_{k|k}$ is the Monte Carlo estimate using control variate method B and $\mu_{k|k}$ is obtained from eqn.(5.37); see also table 6.8. Similarly, the error $\bar{\hat{e}}_{k,C}$ in the case of the nonlinear approximation of the posterior P.D.F. $p(x_k|y^k)$ by a Gaussian P.D.F. is defined by

$$\bar{\hat{e}}_{k,C} = |\bar{\hat{x}}_{k|k} - \mu_{k|k}| , \tag{6.30}$$

but here $\hat{x}_{k|k}$ is the Monte Carlo estimate using method C and $\mu_{k|k}$ is obtained from eqn.(5.67).

6.4.2 Observations and discussion

The scalar multi-stage filtering problem presented in the previous section 6.4.1 allows us to draw the following conclusions.

(1) Both control variate methods B and C reduce the sampling variance of the crude Monte Carlo estimator; e.g. at time k=5 by a factor $\eta_v = 135$ and at time k=10 by $\eta_v = 58$.

Time	Observation	Crude Monte Carlo Method A		Control Variate Method B		Control Variate Method C	
k	y_k	$\bar{\hat{x}}_{k\|k}$	$\overline{\hat{var}(\hat{x}_{k\|k})}$	$\bar{\hat{x}}_{k\|k}$	$\overline{\hat{var}(\hat{x}_{k\|k})}$	$\bar{\hat{x}}_{k\|k}$	$\overline{\hat{var}(\hat{x}_{k\|k})}$
			10^{-5}		10^{-7}		10^{-7}
1	1.1	1.01126	1.934	1.01413	0.169	1.01418	0.198
2	0.79	0.80662	2.146	0.80823	1.604	0.80832	1.630
3	0.68	0.70567	2.578	0.70545	1.639	0.70533	1.877
4	0.58	0.63420	2.883	0.63433	1.976	0.63416	2.064
5	0.5	0.57918	3.158	0.57826	2.353	0.57804	2.509
6	0.44	0.53390	3.321	0.53102	3.138	0.53060	3.153
7	0.4	0.49461	3.634	0.49075	4.218	0.49043	4.303
8	0.36	0.45797	3.854	0.45419	5.346	0.45417	5.408
9	0.33	0.42600	4.208	0.42150	6.704	0.42136	6.881
10	0.28	0.39110	4.606	0.38706	7.952	0.38737	8.041

Table 6.9 Monte Carlo estimates $\bar{\hat{x}}_{k|k}$ of the conditional mean $E[x_k | y^k]$ for the nonlinear system (6.24), (6.25) using three different sampling procedures.

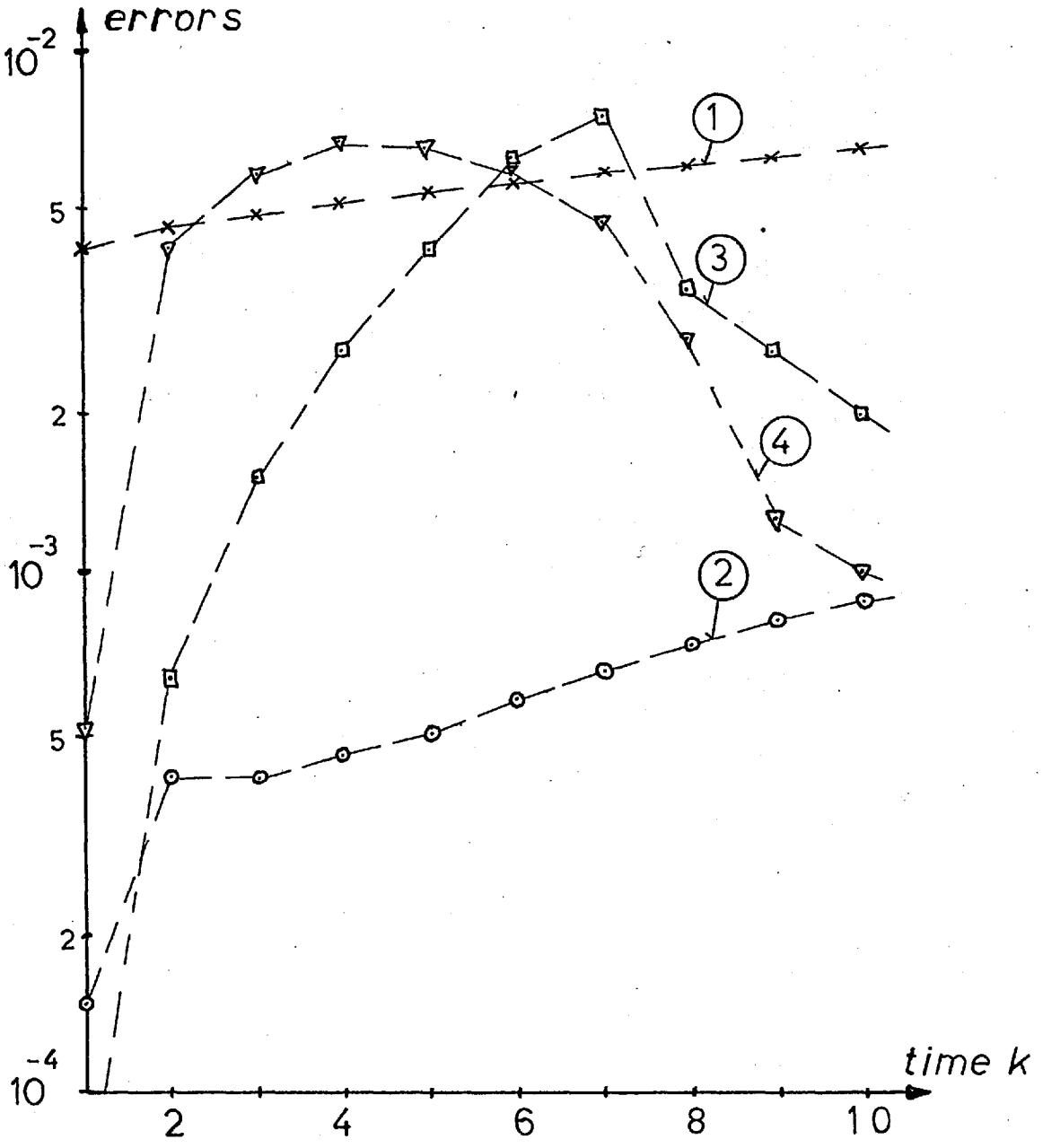

Fig. 6.4 Multi-stage nonlinear filtering errors of conditional mean $E[x_k|y^k]$ for system (6.24) vs. time k.

Sampling errors using

1 crude Monte Carlo method A

2 control variate method B or C

Approximation errors of nonlinear filter equations using the analytic part of

3 method B

4 method C

(2) As shown in fig. 6.4, the statistical linearization procedure of section 5.4 and the linearization procedure to obtain an approximate conditional P.D.F. of section 5.5 yield the same sampling errors for the estimate $\hat{x}_{k|k}$.

(3) The approximation error of the analytic part of method B is smaller for low values of k but for $k > 6$ the analytic solution based on method C gives a better approximation. As the two approximations of method B and C yield the results of linear filtering theory, when applied to a linear system, the approximation errors decrease in both cases for increasing k. Indeed as x_k tends to zero for increasing k, the system considered tends more and more to be a linear system with additive Gaussian noise.

(4) Using the theoretical results derived for the linear system in section 5.4.4, it is plausible to expect that the sampling error of the control variate methods will decrease as the time argument k increases. This inference, however, is only supported by analytic reflections and not by fig. 6.4 as the length of the time interval considered is not long enough.

(5) Both control variate methods require the computation of the parameters belonging to the linear model ; see (5.25), (5.26) or (5.71). This step involves the evaluation of integrals which, in the case of the observation eqn.(6.25), cannot be obtained by analytical methods. As numerical integration methods are required this remark is important in view of establishing a realistic labour ratio.

6.5 Nonlinear state and parameter estimation

6.5.1 Purpose and procedure

The object of this section is to investigate the applicability of Monte Carlo techniques to an essentially nonlinear multivariable problem. In particular, it is desired to estimate the states of a second order system. As some of the parameters of the system are considered to be unknown, we have to increase the dimensionality of the state vector and perform the filtering procedure for the new state vector.

Let us consider the second order nonlinear system

$$x_{k+1}(1) = x_k(1) + T\, x_k(2) \tag{6.31}$$

$$x_{k+1}(2) = x_k(2) - T(2x_k(1) + x_k^3(1)\, \alpha_k + 3x_k(2) - 10.) + w_k \,. \tag{6.32}$$

As usual the arguments in brackets denote the components of the vector $\underline{x}_k$. The quantity α_k is an unknown, time-varying parameter satisfying the relationship

$$\alpha_{k+1} = \alpha_k - T\, \alpha_k\, \beta_k \tag{6.33}$$

where β_k is a constant specifying the rate of change of α_k. Thus, its dynamics are simply

$$\beta_{k+1} = \beta_k \,. \tag{6.34}$$

The output of the plant considered is given by the scalar observation equation

$$y_k = x_k(1) + v_k \,. \tag{6.35}$$

For a specific sequence of observations y^k it is desired to estimate the two state variables $x_k(1)$ and $x_k(2)$ as well as the parameters α_k and β_k. To this end we define the new augmented state vector

$$\underline{x}_k^T = [\, x_k(1) \;\; x_k(2) \;\; \alpha_k \;\; \beta_k \,] \,. \tag{6.36}$$

In order to complete the specification of the problem, we assume the initial condition to be a known four-variate Gaussian P.D.F.

$$p(\underline{x}_1) = n(\underline{x}_1;\; \underline{m}_x,\; \Sigma_x) \tag{6.37}$$

with the mean $\underline{m}_x$ given by

$$\underline{m}_x^T = [\, 2. \quad 0. \quad 2. \quad 0.1] \tag{6.38}$$

and the covariance matrix Σ_x given by

$$\Sigma_x = \begin{bmatrix} 9.10^{-2} & 5.10^{-3} & 1.10^{-4} & 0. \\ 5.10^{-3} & 9.10^{-2} & 5.10^{-3} & 1.10^{-4} \\ 1.10^{-4} & 5.10^{-3} & 5.10^{-2} & 5.10^{-3} \\ 0. & 1.10^{-4} & 5.10^{-3} & 1.10^{-2} \end{bmatrix} . \tag{6.39}$$

This choice indicates that the prior information about $x_k(1)$ and $x_k(2)$ is less precise than that about the parameters α_k and β_k. The disturbance acting on the plant is also assumed to be a four-variate Gaussian P.D.F.

$$p(\underline{w}_k) = n(\underline{w}_k;\; \underline{0}\,,\; \Sigma_{w_k}) \,, \tag{6.40}$$

but the only non-zero element of the covariance matrix Σ_{w_k} is

$$\Sigma_{w_k}(2,2) = 5 \cdot 10^{-2} \,.$$

The variance Σ_{v_k} of the scalar Gaussian disturbance v_k acting on the observation channel (6.35) is assumed to be

$$\Sigma_{v_k} = 0.5 \,. \tag{6.41}$$

Finally, the factor T, being a measure of the nonlinearity of the system, is assumed to be T = 0.15.

In this section we only state the results of the statistical linearization procedure as this method seems to be particularly suited in view of the type of nonlinearity appearing in the system considered. Since the observation system (6.35) is linear, the model (5.26) is identical with eqn.(6.35). That is

$$c_k = \mu_{k|k-1}(1) \text{ and } D_k = [\,1 \;\; 0 \;\; 0 \;\; 0\,] \tag{6.42}$$

where $\mu_{k|k-1}$ is the mean of the P.D.F. $p_a(\underline{x}_k^* \,|\, y^{k-1})$ of eqn.(5.27). As the knowledge of c_k and D_k allows us to compute $\mu_{k|k}$ and $\Sigma_{k|k}$ of eqn.(5.24), we can evaluate the vector $\underline{a}_k$ from eqn.(5.32) as

$$\underline{a}_k(1) = \mu_{k|k}(1) + T\,\mu_{k|k}(2) \tag{6.43}$$

$$\begin{aligned} a_k(2) = {} & \mu_{k|k}(2) - T\,\big[2\mu_{k|k}(1) + 3\Sigma_{k|k}(1,3)\{\Sigma_{k|k}(1,1)+\mu_{k|k}^2(1)\} \\ & + \mu_{k|k}(3)\,\{3\mu_{k|k}(1)\,\Sigma_{k|k}(1,1)+\mu_{k|k}^3(1)\} + 3\mu_{k|k}(2) - 10.\big] \end{aligned} \tag{6.44}$$

$$a_k(3) = \mu_{k|k}(3) - T\,\big[\,\Sigma_{k|k}(3,4) + \mu_{k|k}(3)\,\mu_{k|k}(4)\big] \tag{6.45}$$

$$a_k(4) = \mu_{k|k}(4) \,. \tag{6.46}$$

The elements of B_k are obtained from eqn.(5.34) and (6.43)...(6.46).

The analytic part of the control variate method is completed by computing the correction terms $\underline{\theta}_{an,k}$ and $\theta_{ad,k}$ using eqn.(5.48) and (5.49).

In table 6.10 we compare the ensemble estimates of the conditional mean $\bar{\hat{\underline{x}}}_{k|k}$ obtained via the crude Monte Carlo method (method A, section 5.3) with those obtained from the control variate method (method B, section 5.4). The ensemble values are based on ten repetitions of the sampling experiment. The accuracy of the estimated conditional mean is given by the estimates of the sampling covariance matrix, $\hat{var}(\hat{\underline{x}}_{k|k})_A$ and $\hat{var}(\hat{\underline{x}}_{k|k})_B$ respectively. Sample size N is set N=500.

In fig. 6.5 we compare the sampling errors of the conditional mean estimates using the two Monte Carlo methods mentioned above with the analytic approximation error. This error is defined by eqn.(6.29). The graph displays the errors in the first component of $\bar{\hat{\underline{x}}}_{k|k}$. In fig. 6.6 we plot the corresponding values for the second component of $\bar{\hat{\underline{x}}}_{k|k}$. It becomes apparent from these two figures that the approximation errors are larger than the sampling errors of the control variate method. This is also true for the third component which behaves similarly. As the observation system 6.35 is linear $var(\hat{x}_{k|k})$ is equal to zero for k=1. The crude Monte Carlo method, however, yields a finite sampling error at k=1.

The variance reduction factors η_v for the first three components of $\hat{\underline{x}}_{k|k}$ are contained in table 6.10. In order to compute the efficiency gains of the proposed variance reduction technique, we have to specify a labour ratio. This factor depends on the problem considered because

Time	Observation	Component	Crude Monte Carlo Method		Control Variate Method		Statistical Linearization Approximation	
k	y_k	i	$\bar{\hat{x}}_{k\|k}(i)$	$\overline{\hat{var}(\hat{x}_{k\|k}(i))}$	$\bar{\hat{x}}_{k\|k}(i)$	$\overline{\hat{var}(\hat{x}_{k\|k}(i))}$	$\mu_{k\|k}(i)$	$\Sigma_{k\|k}(i,i)$
				10^{-4}				
1	2.0	1	1.99447	1.303	2.0	0.	2.	0.0763
		2	0.00109	1.798	0.	0.	0.	0.0899
		3	1.99945	1.014	2.0	0.	2.	0.0500
		4	0.09927	0.206	0.1	0.	0.1	0.0100
2	1.8	1	1.96849	1.119	1.97254	0.	1.97254	0.0686
		2	-1.49835	18.656	-1.51657	$5.943.10^{-5}$	-1.53511	1.1827
		3	1.96822	0.975	1.96900	0.	1.96897	0.0465
		4	0.09925	0.212	0.10000	0.	0.10000	0.0100
3	1.6	1	1.73916	0.348	1.74028	$1.211.10^{-6}$	1.73720	0.0179
		2	-2.25090	41.874	-2.26812	$1.567.10^{-4}$	-2.28325	2.5852
		3	1.94113	0.951	1.94186	$1.036.10^{-7}$	1.94122	0.0447
		4	0.09957	0.214	0.10031	0.	0.10031	0.0099
4	1.05	1	1.38515	0.607	1.38167	$6.343.10^{-6}$	1.37339	0.0309
		2	-1.87740	26.716	-1.90122	$1.993.10^{-5}$	-1.92527	1.4747
		3	1.92354	0.940	1.92385	$5.678.10^{-7}$	1.92509	0.0441
		4	0.10041	0.215	0.10142	0.	0.10142	0.0099
5	0.79	1	1.06810	1.729	1.06092	$5.211.10^{-6}$	1.03161	0.0899
		2	-0.78694	5.860	-0.78141	$5.253.10^{-5}$	-0.79382	0.2713
		3	1.90703	1.009	1.90640	$1.541.10^{-4}$	1.90914	0.0448
		4	0.10062	0.227	0.10219	0.	0.10219	0.0099

k	y_k	i	$\bar{\hat{x}}_{k\|k}(i)$	$\overline{\hat{var}(\hat{x}_{k\|k}(i))}$	$\bar{\hat{x}}_{k\|k}(i)$	$\overline{\hat{var}(\hat{x}_{k\|k}(i))}$	$\mu_{k\|k}(i)$	$\Sigma_{k\|k}(i,i)$
6	0.914	1	0.95901	$1.806 \cdot 10^{-4}$	0.95214	$3.964.10^{-6}$	0.91283	0.0991
		2	0.32822	3.230	0.33138	$9.868.10^{-6}$	0.37092	0.1367
		3	1.87987	1.147	1.87884	$2.079.10^{-6}$	1.88086	0.0473
		4	0.10033	0.242	0.10217	0.	0.10217	0.0099
7	1.26	1	1.05663	1.301	1.05116	$3.099.10^{-6}$	1.01259	0.0757
		2	1.02321	5.069	1.02829	$1.408.10^{-5}$	1.09238	0.2303
		3	1.84460	1.299	1.84421	$2.494.10^{-6}$	1.84527	0.0516
		4	0.10059	0.252	0.10225	0.	0.10225	0.0099
8	1.67	1	1.25719	0.812	1.25324	$3.217.10^{-6}$	1.22374	0.0479
		2	1.27398	7.116	1.28047	$2.785.10^{-5}$	1.36021	0.3250
		3	1.80612	1.483	1.80685	$3.253.10^{-6}$	1.80769	0.0572
		4	0.10209	0.263	0.10326	0.	0.10326	0.0099
9	1.99	1	1.47490	0.508	1.47208	$4.168.10^{-6}$	1.45659	0.0256
		2	1.18200	8.441	1.19050	$3.512.10^{-5}$	1.26507	0.4037
		3	1.76767	1.736	1.76953	$4.544.10^{-6}$	1.77107	0.0640
		4	0.10484	0.281	0.10543	0.	0.10543	0.0099
10	2.06	1	1.66388	0.401	1.66220	$4.770.10^{-6}$	1.65730	0.0133
		2	0.82709	8.576	0.83660	$3.912.10^{-5}$	0.89220	0.4223
		3	1.73319	2.024	1.73553	$6.184.10^{-6}$	1.73778	0.0715
		4	0.10740	0.296	0.10778	0.	0.10778	0.0099

Table 6.10 Estimates $\bar{\hat{\underline{x}}}_{k|k}$ of the conditional mean $E[\underline{x}_k \mid y^k]$ for the nonlinear system (6.31)... (6.35) using two different Monte Carlo methods.

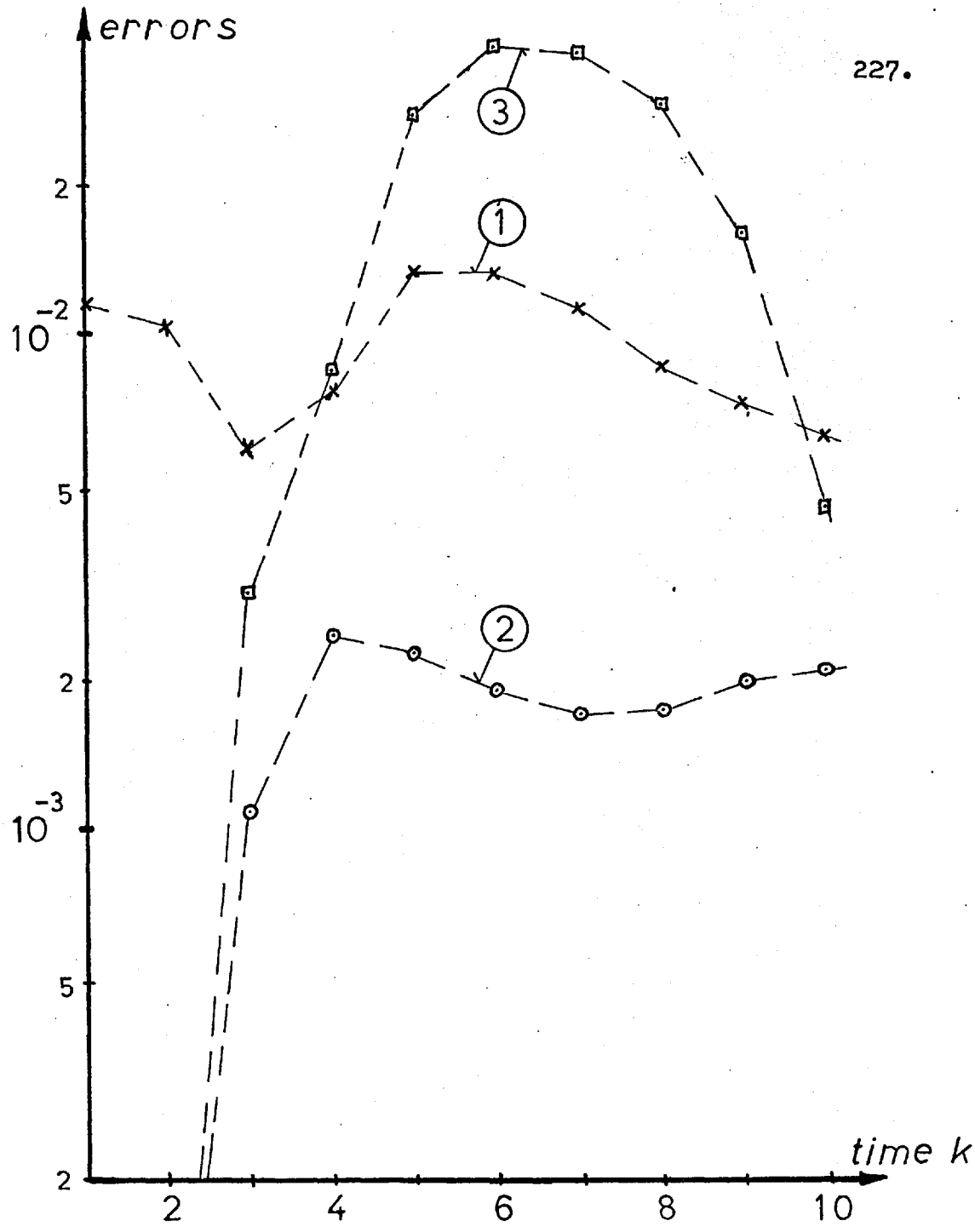

Fig. 6.5 Multi-stage nonlinear filtering errors of first component of conditional mean $E[\underline{x}_k|y^k]$ for system (6.31)...(6.35) vs. time k.

Sampling errors using

1 crude Monte Carlo method A - -*- -*- -

2 control variate method B - -o- -o- -

Approximation error of nonlinear filter equations using the analytic part of

3 method B - -□- -□- -

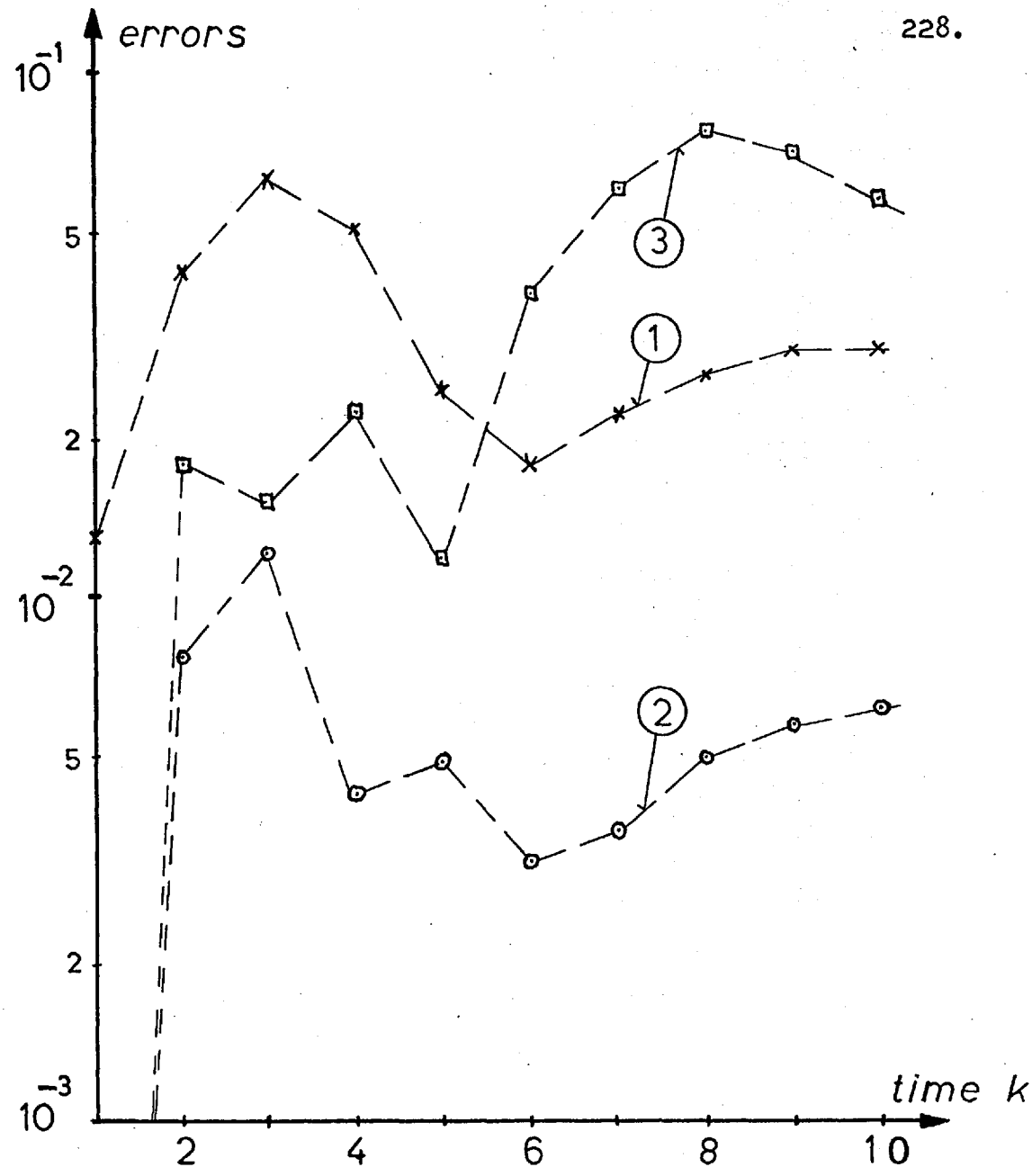

Fig. 6.6 Multi-stage nonlinear filtering errors of second component of conditional mean $E[\underline{x}_k | y^k]$ for system (6.31)...(6.35) vs. time k.

Sampling errors using

1 crude Monte Carlo method A -*--*-

2 control variate method B --o--o--

Approximation error of nonlinear filter equations using the analytic part of

3 method B --□--□--

the evaluation of the linear model parameters $\underline{a}_k$, B_k, $\underline{c}_k$ and D_k may require a numerical evaluation of the integrals (5.32) and (5.33). In our example, however, this computation can be done analytically and therefore the extra work to establish the linear model is small compared with the simulation time. A labour ratio of $\eta_L = 0.8$ is found to be adequate for our example; this factor is used in table 6.10 to obtain the efficiency gains η. The arguments in brackets refer to the component of $\hat{\bar{\underline{x}}}_{k|k}$.

k	$\eta_v(1)$	$\eta(1)$	$\eta_v(2)$	$\eta(2)$	$\eta_v(3)$	$\eta(3)$
1	∞	∞	∞	∞	∞	∞
2	∞	∞	31.4	25.2	∞	∞
3	28.7	22.2	26.8	21.4	918.1	735.5
4	9.6	7.7	134.1	107.3	166.3	133.0
5	33.0	26.4	26.1	20.9	65.4	52.3
6	45.6	36.6	32.8	26.2	55.2	44.2
7	43.3	34.7	36.0	28.8	52.1	41.7
8	25.3	20.2	25.6	20.5	45.5	36.4
9	12.2	9.8	24.1	19.3	38.2	30.6
10	8.5	6.8	21.9	17.5	32.7	26.2

Table 6.11 Variance reduction and efficiency gains of the control variate method over the crude Monte Carlo method.

Finally, in fig.6.7 we plot the sampling error of the estimated conditional mean $\bar{\hat{x}}_{k|k}$ for the first and second component $[\hat{var}(\hat{x}_{k|k}(i))]^{1/2}$, i=1,2, vs. the noise variance $\Sigma_{w_k}(2,2)$. These sampling errors obtained from the control variate method B are compared with the approximation errors $\bar{\hat{e}}_k(i)$, i=1,2.

6.5.2 Observations and discussion

The multivariable example supports the observations for scalar cases which were made previously. Indeed, the following points are verified:

(1) The control variate method yields a significant improvement over the crude Monte Carlo method. The variance reduction depends on the components of $E[\underline{x}_k|y^k]$. In particular, as the relationship (6.11) for the parameter β_k is linear, the sampling variance $var(\hat{x}_{k|k}(4))$ of the last component $\hat{x}_{k|k}(4)$ is zero.

(2) The analytic approximation error can be reduced by means of sampling techniques. This error reduction is, however, realized at the additional expense of generating a random sample $\{\underline{x}_1, w^{k-1}\}_j$ and simulating the given system N times.

(3) If the approximation error $\underline{e}_k$ depends on Σ_{w_k}, then the same is true of the sampling error $[var(\hat{\underline{x}}_{k|k})]^{1/2}$. The largest accuracy improvements of the proposed Monte Carlo methods are achieved for low values of Σ_{w_k}. This observation verifies Sorenson's[24] comment that

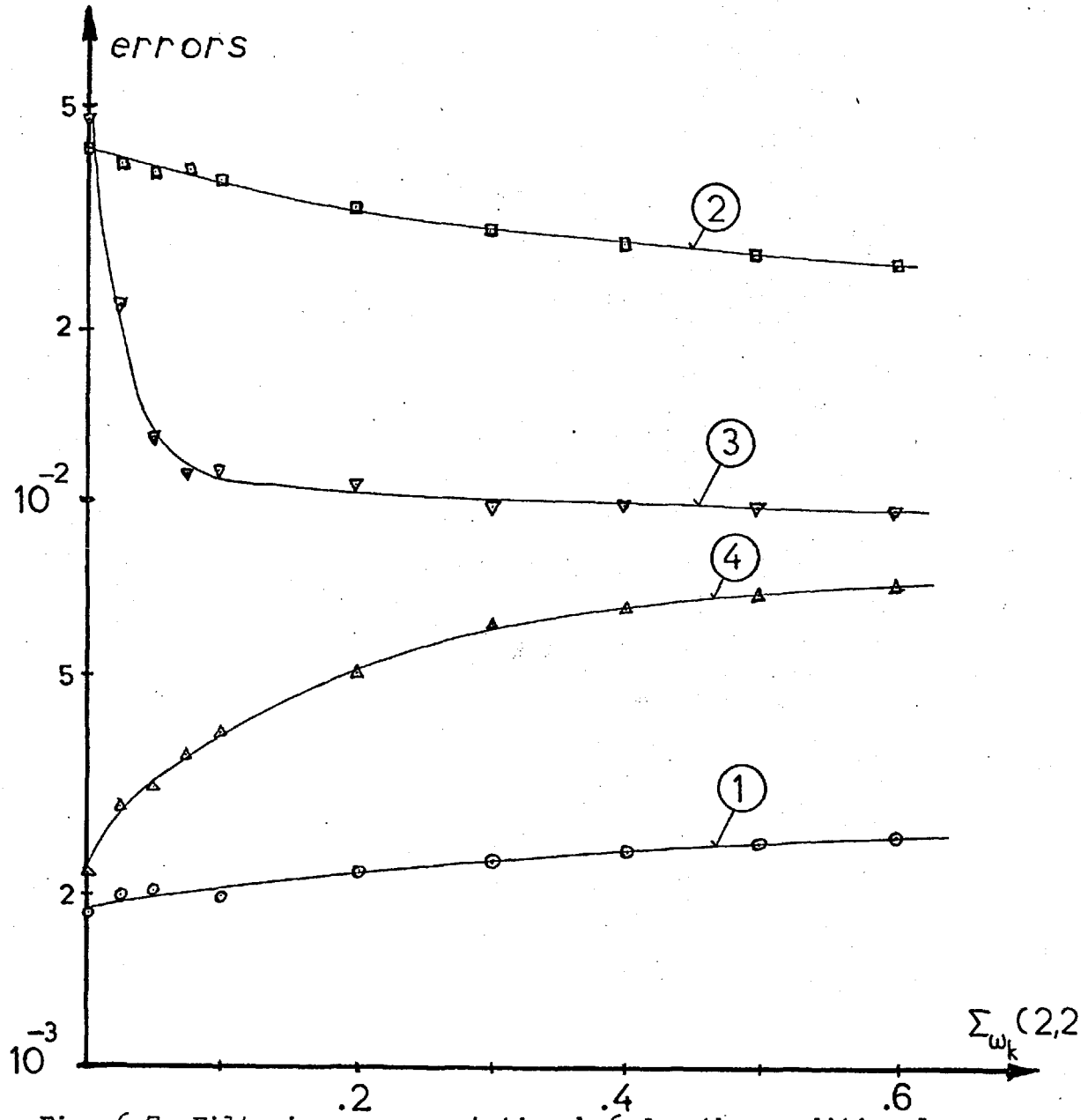

Fig. 6.7 Filtering errors at time k=6 for the conditional mean $E[\underline{x}_k | y^k]$ of system (6.31)...(6.35) vs. the variance $\Sigma_{w_k}(2,2)$ of the plant noise $\underline{w}_k$.

First component of $E[\underline{x}_k | y^k]$:

1 sampling error using control variate method B

2 approximation error using analytic part of method B

Second component of $E[\underline{x}_k | y^k]$:

3 sampling error using control variate method B

4 approximation error using analytic part of method B

the approximation errors of nonlinear filtering equations are more sensitive w.r.t. changes in Σ_{w_k} if the plant noise is small. Fig. 6.7 indicates that for the second component of $E[\underline{x}_k | y^k]$ this remark is true for $\Sigma_{w_k}(2,2) \leq 0.05$. For larger noise variances $\Sigma_{w_k}(2,2)$, the approximation error decreases whereas the sampling error increases. If the accuracy of nonlinear filtering equations is studied for low noise covariance matrices, the computer simulation using the control variate method has to be performed with double precision operations in order to obtain meaningful results.

CHAPTER SEVEN

SUMMARY AND CONCLUDING REMARKS

We first review in section 7.1 the main results of the present work. Suggestions for further research work are contained in section 7.2.

7.1 Summary and main conclusions

Chapter 1: The states $\underline{x}_k$ of nonlinear, discrete-time systems subject to stochastic disturbances evolve according to stochastic difference equations. Under the assumption that the P.D.F. of all random disturbances acting on the system are known, it is desired to predict the P.D.F. of the state of the system an arbitrary number of stages ahead. This is called the prediction problem. It is subdivided into the state prediction and trajectory prediction problem according to whether the time argument k is fixed or variable.

In general, however, the states $\underline{x}_k$ are not directly accessible but related to the output vector $\underline{y}_k$ through a nonlinear stochastic transformation. For a given sequence of observations $\underline{y}^k$ it is desired to obtain the posterior P.D.F. $p(\underline{x}_k|\underline{y}^k)$ as this function contains all the information about the system. This is called the filtering

problem with the additional qualification 'state' or 'trajectory' according to whether k is fixed or varying. Prediction and filtering are sometimes in this thesis collectively referred to as the estimation problem. It is always possible to design a crude Monte Carlo procedure to estimate one or more sample moments of $p(\underline{x}_k)$ and $p(\underline{x}_k | \underline{Y}^k)$ under the following assumptions:

(A1) the structures of the plant $\underline{f}$ and of the observation system $\underline{g}$ are known;

(A2) the statistical properties of the random sequences involved are completely specified;

(A3) the random variables $\underline{x}_1$, $\underline{w}_k$ and $\underline{v}_k$ are statistically independent with respect to each other and with respect to time k.

As the general nonlinear estimation theory still belongs very much to the future, the first objective of our thesis is the promotion of sampling techniques as a true alternative to, and useful improvement on, existing analytic approximations to optimal nonlinear predictors and filters. Our second intention is the development of solutions which enable us to compare different approximate nonlinear estimation methods.

Chapter 2: As our endeavour to obtain practical solutions to the nonlinear estimation problem is entirely based upon Monte Carlo methods, a brief discussion of some of the main features of these techniques is appropriate. A Monte Carlo solution consists of three distinct steps: the first two, choosing or specifying the probability process

and generating sample values of the random variables, are relatively straightforward problems. In particular, we show that they are not complicated by the dimensionality of the state $\underline{x}_k$ to be estimated. As a matter of convenience, all our numerical examples are based either on the uniform or on the Gaussian P.D.F. As far as the crude Monte Carlo procedure is concerned, the type of P.D.F. may be completely arbitrary. However, the form of the underlying P.D.F. becomes relevant for the design of variance reduction techniques.

Increases in the sample size of a crude technique yield but moderate improvements in the accuracy of Monte Carlo estimates. From this observation often follows the 'general conclusion' that Monte Carlo methods should be looked at only as last resort techniques. A large portion of our thesis is devoted to the third step mentioned above: the design of variance reduction techniques to dispel the belief that Monte Carlo methods are inaccurate for solving nonlinear estimation problems.

We refrain from discussing different variance reduction techniques for two reasons, these being that their basic concepts are well documented in various textbooks, and that their general formulation is of little use for our present problems. Indeed, we believe that the greatest variance reductions are made by exploiting specific details of the problem rather than by routine application of general principles. As a result of this attitude we develop a new multi-stage control variate method. The first order algorithm proposed in section 2.4.2

is an extension of a similar procedure derived in the context of importance sampling[37]. No rigorous result is available to prove convergence. But this disadvantage is effectively removed by the second order algorithm of section 2.4.4. where rapid convergence results from all three modifications. As the estimates of a multi-stage estimation procedure are correlated, it is necessary to derive an optimal weighting sequence to obtain a combined estimate with a minimum sampling variance. This topic is dealt with in section 2.4.3. Because this leads to an increased computing requirement, a two-stage estimator is proposed in section 2.4.6. In the first phase a good control variate function is iteratively determined and, in the second phase, it is held constant during the estimation procedure. Although the control variate function could be found by linear regression, the novel feature of our approach is the deterministic property of the matrix of second order derivatives which has to be inverted. The standard example of estimating a scalar integral serves as a numerical illustration (section 2.5) and confirms the fundamental conjecture that the sampling error in the final result is reduced if we can partially replace an estimate by an exact (deterministic) value.

Chapter 3: For nonlinear prediction we start with the problem formulation laid down in chapter one and consider systems where all the states $\underline{x}_k$ are accessible. In the discussion of sections 3.1.1 and 2 we elaborate on the difficulties inherent in the analytic approach to nonlinear prediction. If a solution is to be successfully based on

the Chapman-Kolmogorov equation, the P.D.F. involved have to have the self-reproducing property; that is, their structure must not change with time. If it does, the recursive relation of the Chapman-Kolmogorov equation is of little advantage and indeed the problems arising from the evaluation of $p(\underline{x}_k|\underline{x}_{k-1})$, or of the multidimensional integrals may prevent a direct solution. It is the object of section 3.1.3 to show that the exact nonlinear predictor requires the solution of an infinite set of nonlinear difference equations whereas in the linear Gaussian case the optimal predictor consists of two difference equations.

The nonlinearities between the states $\underline{x}_k$ and the system parameters exclude most standard estimation methods because of manipulative problems. The method of moments is presented in section 3.2.1 as a useful concept for the development of a Monte Carlo procedure. The sample moments of the P.D.F. $p(\underline{x}_k)$ are estimated in section 3.2.2 by simulating the nonlinear system, collecting the appropriate data and performing a statistical analysis. As this procedure does not include variance reduction techniques, it is termed crude Monte Carlo. The discussion shows that Monte Carlo methods are, generally speaking, ill suited to evaluate entire P.D.F. On the other hand, it may not be necessary to know the actual shape of the P.D.F. but one would like to be able, for example, to estimate its location and dispersion. It is this latter aspect which promotes the use of Monte Carlo methods for estimating sample moments.

At the expense of additional assumptions, the crude sampling procedure can be refined by variance reduction techniques. In section 3.3 the antithetic variate method is applied to the nonlinear prediction problem. To the original sample $\{\underline{x}_k\}_j^+$ of size N a negatively correlated sample $\{\underline{x}_k\}_j^-$ can be generated via algebraic operations and not by sampling techniques if the following assumption is satisfied:

(A4) The P.D.F. specifying the initial condition $\underline{x}_1$ and the random disturbance $\underline{w}_k$ must be unimodal and symmetric w.r.t. the mode.

It is not possible to state analytically how much the antithetic variate method will decrease the sampling variance in an arbitrary nonlinear system. Nonetheless, in section 3.3.2 we prove that it gives zero sampling variance for the mean when applied to a linear system with additive Gaussian noise. This result makes it plausible to expect large variance reductions when the antithetic variate method is applied to nearly linear systems.

A new extension of the antithetic variate method in connection with the n-variate Gaussian P.D.F. is derived in section 3.3.3. Under the assumption that

(A5) the P.D.F. involved are Gaussian and possess covariance matrices with all diagonal elements equal,

rearranged random vectors are obtained by permuting the components of the original variate. The resulting estimator remains unbiased while its variance comes close to the smallest value that can be attained with these variables.

A possible disadvantage of the antithetic variate method is its failure to improve significantly the estimators of higher order even sample moments. This deficiency is overcome by the control variate method, introduced in section 3.4, for the nonlinear prediction problem. Instead of increasing the sample size by the algebraic operations mentioned above, the control variate method requires a reference model which is amenable to sampling and analytic techniques. In other words, the number of simulations has to be increased. Statistical linearization is used in section 3.4.1 to set up a linear model of the original nonlinear system. We then obtain an improved sampling procedure if we make the assumption that

(A6) the noise $\underline{w}_k$ acting on the plant is additive and possesses a Gaussian P.D.F.

We prove that zero sampling estimators are obtained for first and second order moments when this method is applied to a linear system.

If assumptions (A4) to (A6) are not realistic, a Monte Carlo method is proposed in section 3.4.2 where the control variate model is derived by a linear regression technique. For the case where the P.D.F. involved are Gaussian, we introduce in section 3.4.3 a two-stage adaptive control variate method for the scalar prediction problem. In section 3.4.4 we show that the n-dimensional case can be reduced to n scalar problems. Compared with linear regression, our new control variate method using a gradient technique has the advantage that the matrix of second order derivatives is deterministic and can be inverted before the sampling procedure starts. As a consequence, the computing

expenditure compared with the linear regression approach is reduced.

In order to illustrate the different Monte Carlo methods a scalar and a multidimensional example are considered in sections 3.5.1 and 2. The conclusions we can derive from this numerical investigation are discussed in section 3.5.3 and are not restated here.

Chapter 4: To obtain a better grasp of the issues involved in nonlinear filtering problems we consider first the single-stage case; that is, we are given only one observation $\underline{y}$. It is desired to estimate some parameters of the posterior P.D.F. $p(\underline{x}|\underline{y})$. The advantage of this problem formulation is that the nonlinear dynamics do not have to be taken into consideration. The Bayesian approach to nonlinear filtering is adopted in this thesis and provides the starting point in section 4.2.1 for deriving a crude Monte Carlo method. The conditional mean $E[\underline{x}|\underline{y}]$ of the posterior P.D.F. has to be estimated as the ratio of two random variables. This gives rise to a number of problems which are discussed in section 4.2.2. It is shown that a finite sample size N makes the chosen Monte Carlo estimator biased. This bias can, however, be neglected for large N as it is much smaller than the sampling error.

It is understood that the single-stage case is but a sub-problem of the multi-stage filtering problem. We include therefore only one possible variance reduction technique in section 4.3, importance sampling. The discussion of other sampling improvements is deferred to chapter five.

The basic justification for importance sampling is that any sample can come from any distribution provided the correct weighting factors are used. In section 4.3.1 it is shown how a pair of new estimators for the numerator θ_n and denominator θ_d of Bayes' theorem is obtained. In section 4.3.2 we prove that there always exist an auxiliary P.D.F. $h^o(\underline{x})$ such that the sampling variance of $\hat{\theta}_d$ is zero. But this proof does not allow us to derive the optimal sampling P.D.F. $h^o(\underline{x})$ in a general nonlinear situation. Considering a linear Gaussian system, however, we derive an optimal sampling P.D.F. for the denominator of Bayes' theorem such that the sampling error is zero. Using this result together with the antithetic variate method (see section 3.3), we obtain a zero sampling variance estimator for the numerator $\hat{\underline{\theta}}_n$, section 4.3.3. The application of importance sampling to nonlinear systems requires the choice of an approximate auxiliary sampling function $h(\underline{x})$. We mention several possibilities in section 4.3.4 and develop one of them in some detail. The usefulness of importance sampling for improving the Monte Carlo method is limited because the weighting factor, being a ratio of two P.D.F., may exhibit unbounded variations outside a region of relative good fit.

The name conditional Monte Carlo would be very suitable for the type of work described in chapters four and five. This name has, however, been used in the past to denote a special variance reduction technique. In section 4.4 we attempt to find a unifying problem formulation for our work and the original conditional Monte Carlo method.

Chapter 5: The problem formulation and the solution of multi-stage nonlinear filtering are similar to the single-stage case of chapter four. But an additional complication is that the state $\underline{x}_k$ now changes from stage to stage according to the non-linear dynamic plant equation. Bayes' theorem provides in section 5.2 a recursive relationship for the posterior P.D.F. Using the Markov property of the stochastic processes considered, an expression is sought suitable for a Monte Carlo approach in the sense that the recursive equation must be replaced by an equivalent relationship containing only prior information about the system.

A crude Monte Carlo procedure for multi-stage filtering is proposed in section 5.3. Monte Carlo methods can be applied to a wide class of problems but until now have been cursed with low accuracy. In order to increase the sampling efficiency of the crude Monte Carlo filtering procedure, it is desired to derive a suitable variance reduction technique. In addition to (A1)-(A3) we assume that

(A7) the noise $\underline{w}_k$ acting on the plant and the noise $\underline{v}_k$ acting on the observation system are additive and possess Gaussian P.D.F.

We adapt a statistical linearization procedure, derived by Sunahara[16] for the continuous-time case, to the nonlinear discrete-time filtering problem. The resulting approximate nonlinear filtering equations are especially suited to the development of a control variate method because, as part of the analytic approximation, we obtain a linear

model. This model may be interpreted as a control variate model because it is a close approximation to the original nonlinear system for the particular sequence of observations considered. The combination of analytic and sampling techniques leads to a new and improved Monte Carlo filtering procedure; sections 5.4.1-3. It is shown in section 5.4.4 that our Monte Carlo method yields a zero sampling variance estimator for the conditional mean when applied to a linear system with an additive Gaussian white noise sequence. This result makes it plausible to expect a significant variance reduction when our procedure is applied to a nearly linear system.

An alternative control variate method for multi-stage nonlinear filtering is presented in section 5.5. The viewpoint is adopted that it is desirable to investigate a given nonlinear filter approximation. In proposing a suitable solution we have to take into consideration that the usual approximations consist of a set of equations for the conditional mean and covariance matrix. The result of Sorenson[24] serves as a starting point in the derivation of a sampling procedure which will allow us to establish the accuracy of the approximation without an inordinately large sample size N. Admittedly, the resulting solution is less elegant than that of section 5.4. Nevertheless, it shows that the Monte Carlo approach adopted in this thesis is versatile enough to cope with a variety of existing nonlinear filter approximations.

Chapter 6: The sampling procedures developed in chapters four and five are applied to a series of examples to illustrate features of our Monte

Carlo algorithms for estimating the conditional mean in both the static and dynamic cases. A number of conclusions is obtained from each example. They are not restated here except for the following four main objectives which are believed to be novel and which have been successfully achieved.

(1) Monte Carlo methods permit direct solutions to nonlinear filtering problems without resorting to linearized expressions for the plant or the measurement system.

(2) The approach adopted in this thesis enables accuracy limits to be directly obtained as part of the solution.

(3) The combination of analytic and sampling techniques improves the efficiency of Monte Carlo solutions significantly. The control variate method is to be preferred to importance sampling because the differences involved are less sensitive to inadequate matches than the ratios. The actual efficiency gain over the crude Monte Carlo solution depends on the nonlinear system considered, on its parameter values and on the length of the observation sequence $\underline{y}^k$. It is shown that the variance reduction is infinite if the system in question is linear and disturbed by an additive Gaussian white noise sequence.

(4) Previously derived analytic approximations do not indicate how accurate they are. Crude Monte Carlo methods have been used to estimate the approximation errors but, due to their low precision, the sample sizes have to be quite large in order to obtain useful approximation error estimates. A significant sample size reduction is made possible

by using the control variate method. It is shown that at the additional expense of simulating the system and model N times, the sampling errors can be made smaller than the approximation errors.

7.2 Suggestions for future research

In the remaining paragraphs, some general questions relating to the application of the theory and the method of approach are discussed. Topics that require additional investigation and areas for future research are included throughout the discussion.

(1.) The study of the adaptive control variate method in section 2.4 has shown that such a procedure may improve the variance reduction significantly compared with an arbitrarily chosen control variate function. On the other hand, such an adaptive procedure requires a careful study of its convergence properties. In section 2.4.4 we showed that the proposed second order algorithm converges with probability one. A possible extension would be to apply the adaptive sampling concept to the nonlinear state estimation problem. Suppose the original control variate model is obtained by statistical linearization. After the first Monte Carlo computation (using control variates) is completed, a new and improved control variate model could be obtained by linearizing the system along the estimated trajectory which - as our numerical examples have shown - is more accurate than the original analytic approximation. This idea leads again to a stochastic approximation procedure with the concomitant problems of

convergence, combination of correlated estimates, etc. As we assume a linear control variate model with Gaussian disturbances, such an application requires the estimation of the mean and covariance matrix for the nonlinear prediction problem and the estimation of the conditional mean and covariance matrix for the nonlinear filtering problem.

(2) The main reason why we use a linear model with additive Gaussian noise is the existence of a well developed linear estimation theory. However, it is understood that our choice may unnecessarily limit the obtainable sampling accuracy. It may be more advantageous to study P.D.F. from a class of functions which differ from the Gaussian P.D.F. but which are still specified by a finite number of parameters. As an example we mention the β-distribution which would allow us to include the effects of the third order moments (i.e. the skewness of the P.D.F.) and which may therefore approximate $p(\underline{x}_k | \underline{Y}^k)$ to greater accuracy.

(3) The Monte Carlo methods for predicting and filtering of nonlinear stochastic processes are based, in this thesis, on discrete-time systems. Some of the reasons for the use of this formulation are discussed in chapter one. A major difficulty created by the use of a discrete-time theory for continuous-time systems is that discrete-time models have first to be determined.

The statistical linearization procedure of section 5.4. was originally derived for continuous-time systems. It seems quite plausible that a control variate method can be derived for a continuous-time system, thus improving the crude Monte Carlo procedure and, possibly,

the analytic approximation provided the disturbances $\underline{w}_k$ and $\underline{v}_k$ can be modelled as discrete-time noise sequences. Such an investigation is well suited to be carried out on a hybrid computer installation. The simulation of the nonlinear dynamic equations would be carried out on the analog equipment and the statistical analysis together with further algebraic operations would be performed on the digital part.

(4) Some of the basic assumptions of chapter one may be relaxed. Allowing correlation between the plant noise $\underline{w}_k$ and the observation noise $\underline{v}_k$ affects Bayes' theorem (5.6). Instead of $p(\underline{x}_{k+1}|\underline{x}_k)$ and $p(\underline{y}_k|\underline{x}_k)$, one must deal with $p(\underline{x}_{k+1}|\underline{x}^k)$ and $p(\underline{y}_k|\underline{x}^k)$. Correlation therefore may complicate some expressions but does not affect the basic solution procedure.

The restriction of assumption (A5) on the generalized antithetic variate method in the formulation of section 3.3.3 may be relaxed by a suitable transformation.

(5) The reasons why estimation problems are interesting in their own right have been discussed in chapter one. Very often, however, they are considered to be part of an optimization problem. No consideration has been given in this thesis to establishing control policies in noisy nonlinear systems. An important extension of the present work would be the study of the combined estimation and optimization problem. Monte Carlo methods have already been successfully used by Mayne[29] for determining optimal control of nonlinear stochastic systems when all states are accessible. It is hoped that Monte Carlo methods can be developed for the optimization of nonlinear stochastic

systems when the observations are nonlinear, noise disturbed, functions of the states. Such a procedure may either result in a solution of the optimal control problem or serve as a method to test and compare several analytic approximations.

REFERENCES

(1) WIENER, N.: "The extrapolation, interpolation and smoothing of stationary time series." J. Wiley, New York, 1949.

(2) KOLMOGOROV, A.: "Interpolation and Extrapolation von stationären zufälligen Folgen." Bul. Acad. Sci. (U.S.S.R.), Ser. Math., 5, p.3-14, 1941.

(3) ZADEH, L. A.: "Progress in information theory in the U.S.A., 1957-1960: prediction and filtering". IRE Trans.PGIT, IT-7, p.139-144, 1961.

(4) KALMAN, R. E.: "A new approach to linear filtering and prediction problems." J. Basic Eng., Trans. ASME, 82D, p.33-45, 1960.

(5) KALMAN, R. E. and BUCY, R. S.: "New results in linear filtering and prediction theory." J. Basic Eng., Trans. ASME, 83 D, p.95-108, 1961.

(6) SORENSON, H. W.: "Kalman filtering techniques." Chap. 5 in Advances in control systems. Vol.3, ed. C.T. Leondes, Academic Press, New York, 1966.

(7) STRATONOVITCH, R. L.: "Conditional Markov processes." Theor. Probability Appl., 5, No.2, p.156-178, 1960.

(8) WONHAM, W. M.: "Stochastic problems in optimal control". IEEE Convention Record, Part 2, p.114-124, 1963.

(9) KUSHNER, H. J.: "On differential equations satisfied by conditional problem densities of Markov processes." J. SIAM Control, Series A 2, No.1, p.106-119, 1964.

(10) KUSHNER, H. J.: "Approximations to optimal nonlinear filters." IEEE Trans. AC-12, No.5, p.546-556, 1967.

(11) FISHER, J. R.: "Optimal nonlinear filtering." Chap.4 in Advances in Control Systems, Vol.5, ed. C.T. Leondes Academic Press, New York, 1967.

(12) BASS, R. W., NORUM, V. D. and SCHWARTZ, L.: "Optimal multichannel nonlinear filtering." J. math. Anal. Applic. 16, p.152-164, 1966.

(13) BUCY, R. S.: "Recent results in linear and nonlinear filtering." Symposium of American Automatic Control Council on stochastic problems in control, Univ. of Michigan, Ann Arbor, 1968.

(14) GUNCKEL, T. L.: "Optimal design of sampled data system with random parameters." Stanford Univ. Electronics Lab., Rep. 2102-2, 1961.

(15) BUCY, R. S.: "Nonlinear filtering theory." IEEE Trans. AC-10, No.2, p.198, 1965.

(16) SUNAHARA, Y.: "An approximate method of state estimation for nonlinear dynamical systems." Brown Univ. Tech. Rept. 1967-8, 1967.

(17) SCHWARTZ, L.: "Approximate continuous nonlinear minimal-variance filtering." Ph.D. thesis, Univ. of Calif. Los Angeles, 1967.

(18) JAZWINSKI, A. H.: "Nonlinear filtering with discrete observations." AIAA 3rd Aerospace Sciences Meeting, New York, 1966.

(19) HO, Y. and LEE, R. C. K.: "A Bayesian approach to problems in stochastic estimation and control." Proc. of Joint Autom. Control Conf., Stanford Univ., p.382-387, 1964.

(20) COX, H.: "On the estimation of state variables and parameters for noisy dynamic systems." IEEE Trans. AC-9, No.1, p.5-12, 1964.

(21) SAGE, A. P. and MASTERS, G. W.: "On-line estimation of states and parameters for discrete-time nonlinear dynamic systems." Proc. Nat. Electr. Conf., Vol. XXII, p.677-682, 1966.

(22) DETCHMENDY, D. M. and SRIDHAR, R.: "Sequential estimation of states and parameters in noisy nonlinear dynamical systems." Proc. of Joint Autom. Control Conf., Rensselaer Polytechnic Institute, New York, p.56-63, 1965.

(23) AOKI, M.: "Optimization of stochastic systems." Academic Press, New York, 1967.

(24) SORENSON, H. W.: "A nonlinear perturbation theory for estimation and control of time-discrete stochastic systems." Report No. 68-2, Univ. of Calif. Los Angeles, 1968.

(25) LAMBERT, F.: "Monte Carlo methods and simulation techniques." Chap. II in Mathematics and engineering applications, OECD Paris 1961.

(26) KAHN, H.: "Application of Monte Carlo." Rand Report, RM-1237-AEC, 1956.

(27) SHREIDER, Yu. A.: "The Monte Carlo method." Pergamon Press, London, 1966.

(28) KWAKERNAAK, H.: "On-line iterative optimization of stochastic control systems." Automatica, 2, No.3, p.195-208, 1965.

(29) MAYNE, D. Q.: "A gradient method for determining optimal control of nonlinear stochastic systems." IFAC Symposium on the theory of self-adaptive control systems, Teddington, 1965.

(30) LITTLE, W. D.: "Hybrid computer solutions of partial differential equations by Monte Carlo methods." Ph.D. thesis, Univ. of British Columbia, 1965.

(31) HANDLER, H.: "High-speed Monte Carlo technique for hybrid computer solution of partial differential equations." Ph.D. thesis, Univ. of Arizona, 1967.

(32) CUMMING, I. G.: "Computing aspects of problems in non-linear prediction and filtering." Ph.D. thesis, Univ. of London, 1967.

(33) CARNEY, T. M. and GOLDWYN, R. M.: "Construction and numerical evaluation of a group of estimators for joint state and parameter identification in autonomous linear systems." Report series EE 66, No.19, Rice University, Houston, 1966.

(34) WILKS, S. S.: "Mathematical statistics." J. Wiley, New York, 1962.

(35) PAPOULIS, A.: "Probability, random variables and stochastic processes." McGraw Hill, New York, 1965.

(36) HAMMERSLEY, D. M. and HANDSCOMB, D. C.: "Monte Carlo methods." Methuen's Monographs, London, 1965.

(37) KAHN, H.: "Use of different Monte Carlo sampling techniques." p.146-190, Symposium on Monte Carlo methods. Univ. of Florida 1954, ed. H.A. Meyer, J. Wiley, New York, 1956.

(38) MARSHALL, A. W.: "Application of multiple stage sampling procedures to Monte Carlo problems." p.123-140, Symposium on Monte Carlo methods, Univ. of Florida 1954, ed. H. A. Meyer, J. Wiley, New York, 1956.

(39) BUTLER, J. W.: "Machine sampling from given probability distributions." p.249-264. Symposium on Monte Carlo methods. Univ. of Florida, ed. H. A. Meyer, J. Wiley, New York, 1956.

(40) TROTTER, H. F. and TUKEY, J. W.: "Conditional Monte Carlo for normal samples." p.64-79. Symposium on Monte Carlo methods, Univ. of Florida, ed. H. A. Meyer, J. Wiley, New York, 1956.

(41) CLARK, C. E.: "Sampling efficiency in Monte Carlo analysis." Ann. Inst. Statist. Math. 15, 1964.

(42) PUGH, E. L.: "A gradient technique for adaptive Monte Carlo." SIAM Review, 8, No.3, 1966.

(43) DVORETZKY, A.: "On stochastic approximation." Proc. of the 3rd Berkeley symposium on mathematical statististics and probability, Vol.I, p.39-56, 1954.

(44) DOOB, J.: "Stochastic processes." J. Wiley, New York, 1953.

(45) DEUTSCH, R.: "Estimation theory." Prentice-Hall, Englewood Cliffs, N.J., 1965.

(46) ANDERSON, T. W.: "Introduction to multivariate statistical analysis." J. Wiley, New York, 1957.

(47) HAMMERSLEY, J. M. and MORTON, K. W.: "A new Monte Carlo technique : antithetic variates." Proc. Camb. phil. soc., 52, p.449-475, 1956.

(48) HAMMERSLEY, J. M. and MAULDON, J. G.: "General principles of antithetic variates." Proc. Camb. phil. soc., 52, p.476-481, 1956.

(49) TUKEY, J. W.: "Antithesis or regression?" Proc. Camb. phil. soc. 53, p.923-924, 1957.

(50) McGHEE, R. B. and WALFORD, R. B.: "A Monte Carlo approach to the evaluation of conditional expectation parameter estimates for nonlinear dynamic systems." IEEE Trans. AC-13, No. 1, p.29-36, 1968.

(51) HAMMERSLEY, J. M.: "Conditional Monte Carlo." J.Assoc. Comp. Mach., 3, p.73-76, 1956.

(52) WENDEL, J. G.: "Groups and conditional Monte Carlo." Ann. Math. Statist., 28, p.1048-1052, 1957.

(53) DENHAM, W. F. and PINES, S.: "Sequential estimation when measurement function nonlinearity is comparable to measurement error." AIAA, 4, No.6, p.1071-1076, 1966.

(54) CLARK, J. M. C.: "The representation of nonlinear stochastic systems with applications to filtering." Ph.D. thesis, Univ. of London, 1966.

(55) KAILATH, T. and FROST, P.: "Mathematical models for stochastic processes." Symposium of American Automatic Control Council on stochastic problems in control, Univ. of Michigan, Ann Arbor, 1968.

(56) LeMAY, J. L., MEDITCH, J. S. and STUBBERUD, A. R.: "Design and application of a digital computer programme for error analyses of a wide class of orbit navigation systems." 3rd IFAC Congress, paper 39A, London, 1966.

(57) GIESE, C. and McGHEE, R. B.: "Estimation of nonlinear system states and parameters by regression methods." Proc. of Joint Autom. Control Conf., Rensselaer Polytechnic Institute, New York, p.46-53, 1965.

(58) LINDLEY, D. V.: "Introduction to probability and statistics from Bayesian point of view." Camb. Univ. Press, Vol. II, 1965.

(59) KENDALL, M. G. and STUART, A.: "The advanced theory of statistics." C. Griffin, London, 1963.

(60) KUSHNER, H. J.: "Nonlinear filtering: the exact dynamical equations satisfied by the conditional mode." IEEE Trans. AC-12, No.3, 1967.

(61) LARSON, R. E. and PESCHON, J.: "A dynamic programming approach to trajectory estimation." IEEE Trans. AC-11, No.3, 1966.

(62) OHAP, R. F.: "A technique for the optimal estimation of the state of a nonlinear system exemplified by an orbit determination problem." M.Sc. Thesis, Univ. of California, Los Angeles, 1964.

(63) PERVOZVANSKII, A. A.: "Random processes in nonlinear control systems." Academic Press, New York, 1965.

(64) WONHAM, W. M.: "Some applications of stochastic differential equations to optimal nonlinear filtering.". J. SIAM Control, Series A 2, No. 3, p.347-369, 1964.

(65) MAYNE, D. Q.: "Estimation and control of stochastic systems." Ph.D. thesis, Univ. of London, 1967.

(66) KROY, W. H. and STUBBERUD, A. R.: "Identification via non-linear filtering." Int. J. Control, 6, No.6, p.499-522, 1967.

(67) BROWN, R. G.: "Smoothing, forecasting and prediction of discrete-time series." Prentice-Hall, Englewood Cliffs, N.J. 1965.

(68) HANDSCHIN, J. E. and MAYNE, D. Q.: "Monte Carlo techniques to estimate the conditional expectation in multi-stage nonlinear filtering." Int. J. Control (to appear).

(69) HANDSCHIN, J. E.: "Monte Carlo techniques for prediction and filtering of nonlinear stochastic processes." Paper submitted to IV. IFAC Congress, Warsaw, 1969.