NLR TR 80058 U

NATIONAAL LUCHT- EN RUIMTEVAARTLABORATORIUM

NATIONAL AEROSPACE LABORATORY NLR

THE NETHERLANDS

NLR TR 80058 U

ON MARTINGALES AND RECURSIVE OPTIMAL STATE ESTIMATION

BY

G. MOEK

NLR

DOCUMENT CONTROL SHEET

	ORIGINATOR'S REF.		SECURITY CLASS.
	NLR TR 80058 U		Unclassified

ORIGINATOR National Aerospace Laboratory NLR
Amsterdam. The Netherlands

TITLE
On martingales and recursive optimal state estimation

PRESENTED AT

AUTHORS	DATE	pp	ref
G. Moek	26-VI-1980	89	39

DESCRIPTORS

Stochastic processes
Estimation
Linear systems
Nonlinear systems
Martingales
Least squares method
Recursive functions
Kalman filter
State vector
White noise
Theorem proving
Coloured noise
Representation theorem
Innovation theorem

ABSTRACT

This report constitutes the deposit of a study of literature on martingales and estimation theory. Some experimental work concerning the efficiency of an estimation algorithm based on martingales is also described. The investigation has been guided by the intent to investigate martingales in relation to the Kalman and Kalman-Bucy filters and to consider the implications for the practical filtering work at NLR.

The optimal estimation problem and the Kalman and Kalman-Bucy filters are recalled. Martingales are defined and some examples are given. A number of theoretical results concerning martingales and linear as well as non-linear estimation theory is summarized in an innovation and presentation theorem. Based upon this a simple derivation of the Kalman and Kalman-Bucy filters is sketched. Reference is made to more general estimation problems (counting process observations, martingale noise) which can be solved along the same lines.

Finally, a wide-sense martingale approach to linear estimation is discussed. For a class of signals, comprising those studied in the Kalman and Kalman-Bucy problem, recursive estimation equations for the signal based on observations in either white or coloured noise are described. A more general model for coloured noise is given. By means of numerical simulation of some model problems the Kalman filter and the wide-sense martingale approach are compared.

NLR TR 80058 U

ON MARTINGALES AND RECURSIVE OPTIMAL STATE ESTIMATION

by

G. Moek

SUMMARY

This report constitutes the deposit of a study of literature on martingales and estimation theory. Some experimental work concerning the efficiency of an estimation algorithm based on martingales is also described. The investigation has been guided by the intent to investigate martingales in relation to the Kalman and Kalman-Bucy filters and to consider the implications for the practical filtering work at NLR.

The optimal estimation problem and the Kalman and Kalman-Bucy filters are recalled. Martingales are defined and some examples are given. A number of theoretical results concerning martingales and linear as well as non-linear estimation theory is summarized in an innovation and representation theorem. Based upon this a simple derivation of the Kalman and Kalman-Bucy filters is sketched. Reference is made to more general estimation problems (counting process observations, martingale noise) which can be solved along the same lines.

Finally, a wide-sense martingale approach to linear estimation is discussed. For a class of signals, comprising those studied in the Kalman and Kalman-Bucy problem, recursive estimation equations for the signal based on observations in either white or coloured noise are described. A more general model for coloured noise is given. By means of numerical simulation of some model problems the Kalman filter and the wide-sense martingale approach are compared.

This research has been performed partially under contract with the Netherlands Agency for Aerospace Programs (NIVR, contract no.1771).

Division: Informatics
Prepared: GM/ MK
Approved: WL/

Completed : 26-VI-1980
Ordernumber: 101.916/554.902
Typ. : GC

CONTENTS

3 Tables
31 Figures

(89 pages in total)

SYMBOLS

1) Latin characters

$a(k+1,k)$	matrix in coloured noise model
$A(k+1,k)$	matrix in linear discrete time state model
$A(t)$	matrix in linear continuous time state model
$A(t,x(t))$	matrix in non-linear continuous time state model
$\hat{A}(t,x(t))$	least squares estimate of $A(t,x(t))$
$B(k)$	matrix in linear discrete time state model
$B(t)$	matrix in linear continuous time state model
$B(t)$	stochastic process
$B(t,x(t))$	matrix in non-linear state model
$C(k)$	matrix in discrete time observation model
$C(t)$	matrix in continuous time observation model
$C(t,x(t))$	matrix in continuous time observation model
$\hat{C}(t,x(t))$	least squares estimate of $C(t,x(t))$
dt	time step
$D(t)$	matrix in continuous time observation model
$E(k)$	matrix (5.17)
f	arbitrary function or functional
$F(k)$	matrix (5.18)
$F(k)$	arbitrary matrix
$F(t,s)$	matrix (4.5)
$g(t,\tau)$	kernel in representation (3.31)
$h(t,\tau)$	kernel in representation (3.33)
$H(k)$	discrete time observation matrix
$H(t)$	continuous time observation matrix
$\tilde{H}(k)$	$H(k) - a(k,k-1)\ H(k-1)$ (4.49)
I	unit matrix
k	symbol for discrete time
$K(k)$	discrete time gain matrix
$K(k+1 \mid k)$	discrete time gain matrix
$K(t)$	continuous time gain matrix
ℓ	symbol for discrete time
$L_2^n(\Omega)$	Hilbert space of square integrable random n-vectors
$L_2(x;k)$	subspace of $L_2^n(\Omega)$ spanned by $x(i), i\leq k$
$L_2(x(k))$	subspace of $L_2^n(\Omega)$ spanned by $x(k)$
$L_2(x;(t))$	subspace of $L_2^n(\Omega)$ spanned by $x(s)$, $s\leq t$

$L_2(x(t))$	subspace of $L_2^n(\Omega)$ spanned by $x(t)$
$L_2^n(y;k)$	subspace of $L_2^n(\Omega)$ spanned by $F(i)y(i)$, $i \leq k$
m	dimension of observation vector y
$m(k)$	discrete time wide sense martingale
m_n	discrete time innovation process
$\hat{m}_n$	discrete time linear innovation process
M_n	discrete time innovation process
$\hat{M}_n$	discrete time linear innovation process
$m(t)$	continuous time innovation process
$\hat{m}(t)$	continuous time linear innovation process
$M(t)$	continuous time innovation process
$\hat{M}(t)$	continuous time linear innovation process
n	dimension of state vector x
$n(k)$	discrete time white noise
$P_m(k)$	covariance matrix of $m(k)$
$P_{\hat{m}}(t)$	covariance matrix of $\hat{m}(t)$
$P_u(k)$	covariance matrix of $u(k)$
$P_u(t)$	covariance matrix of $u(t)$
$P_v(k)$	covariance matrix of $v(k)$
$P_v(t)$	covariance matrix of $v(t)$
$P_w(k)$	covariance matrix of $w(k)$
$P_w(t)$	covariance matrix of $w(t)$
$P_{\tilde{u}}(k\|\ell)$	covariance matrix of $\tilde{u}(k\|\ell)$
$P_{\tilde{u}}(t\|s)$	covariance matrix of $\tilde{u}(t\|s)$
$P_{\tilde{x}}(k\|\ell)$	covariance matrix of $\tilde{x}\,(k\|\ell)$
$P_{\tilde{x}}(t\|s)$	covariance matrix of $\tilde{x}\,(t\|s)$
$P_{\tilde{y}}(k\|k-1)$	expression (4.48)
P^o	covariance matrix of x^o
$Q(k)$	$P_u(k+1)-P_u(k)$
r	dimension of system noise vector w
$R(t)$	$D(t)D(t)^T$
s	dimension of observation noise vector v

s	symbol for time
$S(k)$	correlation between martingale increment and observation noise
t	time
t_k	discrete points of time $(k = 0,1,2,\ldots)$
T	time interval or set of points of time
T(superscript)	transpose of a vector or matrix
$u(k)$	discrete time wide-sense martingale
$\hat{u}(k\|\ell)$	linear least-squares estimate of $u(k)$ given observations $y(i), i\leqslant\ell$
$\tilde{u}(k\|\ell)$	$u(k)-\hat{u}(k\|\ell)$
u_k	martingale difference sequence wrt $\mathcal{F}(k)$
$u(t)$	continuous time wide-sense martingale
$\hat{u}(t\|s)$	linear least-squares estimate of $u(t)$ given observations $y(\tau), \tau\leqslant s$
$\tilde{u}(t\|s)$	$u(t)-\hat{u}(t\|s)$
$U(t)$	martingale wrt $\mathcal{F}(t)$
u^o	initial wide-sense martingale u
$v(k)$	discrete time observation noise
$v(t)$	continuous time observation noise
$w(k)$	discrete time system noise
$w(t)$	continuous time system noise
$W(t)$	Wiener process or Brownian motion
$x(k)$	discrete time state process
$\hat{x}(k\|\ell)$	least-squares estimate of $x(k)$ given observations $y(i), i\leqslant\ell$
$\tilde{x}(k\|\ell)$	$x(k)-\hat{x}(k\|\ell)$
x_n	arbitrary random variables, $n = 0,1,2,\ldots$
$x(t)$	continuous time state process
$\hat{x}(t\|s)$	least-squares estimate of $x(t)$ given observations $y(\tau), \tau\leqslant s$
$\tilde{x}(t\|s)$	$x(t)-\hat{x}(t\|s)$
x^o	initial state x
$y(k)$	discrete time observation process
y_n	arbitrary random variables, $n = 0,1,2,\ldots$
$y(t)$	continuous time observation process
$Y(t)$	continuous time observation process

z_k	discrete time stochastic process uncorrelated wrt $\mathcal{G}_k$
$Z(t)$	continuous time process with uncorrelated increments wrt $\mathcal{G}(t)$
$Z_o(t)$	martingale wrt $\mathcal{F}(t)$
$Z_s(t)$	martingale wrt $\mathcal{F}(t)$

2) Greek and cursive characters

α_k	coefficient in representation 3.24
$\alpha(t)$	kernel in representation 3.23
β_k	coefficient in representation 3.22
$\beta(t)$	kernel in representation 3.21
$\mathcal{B}_k$	arbitrary information field (σ-algebra)
$\mathcal{B}(t)$	arbitrary information field (σ-algebra)
$\gamma(t)$	signal process related to state process (3.38)
$\hat{\gamma}(t\|t)$	filtered least-squares estimate of $\gamma(t)$
δ	Dirac δ-function
$\mathcal{E}\{\cdot\}$	expectation of random variable between accolades
$\mathcal{E}\{x(t)\|\mathcal{B}(s)\}$	conditional mean of $x(t)$ wrt information field $\mathcal{B}(s)$ (σ-algebra)
$\mathcal{E}\{x(k)\|y(i),\ i\leqslant\ell\}$	conditional mean of $x(k)$ wrt information field (σ-algebra) generated by $y(i)$, $i\leqslant\ell$
$\mathcal{E}\{x(t)\|y(\tau),\ \tau\leqslant s\}$	conditional mean of $x(t)$ wrt information field (σ-algebra) generated by $y(\tau)$, $\tau\leqslant s$
$\mathcal{E}\{(x(t)-\mathcal{E}\{x(t)\|\mathcal{B}(s)\})^2\|\mathcal{B}(s)\}$	conditional variance of $x(t)$ wrt information field $\mathcal{B}(s)$
$\hat{\mathcal{E}}\{x(k)\|y(i),\ i\leqslant\ell\}$	wide-sense conditional mean (or linear least-squares estimate) of $x(k)$ given observations $y(i)$, $i\leqslant\ell$
$\hat{\mathcal{E}}\{x(t)\|y(\tau),\ \tau\leqslant s\}$	wide-sense conditional mean (or linear least-squares estimate) of $x(t)$ given observations $y(\tau)$, $\tau\leqslant s$
$\mathcal{F}_k$	information field generated by discrete time observation process $y(i)$, $i\leqslant k$
$\mathcal{F}(t)$	information field generated by continuous time observation process $y(s)$, $s\leqslant t$

$\mathcal{G}_k$	all linear combinations of observations $y(i)$, $i \leq k$
$\mathcal{G}(t)$	all linear combinations of observations $Y(s)$, $s \leq t$
$\Lambda(k,o)$	transition matrix
ρ	correlation coefficient
σ	standard deviation
$\phi(k)$	coefficient in state description (4.6)
$\phi(k,o)$	transition matrix
$\phi(y_1, \ldots, y_n)$	measurable function of $y_1, \ldots, y_n$
$\phi(t)$	coefficient in state description (4.55)
$\phi(t,o)$	transition matrix
ω	point in Ω
Ω	outcome space

3) Miscellaneous symbols

$(\Omega, \mathcal{B}, P)$	probability space
$\|\cdot\|$	norm in Hilbert space $L_2^n(\Omega)$
$(z \mid L_2(x;t))$	projection of z ($\in L_2^n(\Omega)$) on subspace $L_2(x;t)$
$x \sim N(\mu, \sigma^2)$	Gaussian random variable with mean μ and variance σ^2
-1 (as superscript)	inverse of a matrix or function
$\sigma\{y(\tau), \tau \leq t\}$	σ-algebra generated by $y(\tau)$, $\tau \leq t$

1 INTRODUCTION

At the National Aerospace Laboratory NLR the research in optimal estimation is mainly directed to practical applications in aero- and astronautics. Problems as the estimation of the attitude of a spacecraft in the presence of modelling errors using the Kalman filter have been investigated in the past and have led, among others, to the development of a powerful adaptive algorithm to prevent filter divergence. (Refs 35, 36.) From an engineering point of view, it is of course more important to have insight in the operation of the Kalman filter under practical circumstances where often modelling errors occur than to master all the various ways along which the Kalman filter has been derived.

On the other hand, however, it has been stressed in the literature of the past decade that more general and satisfying treatments of the optimal estimation problem can be given using martingale theory. First, martingale theory has been applied to the Kalman and Kalman-Bucy problem where the system disturbances are modelled by Gaussian white noises (references 33, 2, 3, 10, 24, 30, 33). In reference 24 for instance it is stated that new optimal recursive estimation equations have been derived, different from the Kalman and Kalman-Bucy results, but simpler and computationally more efficient. Secondly, martingale theory has been applied to more general estimation problems, i.e. to those where the observations are counting processes and where the system disturbances are martingales (Refs 5, 31, 32, 34, 37).

It was deemed necessary and justifiable, therefore, to study the application of martingale theory to optimal estimation problems and to consider in particular the implications for the current practical filtering work at NLR. Because of the latter aspect this report discusses only sideways the more general problems referred to above.

In chapter two the underlying mathematical concepts are briefly summarized. Also, for later reference, the Kalman and Kalman-Bucy filters are incorporated.

Martingales are stochastic processes with certain specific properties which are particularly convenient in cases where new information is coming in continually. The precise mathematical definition is given in section one of chapter three together with some examples. The only assumption on the process, aside from the defining condition, is that the process is integrable at any point of time. The remaining sections of chapter three are concerned with the role of martingales in optimal estimation. The crucial role of the innovation process associated with a given observation process is discussed. Some fundamental results are summarized in an innovation and representation theorem and their meaning is elucidated.

In chapter four wide sense martingales are considered, i.e. martingales which are quadratically integrable. For a certain class of processes or signals optimal recursive estimation equations are described. It is shown, using a suitable transformation that the processes studied in the Kalman and Kalman-Bucy problem belong to this class. The new estimation equations applied to the transformed Kalman and Kalman-Bucy problem are seen to be somewhat simpler than the original ones. The case of coloured observation noise is easily incorporated in the wide sense martingale approach. The effectiveness of this approach as regards computation time and implementation is analyzed in chapter five for some model problems.

Finally, in chapter six the main conclusions are summarized.

2 THE GENERAL CONCEPT OF OPTIMAL ESTIMATION

2.1 The estimation problem

In estimation problems one is concerned with a stochastic process $x(t)$ (i.e. a family of random variables $x(t)$) which cannot be observed directly but which is related to another process $y(t)$ which is accessible to direct observation. The processes $x(t)$ and $y(t)$ are called state process and observation process. The general problem is to find at any time t an estimate $\hat{x}(t|s)$ of the state $x(t)$ given the record of observations $y(\tau)$, $\tau \leqslant s$. For $t>s$, $t = s$ and $t<s$ the estimation problem is called prediction, filtering and smoothing. In a discrete time estimation problem both the state and the observation process are given at discrete points of time whereas in a continuous time-estimation problem both are given for a time interval. The estimate $\hat{x}(t|s)$ of $x(t)$ is required to be optimal in least-squares sense, i.e. it is wanted to find the estimate $\hat{x}(t|s)$ as a function of the observation record such that

$$\mathcal{E}\left\{\left(x(t) - \hat{x}(t|s)\right)^2\right\} \tag{2.1}$$

is minimum (the symbol $\mathcal{E}\{x\}$ denotes the expectation of x).

Each new observation y(t) that comes available adds new information to the already existing knowledge. Mathematically this means that there is an increasing family of information fields (σ-algebras) $\mathcal{F}(t)$ i.e. $\mathcal{F}(t_1) \subset \mathcal{F}(t_2)$ for $t_1 < t_2$, generated by the observation process y(t). The state x(t) at time t is generally not completely determined by the observations contained in the information field $\mathcal{F}(s)$. Mathematically this means that x(t) is not measurable-$\mathcal{F}(s)$. However, there exists a kind of projection of x(t) on the information field $\mathcal{F}(s)$ generated by the observation process $y(\tau)$, $\tau \leqslant s$ which is completely determined by the observations, i.e. which is measurable-$\mathcal{F}(s)$. This version of x(t) which in general is a non-linear functional of the set $y(\tau)$, $\tau \leqslant s$, is called the conditional mean of x(t) with respect to the observations and is denoted by the symbol $\mathcal{E}\{x(t)|y(\tau)\ ,\ \tau \leqslant s\}$. (Actually, this is only one version of the conditional mean, the others being equal to this one with probability one.)

The conditional mean concept plays an important role in the definition of martingales. Its importance for least-squares estimation follows from the following property. Let $x = x(\omega)$ be a random variable and $f(\omega)$ an ω-function which is measurable-$\mathcal{F}(s)$ and such that $\mathcal{E}\{|f(\omega)x|\}<\infty$, $\mathcal{E}\{|x|\}<\infty$ holds. Then,

$$\mathcal{E}\{f(\omega)x|\mathcal{F}(s)\} = f(\omega)\mathcal{E}\{x|\mathcal{F}(s)\} \quad . \tag{2.2}$$

From this equality it follows immediately that

$$\mathcal{E}\{(x-\mathcal{E}\{x|\mathcal{F}(s)\})f(\omega)\} = 0 \tag{2.3}$$

i.e. $x-\mathcal{E}\{x|\mathcal{F}(s)\}$ is orthogonal to any function $f(\omega)$ satisfying the assumptions stated above. It follows that

$$\mathcal{E}\{(x-\mathcal{E}\{x|\mathcal{F}(s)\})\ \mathcal{E}\{x|\mathcal{F}(s)\}\} = 0 \quad .$$

Hence, for any f which is $\mathcal{F}(s)$-measurable

$$\begin{aligned}\mathcal{E}\{(x-f)^2\} &= \mathcal{E}\{(x-\mathcal{E}\{x|\mathcal{F}(s)\})^2\} \\ &\quad + \mathcal{E}\{(\mathcal{E}\{x|\mathcal{F}(s)\}-f)^2\} \\ &\quad +2\mathcal{E}\{(x-\mathcal{E}\{x|\mathcal{F}(s)\})(\mathcal{E}\{x|\mathcal{F}(s)\}-f)\} \\ &= \mathcal{E}\{(x-\mathcal{E}\{x|\mathcal{F}(s)\})^2\} + \mathcal{E}\{(\mathcal{E}\{x|\mathcal{F}(s)\}-f)^2\} \quad .\end{aligned} \tag{2.4}$$

This shows that if $\mathcal{F}(s)$ is the information field generated by the observations $y(\tau), \tau \leqslant s$, the conditional mean $\mathcal{E}\{x(t)|\mathcal{F}(s)\} = \mathcal{E}\{x(t)|y(\tau)\ ,\ \tau \leqslant s\}$ is the least-squares estimate of x(t), that is, it minimizes the expression (2.4) over all possible functionals f of the observation record $y(\tau), \tau \leq s$ (Ref. 8). The uncertainty in the estimate of x(t) is characterized by the conditional variance

$$\mathcal{E}\{(x(t)-\mathcal{E}\{x(t)|\mathcal{F}(s)\})^2|\mathcal{F}(s)\} \tag{2.5}$$

which as the conditional mean generally depends non-linearly on the observations $y(\tau)$, $\tau \leq s$. Moreover, both the conditional mean and variance will usually depend on all higher order conditional moments which makes the general (i.e. non-linear) optimal estimation problem very difficult to solve.

A considerable simplification of the estimation problem occurs if it is only required to find the linear least-squares estimate of x(t) given the observations $y(\tau)$, $\tau \leq s$. If both the state and observation process are Gaussian, then the linear least-squares estimate is at the same time the least-squares estimate. For instance, for two Gaussian random variables x_1 and x_2 it follows that the conditional mean of x_1 given x_2 is the following linear function of x_2:

$$\mathcal{E}\{x_1|x_2\} = \mathcal{E}\{x_1\} + \rho_{x_1x_2}\frac{\sigma_{x_1}}{\sigma_{x_2}}\left(x_2-\mathcal{E}\{x_2\}\right) \tag{2.6}$$

where $\rho_{x_1x_2}$ denotes the correlation coefficient between x_1 and x_2 and σ_{x_1} and σ_{x_2} respectively are the standard deviations of x_1 and x_2. Moreover, the conditional variance of x_1 given x_2 is independent of x_2 and, hence, deterministic:

$$\mathcal{E}\left\{\left(x_1-\mathcal{E}\{x_1|x_2\}\right)^2\Big|x_2\right\} = \sigma^2_{x_1}\left(1-\rho^2_{x_1x_2}\right) \quad . \tag{2.7}$$

If x_1 and x_2 are non Gaussian, then the right-hand side of equation (2.6) gives the linear least-squares estimate of x_1 given x_2 and is sometimes called the wide sense conditional mean of x_1 given x_2. From (2.6) it is clear that if x_1 and x_2 are uncorrelated then the linear least-squares estimate of x_1 cannot be improved by observing x_2. The preceeding discussion can easily be generalized for the state and observations being vectors.

2.2 The Kalman and Kalman-Bucy filter

In this section the Kalman filter for the discrete time estimation problem and the Kalman-Bucy filter for the continuous time estimation problem are recalled. These filters yield recursive equations for computing $\hat{x}(t|t)$ and the associated covariance matrix $P_{\tilde{x}}(t|t) = \mathcal{E}\{(x(t)-\hat{x}(t|t))\ (x(t)-\hat{x}(t|t))^T\}$ and will be referred to in later chapters.

For the Kalman filter it is assumed that the state process $x(k)$(nx1-vector) is defined at discrete points of time t_k and satisfies a linear difference equation driven by Gaussian white noise $w(k)$ (rx1-vector), that is

$$x(k+1) = A(k+1,k)x(k) + B(k)w(k) \tag{2.8}$$

where $w(k) \sim N(o,P_w(k))$ and the initial condition $x(o)$ is $x(o) \sim N(x^o,P^o)$. The observation process $y(k)$ (mx1-vector, $m \leq n$) is described by

$$y(k) = C(k)x(k) + v(k) \tag{2.9}$$

where $v(k)$(sx1-vector) is another Gaussian white noise, $v(k) \sim N(o,P_v(k))$. Furthermore it is assumed that the initial condition x^o and the noise sequences $w(k)$ and $v(k)$ are independent. $A(k+1,k)$, $B(k)$ and $C(k)$ are appropriately dimensional matrices. From the above assumptions it follows that the state and observation processes are Gaussian for all discrete points of time t_k. Hence, the best estimate $\hat{x}(k|\ell)$ in least-squares sense of $x(k)$ given the observations $y(i)$, $i \leq \ell$ is a linear functional of the observations. The filtering solution $\hat{x}(k+1|k+1)$ is found from the measurement update equation:

$$\hat{x}(k+1|k+1) = \hat{x}(k+1|k) + K(k+1)\left(y(k+1)-C(k+1)\hat{x}(k+1|k)\right) \tag{2.10}$$

where the predicted estimate $\hat{x}(k+1|k)$ of $x(k+1)$ given the observations $y(i)$, $i \leq k$ is

$$\hat{x}(k+1|k) = A(k+1,k)\hat{x}(k|k) \tag{2.11}$$

and the Kalman gain matrix $K(k+1)$ is

$$K(k+1) = P_{\tilde{x}}(k+1|k)\ C(k+1)^T \left[C(k+1)P_{\tilde{x}}(k+1|k)\ C(k+1)^T + P_v(k+1)\right]^{-1} . \tag{2.12}$$

The predicted covariance matrix $P_{\tilde{x}}(k+1|k)$ is

$$P_{\tilde{x}}(k+1|k) = A(k+1,k)\ P_{\tilde{x}}(k|k)\ A(k+1,k)^T + B(k)\ P_w(k)\ B(k)^T \tag{2.13}$$

and the covariance matrix $P_{\tilde{x}}(k+1|k+1)$ associated with the filtered estimate $\hat{x}(k+1|k+1)$ satisfies

$$P_{\tilde{x}}(k+1|k+1) = P_{\tilde{x}}(k+1|k) - K(k+1)\ C(k+1)\ P_{\tilde{x}}(k+1|k)\ . \qquad (2.14)$$

Equations (2.10), ..., (2.14) constitute the Kalman filter for the discrete time-filtering problem (2.8) and (2.9).

For the Kalman-Bucy filter it is assumed that the state process $x(t)$ (nx1-vector) is defined for a time interval $0 \leqslant t \leqslant 1$, say, and satisfies a linear differential equation

$$\frac{dx(t)}{dt} = A(t)a(t) + B(t)w(t) \qquad (2.15)$$

where $w(t)$ (rx1-vector) in a Gaussian white noise, $w(t) \sim N(o, P_w(t))$ and the initial condition $x(0)$ is $x(o) \sim N(x^o, P^o)$. The observation process $y(t)$ (mx1-vector) is described by

$$y(t) = C(t)x(t) + v(t) \qquad (2.16)$$

where $v(t)$ (sx1-vector) is another Gaussian white noise $v(t) \sim N(o, P_v(t))$. The initial condition $x(o)$ and the noises $w(t)$ and $v(t)$ are assumed to be independent. It follows again under the above assumptions that the state and observation processes are Gaussian for all t, $0 \leqslant t \leqslant 1$. The filtering solution $\hat{x}(t|t)$ and the covariance matrix $P_{\tilde{x}}(t|t)$ follow from the Kalman-Bucy filter described by equations (2.17), ..., (2.19) below:

$$\frac{d\hat{x}(t|t)}{dt} = A(t)\hat{x}(t|t) + K(t)\{y(t) - C(t)\hat{x}(t|t)\} \qquad (2.17)$$

$$\frac{d}{dt} P_{\tilde{x}}(t|t) = A(t)P_{\tilde{x}}(t|t) + P_{\tilde{x}}(t|t)A(t)^T + B(t)P_w(t)B(t)^T +$$
$$- P_{\tilde{x}}(t|t)C(t)^T P_v^{-1}(t)C(t)P_{\tilde{x}}(t|t) \qquad (2.18)$$

where the Kalman gain matrix is given by

$$K(t) = P_{\tilde{x}}(t|t)\ C(t)^T\ P_v^{-1}(t)\ . \qquad (2.19)$$

Though the filters described above for discrete and continuous time look quite the same, their original derivations are quite different. One of the aspects of the approaches described in the following chapters is a more parallel treatment of the continuous and discrete time filtering problem.

The mixed problem of a state process $x(t)$ satisfying a differential equation (2.15) with observations given at discrete points of time like (2.9) can easily be embedded in the discrete time problem. The model problems treated in chapter five are of the mixed or continuous discrete type.

3 MARTINGALES AND ESTIMATION THEORY

In this chapter martingales are considered in relation to optimal estimation theory. The purpose is to describe how martingale theory is used to derive recursive optimal estimation equations.

Martingales are defined in section one and a few examples of stochastic processes being martingales are given there. An important concept in estimation theory which is discussed in section two, is the innovation process associated with a given process. As will be seen, the innovation process represents the new information coming in with each new realization of the process at hand. In section three a number of results is summarized in the form of innovation and representation theorems. For example, it is stated that the (non-linear) innovation process is a martingale. The remaining sections are concerned with a further study in depth.

3.1 Martingales

A martingale is a stochastic process $x(t)$, $t\in T$ with a specific property for the conditional expectation with respect to a given family $\mathcal{B}(t)$, $t\in T$ of information fields. $\mathcal{B}(t)$ denotes the collection of events representing the information available at time t and is in a particular problem usually the collection of events generated by one or more of the stochastic processes involved up to time t (mathematically, $\mathcal{B}(t)$ is a σ-algebra of events). Let the family $\mathcal{B}(t)$, $t\in T$ be increasing in the sense that $\mathcal{B}(t_1)\subset\mathcal{B}(t_2)$ whenever $t_1<t_2$. The process $x(t)$, $t\in T$ is called a martingale with respect to $\mathcal{B}(t)$, $t\in T$ or a $\mathcal{B}(t)$ martingale if for each t $x(t)$ is $\mathcal{B}(t)$ measurable, $\mathcal{E}\{|x(t)|\}<\infty$ and

$$\mathcal{E}\{x(t+dt) - x(t)|\mathcal{B}(t)\} = 0 \quad dt>0 \quad . \tag{3.1}$$

Because of the assumption that $x(t)$ is measurable $\mathcal{B}(t)$, i.e. $x(t)$ is completely determined by $\mathcal{B}(t)$ for every t, it follows that

$\mathcal{E}\{x(t)|\mathcal{B}(t)\} = x(t)$ and equation (3.1) may be written then as

$$\mathcal{E}\{x(t+dt)|\mathcal{B}(t)\} = x(t) \qquad dt>0 \tag{3.2}$$

for every t. If the family of information fields $\mathcal{B}(t)$, $t\in T$ is that generated by the stochastic process $x(t)$, $t\in T$ itself, then $x(t)$ is simply said to be a martingale. It follows that if $x(t)$ is a $\mathcal{B}(t)$ martingale then it is a martingale because

$$\begin{aligned}\mathcal{E}\{x(t) - x(s)|x(\tau),\ \tau\leqslant s\} &= \mathcal{E}\{\mathcal{E}\{x(t) - x(s)|\mathcal{B}(s)\}|x(\tau),\ \tau\leqslant s\} \\ &= \mathcal{E}\{o|x(\tau),\ \tau\leqslant s\} = 0 \qquad t\geqslant s \ .\end{aligned} \tag{3.3}$$

It follows from $\mathcal{E}\{\mathcal{E}\{x|\mathcal{B}\}\} = \mathcal{E}\{x\}$ that the mean value function $\mathcal{E}\{x(t)\}$ of any martingale is a constant. Moreover, the martingale increment $x(t) - x(s)$ is orthogonal to $x(\tau)$, $\tau\leqslant s$ i.e. $\mathcal{E}\{(x(t) - x(s))x(\tau)\} = 0$, $\tau\leqslant s$ and $\mathcal{E}\{(x(t) - x(s))\,(x(\tau) - x(\sigma))\} = 0$, $\sigma\leqslant\tau\leqslant s\leqslant t$. Therefore, a martingale has uncorrelated increments. From an estimation point of view it is important to note that the least-squares estimate of a martingale $x(t)$ given the information field $\mathcal{B}(s)$ $s<t$ is the same as the least-squares estimate of $x(s)$ given the information contained in $\mathcal{B}(s)$. Thus, the prediction problem is trivial for a martingale.

The following examples show how to handle definition (3.1). Suppose that a sequence of random variables x_n is given which can be written as $x_n = \sum_{i=1}^{n} y_i$ where the sequence y_i has the properties

$$\mathcal{E}\{|y_i|\}<\infty \ , \ \mathcal{E}\{y_i|y_1, \ldots, y_{i-1}\} = 0 \ . \tag{3.4}$$

The sequence x_n is then a martingale because

$$\begin{aligned}&\mathcal{E}\{x_{n+1}|x_1, \ldots, x_n\} = \mathcal{E}\{y_{n+1} + x_n|x_1, \ldots, x_n\} \\ &= \mathcal{E}\{y_{n+1}|y_1, \ldots, y_n\} + \mathcal{E}\{x_n|x_1, \ldots, x_n\} = x_n \ .\end{aligned} \tag{3.5}$$

Conversely, if x_n is a martingale, then the difference sequence y_n defined by

$$\begin{aligned}y_1 &= x_1 \\ y_n &= x_n - x_{n-1} \qquad n>1\end{aligned} \tag{3.6}$$

satisfies (3.4). A sequence y_i satisfying (3.4) has been called a martingale difference sequence (MD sequence, Refs 2, 31). Hence, a sequence x_n is a martingale if and only if the difference sequence y_n defined by (3.6) is a MD sequence. Property (3.4) is intermediate between independence and uncorrelatedness of the y_i's because it expresses that y_i $i>1$ is orthogonal to every (Baire) function $\phi(y_1, \ldots, y_{i-1})$ of $y_1, \ldots, y_{i-1}$ (compare chapter 2.1 equation (2.3)). Property (3.4) expresses that the least-squares estimate of y_i given $y_1, \ldots, y_{i-1}$ does not depend on $y_1, \ldots, y_{i-1}$ i.e. observing $y_1, \ldots, y_{i-1}$ does not help to estimate y_i. On the other hand, if the y_i's are independent and zero mean then (3.4) holds.

Suppose $\mathcal{B}_n$ is an increasing sequence of information fields (σ-algebras) and y is a random variable. Define a sequence x_n by $x_n = \mathcal{E}\{y|\mathcal{B}_n\}$. Then, x_n is a $\mathcal{B}_n$ martingale because for $n \geqslant m$

$$\begin{aligned}\mathcal{E}\{x_n|\mathcal{B}_m\} &= \mathcal{E}\{\mathcal{E}\{y|\mathcal{B}_n\}|\mathcal{B}_m\} \\ &= \mathcal{E}\{y|\mathcal{B}_m\} = x_m \quad .\end{aligned} \tag{3.7}$$

Let $W(t)$, $t \geqslant 0$ be a Brownian motion or Wiener process. Then, $W(t)$, $t \geqslant 0$ has independent increments, zero mean and is a martingale because for $t \geqslant s$:

$$\begin{aligned}\mathcal{E}\{W(t)|W(\tau),\tau\leqslant s\} &= \mathcal{E}\{W(t) - W(s) + W(s)|W(\tau),\tau\leqslant s\} \\ &= \mathcal{E}\{W(t)-W(s)\} + W(s) = W(s) \quad .\end{aligned} \tag{3.8}$$

More generally, any constant mean independent increment process constitutes a martingale.

As a final remark it is noted that a Markov process does not need to be a martingale (the Brownian motion is both a Markov process and a martingale). For a Markov process holds

$$\mathcal{E}\{x(t)|x(\tau),\tau\leqslant s\} = \mathcal{E}\{x(t)|x(s)\} \qquad t \geqslant s \tag{3.9}$$

which needs not be the same as $x(s)$.

3.2 Martingale decomposition and innovation processes

If a process $x(t)$, $t \in T$ and an arbitrary (increasing) family of information fields (σ-algebras) $\mathcal{B}(t)$, $t \in T$ are given, martingale decomposition means that the $x(t)$ process is written as the sum of two processes, one of which is a martingale. To see this, write $dx(t) = dB(t) + dM(t)$ where $dB(t)$ is defined by $dB(t) = \mathcal{E}\{dx(t)|\mathcal{B}(t)\}$. Then

$$\begin{aligned}\mathcal{E}\{d\,M(t)|\mathcal{B}(t)\} &= \mathcal{E}\{dx(t) - dB(t)|\mathcal{B}(t)\} \\ &= \mathcal{E}\{dx(t)|\mathcal{B}(t)\} - \mathcal{E}\{\mathcal{E}\{dx(t)|\mathcal{B}(t)\}|\mathcal{B}(t)\} \\ &= \mathcal{E}\{dx(t)|\mathcal{B}(t)\} - \mathcal{E}\{dx(t)|\mathcal{B}(t)\} = 0 \quad .\end{aligned} \tag{3.10}$$

It is exactly by choosing $\mathcal{B}(t)$ to be the information field generated by the process $x(t)$ itself that the martingale $M(t)$ corresponding with $x(t)$ is called the innovation process of $x(t)$. This is because

$$dM(t) = dx(t) - \mathcal{E}\{dx(t)|x(\tau), \tau \leqslant t\}$$

clearly represents the new information in $dx(t)$ not already known at time t. In estimation theory, the innovation process associated with the observation process plays an important role.

As an illustration, let the discrete time process x_n defined by $x_n = \sum_{i=1}^{n} y_i$ and the continuous time process $x(t) = \int_o^t y(s)ds$ be given. It is wanted to determine the innovation process of x_n and of $x(t)$ in the sense just described.

Suppose an arbitrary sequence y_n is given. To start with, the sequence M_n defined by

$$M_n = x_n - \sum_{k=1}^{n} \mathcal{E}\{y_k|y_1, \ldots, y_{k-1}\} \tag{3.11}$$

is a martingale with respect to x_n and is called the innovation process of x_n. Also the difference sequence m_n

$$\begin{aligned} m_n &= M_n - M_{n-1} & n \geqslant 2 \\ &= y_n - \mathcal{E}\{y_n|y_1, \ldots, y_{n-1}\} \\ m_1 &= M_1 \end{aligned} \tag{3.12}$$

is a martingale difference (MD) sequence and is called the innovation process of the sequence y_n. Equation (3.12) expresses that the process y_n consists of a predictable part $\mathcal{E}\{y_n|y_1, \ldots, y_{n-1}\}$ completely known at time t_{n-1} and a MD sequence which is completely unpredictable from knowing $y_1, \ldots, y_{n-1}$. Similarly, the process $M(t)$ defined by

$$M(t) = x(t) - \int_o^t \mathcal{E}\{y(s)|y(\tau), \tau \leqslant s\} \; ds \tag{3.13}$$

is an $x(t)$ martingale and is called the innovation process of $x(t)$. The innovation processes described so far are based on the conditional mean with respect to a given information field. The conditional mean or least-squares estimate is, in general, a non-linear functional of the elements of the information field. Analogously, the linear innovation process is defined on the basis of the linear least-squares estimate or wide-sense conditional mean. Let $\hat{\mathcal{E}}\{y_n|y_1, \ldots, y_{n-1}\}$ denote the wide-sense conditional mean of y_n given $y_1, \ldots, y_{n-1}$. The linear innovation process $\hat{M}_n$ of the process $x_n = \sum_{i=1}^{n} y_n$ is defined by

$$\hat{M}_n = x_n - \sum_{k=1}^{n} \hat{\mathcal{E}}\{y_k | y_1, \ldots, y_{k-1}\} \quad . \tag{3.14}$$

It follows (compare the next section) that the linear innovation process $\hat{M}_n$ is a process with uncorrelated increments. The difference sequence $\hat{m}_n$

$$\begin{aligned} \hat{m}_n &= \hat{M}_n - \hat{M}_{n-1} = y_n - \hat{\mathcal{E}}\{y_n | y_1, \ldots, y_{n-1}\} \qquad n \geqslant 2 \\ \hat{m}_1 &= \hat{M}_1 \end{aligned} \tag{3.15}$$

is a sequence of uncorrelated random variables and is called the innovation process of y_n. Also, the linear innovation process $\hat{M}(t)$ of $x(t) = \int_0^t y(s)ds$ is defined by

$$\hat{M}(t) = x(t) - \int_0^t \hat{\mathcal{E}}\{y(s) | y(\tau), \tau \leqslant s\} \; ds \tag{3.16}$$

and has uncorrelated increments.

3.3 Innovation and representation theorems

In this section a number of results are stated which illustrate the role of martingales in estimation theory. These results are called innovation and representation theorems and apply, partly, for both continuous and discrete time, linear and non-linear estimation. They express that the observation process is equivalent to the innovation process for which, due to its simpler structure, useful representations hold. Suppose that the continuous time observation, innovation and linear innovation processes are $Y(t)$, $M(t)$ and $\hat{M}(t)$. Let the corresponding processes for discrete time be denoted by y_i, m_i and $\hat{m}_i$. (It is assumed that the continuous time observations are in an integral form, like $Y(t) = \int_0^t f(s)ds$. The continuous time innovations are defined then in the sense of (3.13) and (3.16). For discrete time the observations and innovations are taken in the sense of (3.12) and (3.15).)

The innovation theorem states this:

a) The innovation process is a martingale that is equivalent to the observation process, i.e. the information field generated by the innovations is the same as that generated by the observations,

$$\mathcal{F}(t) \equiv \sigma\{Y(s),\ s\leqslant t\} = \sigma\{M(s),\ s\leqslant t\} \tag{3.17}$$

and for discrete time

$$\mathcal{F}_k \equiv \sigma\{y_j,\ j\leqslant k\} = \sigma\{m_j,\ j\leqslant k\} \quad . \tag{3.18}$$

a) The linear innovation process has uncorrelated increments and is linearly equivalent to the observations, i.e.

$$\begin{aligned} \mathcal{G}(t) &\equiv \{\text{all linear combinations of } Y(s),\ s\leqslant t\} \\ &= \{\text{all linear combinations of } \hat{M}(s),\ s\leqslant t\} \end{aligned} \tag{3.19}$$

and for discrete time

$$\begin{aligned} \mathcal{G}_k &\equiv \{\text{all linear combinations of } y_j,\ j\leqslant k\} \\ &= \{\text{all linear combinations of } \hat{m}_j,\ j\leqslant k\} \end{aligned} \tag{3.20}$$

The innovation theorem holds for all discrete time cases and for linear observations in all continuous time cases (Refs 12, 17). Under certain conditions on the observation process it also holds for continuous time non-linear innovations for signals in Gaussian white noise. This is discussed more extensively in section 3.5. The innovation theorem also holds for continuous time non-linear innovations for point processes (Refs 31, 33). Point processes are not discussed further (see references 5, 31, 32, 34, 37).

The representation theorem reads as follows:

a) Every process $Z(t)$, with uncorrelated increments wrt $\mathcal{G}(t)$ (3.19) can be written as a Wiener integral of the linear innovations $\hat{M}(t)$,

$$Z(t) = \int_0^t \beta(s)d\hat{M}(s) \tag{3.21}$$

(where the kernel β is deterministic) and for discrete time, every uncorrelated sequence z_k wrt $\mathcal{G}_k$ (3.20) can be written as

$$z_k = \beta_k\ \hat{m}_k \tag{3.22}$$

with a deterministic sequence β_k.

b) Every martingale $U(t)$, wrt $\mathcal{F}(t)$ (3.17) can be written as a stochastic integral of the innovations $M(t)$,

$$U(t) = \int_0^t \alpha(s)dM(s) \tag{3.23}$$

where $\alpha(t)$ is a $\mathcal{F}(t)$ predictable process and for discrete time every martingale difference sequence u_k wrt $\mathcal{F}_k$ can be written as

$$u_k = \alpha_k m_k \qquad (3.24)$$

with an $\mathcal{F}_k$ predictable sequence α_k.
(Note: The sequence α_k is called $\mathcal{F}_k$ predictable if for every k α_k belongs to $\mathcal{F}_{k-1}$. A similar definition holds for continuous time.)

The representation theorem holds for all linear cases in discrete and continuous time. It also holds for the continuous time non-linear case for both signals in Gaussian white noise and signals observed through point processes. Finally it holds for the discrete time non-linear case for point processes but not for the discrete time non-linear case for signals in Gaussian white noise.

3.4 Application to linear discrete and continuous estimation

In this section it is shown how the innovation and representation theorem can be used to derive recursive estimation equations in the linear case.

Consider first the linear discrete time model of chapter 2.2

$$x(k+1) = A(k+1,k)x(k) + B(k)w(k) \qquad (2.8)$$

$$y(k) = C(k)x(k) + v(k) \qquad (2.9)$$

where the initial condition x^o and the white noise sequences v(k) and w(k) are assumed to be uncorrelated. Denote by $\hat{x}(k|\ell)$ the linear least-squares estimate of x(k) given the observations y(1), ..., y(ℓ). Let the sequence of linear spaces $\mathcal{G}_k$ be defined as in (3.20) and consider the sequence of random variables z(k) where

$$\begin{aligned} z(k+1) &= \hat{x}(k+1|k+1) - A(k+1,k)\hat{x}(k|k) \\ &= \hat{x}(k+1|k+1) - \hat{x}(k+1|k) \quad . \end{aligned} \qquad (3.25)$$

To show that the sequence z(k+1) is an uncorrelated sequence wrt the sequence of linear spaces $\mathcal{G}_{k+1}$ it must be demonstrated that z(k+1) belongs to $\mathcal{G}_{k+1}$ and moreover that it is uncorrelated with the elements of $\mathcal{G}_k$, i.e. with y(j) for j$\leq$k. It is clear that z(k+1) is an element of $\mathcal{G}_{k+1}$ and $\mathcal{E}\{z(k+1)\} = 0$. It follows that z(k+1) is uncorrelated with all elements of $\mathcal{G}_k$ because

$$\mathcal{E}\left\{z(k+1)y(j)^T\right\} = -\mathcal{E}\left\{\Big(x(k+1) - \hat{x}(k+1|k+1)\Big)y(j)^T\right\} +$$
$$+ \mathcal{E}\left\{\Big(x(k+1) - A(k+1,k)\hat{x}(k|k)\Big)y(j)^T\right\}$$
$$= A(k+1,k)\mathcal{E}\left\{\Big(x(k) - \hat{x}(k|k)\Big)y(j)^T\right\} + B(k)\mathcal{E}\left\{w(k)y(j)^T\right\}$$
$$= 0 \quad \text{for} \quad j \leq k \quad . \tag{3.26}$$

Therefore, according to the representation theorem the sequence $z(k+1)$ can be written as

$$z(k+1) = K(k+1)\hat{m}(k+1) \tag{3.27}$$

where the sequence $K(k+1)$ is deterministic and the linear innovation sequence $\hat{m}(k+1)$ is

$$\hat{m}(k+1) = y(k+1) - C(k+1)\hat{x}(k+1|k) \quad . \tag{3.28}$$

The gain $K(k+1)$ can be solved from (3.27) as

$$K(k+1) = \frac{\mathcal{E}\left\{z(k+1)\hat{m}(k+1)^T\right\}}{\mathcal{E}\left\{\hat{m}(k+1)\hat{m}(k+1)^T\right\}} \quad . \tag{3.29}$$

Combining the results so far it is found that

$$\hat{x}(k+1|k+1) = \hat{x}(k+1|k) + K(k+1)\{y(k+1) - C(k+1)\hat{x}(k+1|k)\} \tag{3.30}$$

where the gain $K(k+1)$ can be derived from (3.29) using the system description. Elaboration of this yields exactly the Kalman gain already given in chapter 2.2:

$$K(k+1) = P_{\tilde{x}}(k+1|k)C(k+1)^T \left[C(k+1)P_{\tilde{x}}(k+1|k)C(k+1)^T + P_v(k+1)\right]^{-1} \tag{2.12}$$

where the predicted covariance matrix $P_{\tilde{x}}(k+1|k)$ satisfies equation (2.13).

Consider now the linear continuous time problem

$$\frac{dx(t)}{dt} = A(t)x(t) + B(t)w(t) \tag{2.15}$$

$$y(t) = C(t)x(t) + v(t) \tag{2.16}$$

where the noises $w(t)$ and $v(t)$ are white and Gaussian and uncorrelated as described in chapter 2.2. The description (2.15) and (2.16) is not the rigorous integral form for the continuous time estimation problem. Therefore, the innovation and representation theorem as described in section 3.3 cannot be applied directly. An intuitive derivation quite similar to the discrete case treated above may be given, however. A rigorous treatment is described in section 3.6. The linear least-squares estimate $\hat{x}(t|t)$ given the observations $y(s)$, $s \leq t$ is expressed as

$$x(t|t) = \int_0^t g(t,\tau)y(\tau)d\tau \tag{3.31}$$

and the linear innovation process $\hat{m}(t)$ of $y(t)$ is defined by

$$\hat{m}(t) = y(t) - C(t)\hat{x}(t|t) \quad . \tag{3.32}$$

It can be shown that the linear innovation process $\hat{m}(t)$ is equivalent to the observations $y(t)$ and moreover that it is a Gaussian white noise (derivative of a process with orthogonal increments). Therefore, the representation (3.31) can be expressed equivalently in terms of the innovation process $\hat{m}(t)$ as

$$\hat{x}(t|t) = \int_0^t h(t,\tau)\hat{m}(\tau)d\tau \quad . \tag{3.33}$$

The step from (3.31) to (3.33) is the essential one. Because the linear innovation process $\hat{m}(t)$ is a white noise, the kernel $h(t,\tau)$ can be determined explicitely using the definition of the least-squares estimate $\hat{x}(t|t)$, i.e.

$$\mathcal{E}\left\{\left(x(t) - \hat{x}(t|t)\right)\hat{m}(s)^T\right\} = 0 \qquad s \leq t \quad . \tag{3.34}$$

Substitution of the representation (3.33) of $\hat{x}(t|t)$ in (3.34) yields

$$\begin{aligned}\mathcal{E}\left\{x(t)\hat{m}(s)^T\right\} &= \int_0^t h(t,\tau)\mathcal{E}\left\{\hat{m}(\tau)\hat{m}(s)^T\right\}d\tau \\ &= h(t,s)P_{\hat{m}}(s)\end{aligned} \tag{3.35}$$

where $P_{\hat{m}}(s)$ is given by $\mathcal{E}\left\{\hat{m}(s)\hat{m}(s)^T\right\} = P_{\hat{m}}(s)\ \delta(o)$. The latter equality in (3.35) follows from the whiteness of the linear innovation process $\hat{m}(t)$. It has been shown (Ref. 12) that the linear innovation process $\hat{m}(t)$ is not only a white noise but also has the same covariance as the observation noise, i.e. $P_{\hat{m}}(s) = P_v(s)$. Therefore, the least-squares estimate $\hat{x}(t|t)$ can be written as

$$\hat{x}(t|t) = \int_0^t \mathcal{E}\left\{x(t)\hat{m}(\tau)^T\right\} P_v^{-1}(\tau)\hat{m}(\tau)d\tau$$

or

$$\hat{x}(t|t) = \int_0^t \mathcal{E}\left\{x(t)\hat{m}(\tau)^T\right\} P_v^{-1}(\tau) \Big(y(\tau)-C(\tau)\hat{x}(\tau|\tau)\Big)d\tau \qquad (3.36)$$

Notice the similarity between this representation and the discrete time equations (3.30) and (3.29). Differentiating equation (3.36) with respect to t and using the system dynamics it follows that $\hat{x}(t|t)$ satisfies the differential equation

$$\frac{d}{dt}\,\hat{x}(t|t) = A(t)\hat{x}(t|t) + K(t)\{y(t) - C(t)\hat{x}(t|t)\} \qquad (3.37)$$

where $K(t)$ and $P_{\tilde{x}}(t|t)$ satisfy equations (2.19) and (2.18). Hence, equation (3.37) together with (2.18) and (2.19) constitutes exactly the Kalman-Bucy filter.

In this section the derivation of the Kalman and Kalman-Bucy filter using the theory previously described has been sketched. Essentially, it amounts to showing that the observations and innovations are equivalent (linearly). Due to the fact that the innovation process has simpler structure than the original observations, the estimation problem based on the innovations is relatively easy to solve then.

3.5 Application to non-linear estimation

It has been described in section three of this chapter that for the continuous time non-linear case the representation theorem holds and under certain conditions on the observation process also the innovation theorem. The latter states that the innovation process is a martingale and that the information fields (σ-algebras) generated by the innovation process and the observation process are equivalent, i.e. there is no loss of information when the original observations are replaced by the innovations. In this section mainly the innovation approach of Kailath e.a. (Ref. 9) is outlined.

Let the observation process $y(t)$ be given by

$$y(t) = \gamma(t) + v(t) \qquad 0 \leqslant t \leqslant 1 \qquad (3.38)$$

where $\gamma(t)$ and $v(t)$ are independent stochastic processes, $v(t)$ being zero mean Gaussian white noise and $\mathcal{E}\left\{v(t)v(s)^T\right\} = I\ \delta(t-s)$. The signal $\gamma(t)$ is related to the state x by an expression of the form

$$\gamma(t) = C(x(s),\ s\leqslant t)$$

i.e. $\gamma(t)$ is a functional of past and present values of the state. The signal $\gamma(t)$ is assumed to be zero mean, not necessarily Gaussian but with the properties

$$\int_0^1 \mathcal{E}\left\{\gamma(t)^T\gamma(t)\right\} dt<\infty$$

and

$$|\gamma(t)|\leqslant M<\infty \qquad 0\leqslant t\leqslant 1 \quad . \tag{3.39}$$

The innovation process $m(t)$ corresponding with the observation process $y(t)$ is defined by

$$m(t) = y(t) - \hat{\gamma}(t|t) \tag{3.40}$$

where $\hat{\gamma}(t|t)$ denotes the least-squares estimate of $\gamma(t)$ given the observations up to time t (compare equation (3.32)). Under the assumptions stated the equivalence of the innovations and observations has been proved in reference 9. Moreover, it has been proved that the innovation process $m(t)$ is a Gaussian white noise with the same covariance as the observation noise, i.e. $\mathcal{E}\left\{m(t)m(s)^T\right\} = I\ \delta(t-s)$. This result is surprising because if $\gamma(t)$ is non-Gaussian then neither $y(t)$ nor $\hat{\gamma}(t|t)$ is Gaussian. From the equivalence of observations and innovations it follows that the least squares estimate $\hat{x}(t|s)$ of $x(t)$ given the observations is the same as that given the innovations, i.e.

$$\hat{x}(t|s) = \mathcal{E}\{x(t)|y(\tau),\ \tau\leqslant s\} = \mathcal{E}\{x(t)|m(\tau),\ \tau\leqslant s\} \quad . \tag{3.41}$$

By definition, the least-squares estimate $\hat{x}(t|s)$ is orthogonal to every functional measurable with respect to the information field (σ-algebra) generated by the observations or, equivalently, the innovations. Therefore,

$$\mathcal{E}\left\{\left(x(t)-\hat{x}(t|s)\right)f^T\right\} = 0 \tag{3.42}$$

for every functional f measurable with respect to the information field $\mathcal{F}(s) \equiv \sigma\{y(\tau),\ \tau\leqslant s\} = \sigma\{m(\tau),\ \tau\leqslant s\}$. The importance of these steps is that (non-linear) functionals of Gaussian white noise have special representations in the form of stochastic integrals. Applied to $\hat{x}(t|s)$ this means that $\hat{x}(t|s)$ can be written as

$$\hat{x}(t|s) = \int_0^s h(t,\tau,\ \{m(\sigma),\ 0\leqslant\sigma<\tau\})m(\tau)\ d\tau \tag{3.43}$$

where the kernel $h(.,\tau,.)$ at time τ depends on all innovations up to time τ. The representation (3.43) is clearly a generalization of the representation (3.33) for the linear case. Substituting (3.43) in the condition (3.42) and using representations for f similar to (3.43) it can be shown that $\hat{x}(t|s)$ can be written as

$$\hat{x}(t|s) = \int_0^s \mathcal{E}\left\{x(t)m(\tau)^T|m(\sigma),\ 0\leqslant\sigma<\tau\right\}m(\tau)d\tau \tag{3.44}$$

which generalizes the representation (3.36) for the linear case. However, the essential difference is that in the non-linear least-squares estimate (3.44) the kernel is itself a conditional expectation which is not suitable for direct computation.

The general representation (3.44) can be evaluated for various special cases. The most direct generalization of the Kalman-Bucy problem is the following, where the state and observation process satisfy

$$\frac{d}{dt}\,x(t) = A(t,x(t)) + B(t,x(t))w(t) \tag{3.45}$$

$$y(t) = C(t,x(t)) + v(t) \quad . \tag{3.46}$$

The matrices A, B and C depend non-linearly on the current value of the state $x(t)$ and the Gaussian white noises $w(t)$ and $v(t)$ are assumed to be independent. By differentiating the representation (3.44) and using (3.45) and (3.46) it can be shown that $\hat{x}(t|t)$ is the solution of the differential equation

$$\frac{d}{dt}\,\hat{x}(t|t) = \hat{A}(t,x(t)) + K(t)m(t) \tag{3.47}$$

where

$$\begin{aligned} \hat{A}(t,x(t)) &= \mathcal{E}\{A(t,x(t))|m(\sigma),\ \sigma<t\} \\ \hat{C}(t,x(t)) &= \mathcal{E}\{C(t,x(t))|m(\sigma),\ \sigma<t\} \\ K(t) &= \mathcal{E}\{x(t)m(t)^T|m(\sigma),\ \sigma<t\} \end{aligned} \tag{3.48}$$

and

$$m(t) = y(t) - \hat{C}(t,x(t)) \tag{3.49}$$

$\hat{A}(t,x(t))$ and $\hat{C}(t,x(t))$ represent the least-squares estimates of $A(t,x(t))$ and $C(t,x(t))$ respectively.

The representation (3.44) for the least-squares estimate $\hat{x}(t|s)$ is expressed completely in terms of the innovations and based on the equivalence of observations and innovations under the conditions described. In reference 10 important theoretical results like (3.47) are given, based on martingale theory and which hold under weaker conditions on the signal $\gamma(t)$, but which are expressed in both the observations and the innovations.

3.6 Martingale approach of Balakrishnan

In "A martingale approach to linear recursive state estimation" (Ref. 3) Balakrishnan has given an elegant proof of the innovation and representation theorem as described in chapter 3.3 for the linear continuous time filtering problem. This section gives an outline of his approach and a remark on some related work.

Let the state be described by

$$x(t) = \int_0^t A(s)x(s)ds + \int_0^t B(s)dW(s) \tag{3.50}$$

and the observations by

$$Y(t) = \int_0^t C(s)x(s)ds + \int_0^t D(s)dW(s) \tag{3.51}$$

where $W(t)$ is a Brownian motion. Notice that the form in which the noise is given in the system (3.50) and (3.51) implies a correlation between the system and observation noise. The state $x(t)$ is assumed to be an $(n\times 1)$-vector, whereas the Brownian motion $W(t)$ and the observation process $Y(t)$ are $(r\times 1)$- and $(m\times 1)$-vectors respectively. Let $\mathcal{F}(t)$ be the information field generated by the observation process $Y(s), s\leqslant t$, i.e. $\mathcal{F}(t) \equiv \sigma\{Y(s), s\leqslant t\}$ and let $\hat{x}(t|t) = \mathcal{E}\{x(t)|Y(s), s\leqslant t\}$ be the filtered least-squares estimate of $x(t)$ given the observations up to time t.

The innovation process $Z_o(t)$ associated with the observation process $Y(t)$ is defined by

$$Z_o(t) = Y(t) - \int_0^t C(s)\hat{x}(s|s)ds \tag{3.52}$$

and is an $\mathcal{F}(t)$ martingale. Under the assumption that $D(t)D(t)^T>0$ for every t it has been proved that for every t
$\mathcal{F}(t) \equiv \sigma\{Y(s), s\leqslant t\} = \sigma\{Z_o(s), s\leqslant t\}$ i.e. the information fields (σ-algebras) generated by the observation and innovation processes are the same. Another process $Z_s(t)$ is defined by

$$Z_s(t) = \hat{x}(t|t) - \int_0^t A(s)\hat{x}(s|s)ds \tag{3.53}$$

where the subscript s in $Z_s(t)$ refers to the state. The process $Z_s(t)$ can be shown to be also an $\mathcal{F}(t)$ martingale. According to the representation theorem of section 3.3 the state martingale $Z_s(t)$ can therefore be written as

$$Z_s(t) = \int_0^t \alpha(s)dZ_o(s) \tag{3.54}$$

where $\alpha(s)$ has to be determined yet from the system description. The way this has been done by Balakrishnan is as follows. First, it was proved that the filtered least-squares estimate $\hat{Z}_s(t|t)$ of the state martingale given the innovations up to time t can be written as

$$\begin{aligned}\hat{Z}_s(t|t) &= \mathcal{E}\{Z_s(t)|Z_o(s),\ s\leq t\}\\ &= \int_0^t r_{12}(s)dZ_o(s)\end{aligned} \tag{3.55}$$

where

$$r_{12}(t) = P_{12}(t)/P_{22}(t) \tag{3.56}$$

and

$$\begin{aligned}P_{12}(t) &= \lim_{\Delta\to 0}\frac{1}{\Delta}\,\mathcal{E}\left\{\int_t^{t+\Delta} dZ_s(s)\int_t^{t+\Delta} dZ_o(s)^T|\mathcal{F}(t)\right\}\\ &= P_{\tilde{x}}(t|t)C(t)^T + B(t)D(t)^T\end{aligned} \tag{3.57}$$

$$P_{\tilde{x}}(t|t) = \mathcal{E}\left\{\Big(x(t) - \hat{x}(t|t)\Big)\Big(x(t) - \hat{x}(t|t)\Big)^T\right\} \tag{3.58}$$

$$\begin{aligned}P_{22}(t) &= \lim_{\Delta\to 0}\frac{1}{\Delta}\,\mathcal{E}\left\{\int_t^{t+\Delta} dZ_o(s)\int_t^{t+\Delta} dZ_o(s)^T|\mathcal{F}(t)\right\}\\ &= D(t)D(t)^T\ .\end{aligned} \tag{3.59}$$

But because $Z_s(t)$ is an $\mathcal{F}(t)$ martingale it follows that $Z_s(t)$ is $\mathcal{F}(t)$ measurable for every t and hence measurable with respect to the σ-algebra generated by the innovation process $Z_o(s), s\leq t$, which, as was shown above, is the same as $\mathcal{F}(t)$. Thus it follows that the least-squares estimate $\hat{Z}_s(t|t)$ of the state martingale $Z_s(t)$ is $Z_s(t)$ itself and the latter can be written as

$$Z_s(t) = \int_o^t r_{12}(s)dZ_o(s) \tag{3.60}$$

which is the required representation (3.54). Combining the results found so far it follows that

$$\hat{x}(t|t) = \int_o^t A(s)\hat{x}(s|s)ds + \int_o^t \left(P_{\tilde{x}}(s|s)C(s)^T + B(s)D(s)^T\right) * \left\{D(s)D(s)^T\right\}^{-1} dZ_o(s) \tag{3.61}$$

or after differentiation with respect to t:

$$\frac{d}{dt}\,\hat{x}(t|t) = A(t)\hat{x}(t|t) + \left\{P_{\tilde{x}}(t|t)C(t)^T + B(t)D(t)^T\right\} * \left\{D(t)D(t)^T\right\}^{-1} \frac{d}{dt}\,Z_o(t)$$

where

$$dZ_o(t) = dY(t) - C(t)\hat{x}(t|t)dt \tag{3.62}$$

and which combined with the equation

$$\frac{d}{dt}\,P_{\tilde{x}}(t|t) = A(t)P_{\tilde{x}}(t|t) + P_{\tilde{x}}(t|t)A(t)^T + B(t)B(t)^T + -\left\{P_{\tilde{x}}(t|t)C(t)^T + B(t)D(t)^T\right\}\left\{D(t)D(t)^T\right\}^{-1}\left\{C(t)P_{\tilde{x}}(t|t) + D(t)B(t)^T\right\} \tag{3.63}$$

for $P_{\tilde{x}}(t|t)$ is exactly the Kalman-Bucy filter for the continuous time linear filtering problem with correlated system and observation noise.

In reference 2 a similar derivation of the Kalman-Bucy filter is given, however, based on a different state martingale $Z_s(t)$. The state martingale used there is $\hat{x}(t|s)$, i.e. the least-squares estimate of $x(t)$ given the observations $Y(\tau)$, $\tau \leqslant s$ up to time s. For fixed t and s varying it can be shown that $\hat{x}(t|s)$ is a martingale in s with respect to $\mathcal{F}(s) = \sigma\{Y(\tau), \tau \leqslant s\}$. This state martingale has also been used in reference 1 to solve the smoothing problem where it proved to be a more natural one than the state martingale of Balakrishnan.

4 WIDE-SENSE MARTINGALE APPROACH

In the linear-estimation problems considered in the preceeding chapters it was assumed that the state x(t) satisfied a linear difference or differential equation driven by white noise. It is well known (Refs 4, 26) that the solution x(t) generated then is a wide-sense Markov process (defined in section 4.1). Moreover, it is known that a wide-sense Markov process can be written as a linear transformation of a wide-sense martingale (defined in section 4.1). The wide-sense martingale approach described in this chapter and developed for both discrete and continuous time estimation has resulted in recursive estimation equations, based on noisy observations, for a signal $x(t) = \phi(t)u(t)$, x(t) and u(t) being wide-sense Markov and wide-sense martingale respectively. Because of the linear relation between both signals the whole estimation problem can be carried through for the wide-sense martingale u(t) which gives due to the martingale property simpler and more clear derivations as well as results.

It will be shown that the signals studied in the Kalman and Kalman-Bucy problem can easily be brought into the form $x(t) = \phi(t)u(t)$ considered here. It might therefore seem contradictory that the recursive estimation equations are simpler than the original ones. However, as will be seen, there is a linear relation between them and it is just this linear transformation that causes the simplification.

After a discussion of wide-sense martingales and Markov processes in section 4.1, a very general form of the linear discrete time estimation problem, including coloured observation noise is discussed in section 4.2. The linear continuous time estimation problem for observations in additive white noise is discussed in section 4.3.

4.1 Wide-sense martingales and wide-sense Markov processes

Many concepts in the theory of stochastic processes have both a wide-sense and a strict-sense version although the phrase strict sense is usually omitted. For example, martingale is shorthand for strict-sense martingale. In the definition of martingales given in chapter 3.1 it was required for the stochastic process to be integrable, i.e. for all t $\mathcal{E}\{|x(t)|\}<\infty$. (This is because the conditional expectation is only defined for integrable random variables.) Wide-sense properties always require the process to be square integrable, i.e. for all t $\mathcal{E}\{|x(t)|^2\}<\infty$.

The most natural way to study wide-sense processes then is in the Hilbert space of quadratically integrable random variables on a probability space $(\Omega,\mathcal{B},P)$.

Thus, let $L_2^n(\Omega)$ be the Hilbert space of (nx1)-dimensional random vectors $x(t)$ that are square integrable, i.e. $\mathcal{E}\left\{x(t)^T x(t)\right\}<\infty$ for each t. Further, let $L_2(x;t)$ denote the subspace of L_2^n spanned by $x(s)$, $s\leq t$ and $L_2(x(t))$ the subspace spanned by $x(t)$. The norm $\|.\|$ defined in the Hilbert space $L_2^n(\Omega)$ is given by $\|x(t)\|=\mathcal{E}\left\{x(t)^T x(t)\right\}$ for every $x(t)$ in $L_2^n(\Omega)$. It follows that for every z in $L_2^n(\Omega)$ there is a unique element $\hat{z}$ in $L_2(x;t)$ such that

$$\| z-\hat{z}\| = \min_{u\in L_2(x;t)} \| z-u\| \tag{4.1}$$

i.e. $\hat{z}$ is the linear least-squares estimate (or wide-sense conditional mean) of z based on $L_2(x;t)$. The linear least-squares estimate $\hat{z}$ is orthogonal to the elements of $L_2(x;t)$, i.e.

$$\mathcal{E}\left\{(z-\hat{z})u^T\right\} = 0 \quad \text{for every } u\in L_2(x;t) \tag{4.2}$$

and is therefore called the projection of z on $L_2(x;t)$. From here on the symbols $(z|L_2(x;t))$ and $(z|L_2(x(t)))$ are used for the projections of z on $L_2(x;t)$ and $L_2(x(t))$ respectively.

Identifying $L_2(x;t)$ with the past of the process $x(t)$ up to the present inclusive and $L_2(x(t))$ with the present, wide-sense Markov processes and wide-sense martingales are defined as follows. A square integrable vector process $x(t)$ is called a wide-sense Markov process if for $s\leq t$

$$(x(t)|L_2(x;s)) = (x(t)|L_2(x(s))) \tag{4.3}$$

i.e. the projection of the future on the past up to and including the present is the same as the projection on the present only. A second order process $x(t)$ is called a wide-sense martingale if for $s\leq t$

$$(x(t)|L_2(x;s)) = x(s) \tag{4.4}$$

i.e. the projection of the future on the past up to the present inclusive is the present itself. For a wide-sense Markov process it follows from equation (4.3) that

$$(x(t)|L_2(x;s)) = F(t,s)x(s) \tag{4.5}$$

where $F(t,s)$ is an nxm-matrix function of t and s. Hence, the wide-sense martingales constitute a subclass of the wide-sense Markov processes. More generally, it has been shown in the literature that when the covariance matrix function of a wide-sense Markov process $x(t)$ is non-singular for all t,s then $x(t)$ can be written as a non-singular linear transformation of a wide-sense martingale $u(t)$, that is, $x(t) = \phi(t)u(t)$ and $L_2(x;t) = L_2(u;t)$. It is for this class of processes that the derivation of recursive estimation equations is given in the following sections.

4.2 Wide-sense martingale approach; discrete time

In this section recursive estimation equations for a certain class of discrete time signals are described.

4.2.1 Definition of the estimation problem

Let the state $x(k)$((nx1)-vector) be given by

$$x(k) = \phi(k)u(k) \qquad k = 0,1,2,... \tag{4.6}$$

where $\phi(k)$ is a known (nxn)-matrix time function and $u(k)$((nx1)-vector) is a discrete time wide-sense martingale, zero mean and with covariance matrix $P_u(k) = \mathcal{E}\left\{u(k)u(k)^T\right\}$. The state $x(k)$ is then a wide-sense Markov process. Let the observation process $y(k)$ ((mx1)-vector) be given in the form

$$y(k) = C(k)x(k) + v(k) \tag{2.9}$$

or

$$y(k) = H(k)u(k) + v(k) \tag{4.7}$$

where $H(k) = C(k)\phi(k)$ and $v(k)$ ((mx1)-vector) is additive noise. Assume that the initial condition $u(o) = u^o$ is uncorrelated with the observation noise $v(k)$, $k = 0,1,2,...$. For the observation noise three cases are distinguished. These are: white observation noise uncorrelated or correlated with the martingale increment and coloured observation noise correlated with the martingale increment. A usual approach to coloured observation noise is to assume that it is generated by a linear difference equation driven by white noise. The approach described here is to give the coloured observation noise by an equation of the form $v(k) = \psi(k)m(k)$ where $m(k)$((mx1)-vector) is a wide-sense martingale and hence, $v(k)$ a wide-sense Markov process.

The linear least-squares estimate $\hat{x}(k|\ell)$ of $x(k)$ is given by the projection of $x(k)$ on the subspace $L_2^n(y;\ell)$ of $L_2^n(\Omega)$, i.e.

$$\hat{x}(k|\ell) = (x(k)|L_2^n(y;\ell)) \quad . \tag{4.8}$$

The subspace $L_2^n(y;\ell)$ of $L_2^n(\Omega)$ is spanned by vectors $F(i)y(i)$, $i = 1, \ldots, \ell$ where $F(i)$ are arbitrary $n \times m$ matrices. It follows immediately from the linear relation (4.6) between the signals $x(k)$ and $u(k)$ that the linear least-squares estimate $\hat{x}(k|\ell)$ and $\hat{u}(k|\ell)$ of $x(k)$ and $u(k)$ are related to each other as

$$\begin{aligned} \hat{x}(k|\ell) &= \phi(k)\hat{u}(k|\ell) \\ \tilde{x}(k|\ell) &= \phi(k)\tilde{u}(k|\ell) \\ P_{\tilde{x}}(k|\ell) &= \phi(k)P_{\tilde{u}}(k|\ell)\phi(k)^T \end{aligned} \tag{4.9}$$

where $\hat{u}(k|\ell) = \left(u(k)|L_2^n(y;\ell)\right)$, $\tilde{u}(k|\ell) = u(k) - \hat{u}(k|\ell)$ and $P_{\tilde{u}}(k|\ell) = \mathcal{E}\left\{\tilde{u}(k|\ell)u(k|\ell)^T\right\}$. Therefore, the estimation problem for $x(k)$ can be solved once the estimation problem for $u(k)$ has been solved, using the equations (4.9).

After the description of the recursive estimation equations for the wide-sense martingale $u(k)$, the Kalman problem is discussed at the end of each subsection below.

4.2.2 White observation noise uncorrelated with the martingale increment

Let the observation noise $v(k)$ be zero mean white noise, $\mathcal{E}\left\{v(k)v(\ell)^T\right\} = P_v(k)\delta_{k\ell}$ and uncorrelated with the martingale increment, i.e.

$$\mathcal{E}\left\{\left(u(k+1)-u(k)\right)v(j)^T\right\} \equiv 0 \qquad \text{for all } j,k \quad . \tag{4.10}$$

It follows then immediately that for $k>\ell$ $\left(u(k) - u(\ell)|L_2^n(y;\ell)\right) = 0$ and the predicted estimate $\hat{u}(k|\ell)$ is found to be

$$\begin{aligned} \hat{u}(k|\ell) = \left(u(k)|L_2^n(y;\ell)\right) &= \left(u(k) - u(\ell)|L_2^n(y;\ell)\right) + \\ &+ \left(u(\ell)|L_2^n(y;\ell)\right) = \hat{u}(\ell|\ell) \end{aligned} \tag{4.11}$$

i.e. the predicted estimate $\hat{u}(k|\ell)$ for all $k>\ell$ is equal to the filtered estimate $\hat{u}(\ell|\ell)$. In reference 21 the smoothing, filtering and prediction problem are treated in detail (see also Ref. 22). Here we are mainly interested in the filtered estimate $\hat{u}(k|k)$ which satisfies

$$\hat{u}(k|k) = \hat{u}(k|k-1) + K(k)\Big[y(k) - H(k)\hat{u}(k|k-1)\Big] \tag{4.12}$$

$$K(k) = P_{\tilde{u}}(k|k-1)H(k)^T \Big[H(k)P_{\tilde{u}}(k|k-1)H(k)^T + P_v(k)\Big]^{-1} \tag{4.13}$$

$$P_{\tilde{u}}(k|k) = P_{\tilde{u}}(k|k-1) - K(k)H(k)P_{\tilde{u}}(k|k-1) \tag{4.14}$$

where $H(k) = C(k)\phi(k)$ and the predicted estimate $\hat{u}(k|k-1)$ and the associated covariance matrix satisfy

$$\hat{u}(k|k-1) = \hat{u}(k-1|k-1) \tag{4.15}$$

$$P_{\tilde{u}}(k|k-1) = P_{\tilde{u}}(k-1|k-1) + P_u(k) - P_u(k-1) \quad . \tag{4.16}$$

The following initial conditions complete the filtering equations for the estimation problem considered here:

$$\hat{u}(1|o) = 0 \qquad P_{\tilde{u}}(1|0) = P_u(1) \quad . \tag{4.17}$$

Notice that the structure of the filtering equations (4.12), ..., (4.14) is the same as that of the Kalman filter (2.10), (2.12) and (2.14) for a state $x(k)$ given by $x(k+1) = x(k) + B(k)w(k)$. It is interesting to establish that the stochastic process $\hat{u}(k|k)$ $k = 0,1,2,...$ which is defined by the filtered estimate is a zero mean wide-sense martingale. To prove this, observe that for $k \geqslant \ell$

$$\Big(\hat{u}(k|k)|L_2(\hat{u};\ell)\Big) = \Big(\Big(u(k)|L_2^n(y;k)\Big)|L_2(\hat{u};\ell)\Big)$$

$$= \Big(\Big((u(k)|L_2^n(y;k)\Big)|L_2^n(y;\ell)\Big)|L_2(\hat{u};\ell)\Big)$$

$$= \Big(\Big(u(k)|L_2^n(y;\ell)\Big)|L_2(\hat{u};\ell)\Big)$$

$$= \Big(\hat{u}(k|\ell)|L_2(\hat{u};\ell)\Big) = \Big(\hat{u}(\ell|\ell)|L_2(\hat{u};\ell)\Big)$$

$$= \hat{u}(\ell|\ell)$$

because for $k \geqslant \ell$ $L_2^n(y;k) \supset L_2^n(y;\ell) \supset L_2(\hat{u};\ell)$.

Consider now the Kalman problem. As described in chapter 2.2 the state $x(k)$ is then given by a linear difference equation driven by white noise $w(k)$:

$$\begin{aligned} x(k+1) &= A(k+1,k)x(k) + B(k)w(k) \\ x(o) &= x^o \quad . \end{aligned} \tag{2.8}$$

The observations are given by equation (2.9). The initial condition x^o and the noise sequences $w(k)$ and $v(k)$ are assumed to be independent.

The solution of the difference equation (2.8) can be written in the form

$$x(k) = \phi(k,o)\left\{x^o + \sum_{i=1}^{k} \phi(o,i)B(i-1)w(i-1)\right\} \tag{4.18}$$

where

$$\phi(k,o) = \prod_{i=1}^{k} A(i,i-1) \ , \ \phi(o,k) = \phi(k,o)^{-1} \quad . \tag{4.19}$$

Hence, by the definition

$$\begin{aligned} u(k) &= x^o + \sum_{i=1}^{k} \phi(o,i)B(i-1)w(i-1) \\ u(o) &= x(o) = x^o \end{aligned} \tag{4.20}$$

the signal $x(k)$ can be written in the form $x(k) = \phi(k,o)u(k)$. To show that $u(k)$ is a wide-sense martingale it must be proved that for every $k,\ell,k\geqslant\ell$

$$\Big(u(k)|L_2(u;\ell)\Big) = u(\ell) \qquad k\geqslant\ell \quad . \tag{4.21}$$

It follows from definition (4.20) that for $k>\ell$

$u(k)-u(\ell) = \sum_{i=\ell+1}^{k} \phi(o,i)B(i-1)w(i-1)$ and therefore,

$$\mathcal{E}\left\{\Big(u(k)-u(\ell)\Big)u(j)^T\right\} = \sum_{i=\ell+1}^{k} \phi(o,i)B(i-1)\mathcal{E}\left\{w(i-1)x^{o^T}\right\} +$$

$$+ \sum_{i=\ell+1}^{k} \phi(o,i)B(i-1) \sum_{n=1}^{j} \phi(o,n)B(n-1)\mathcal{E}\left\{w(i-1)w(n-1)^T\right\} .$$

Because the sequence $w(k)$ is white and independent of x^o it follows that

$$\mathcal{E}\left\{\Big(u(k)-u(\ell)\Big)u(j)^T\right\} = 0 \qquad k>\ell\geqslant j \quad . \tag{4.22}$$

Hence (4.21) holds. The wide-sense martingale $u(k)$ satisfies the linear difference equation driven by white noise

$$u(k+1) = u(k) + \phi(o,k+1)B(k)w(k) \quad .$$

Furthermore,

$$\mathcal{E}\left\{\Big(u(k+1)-u(k)\Big)v(j)^T\right\} = \phi(o,k+1)B(k)\mathcal{E}\left\{w(k)v(j)^T\right\} = 0 \tag{4.23}$$

because the noise sequences $w(k)$ and $v(k)$ are independent. Thus, the martingale increment is uncorrelated with the observation noise in the Kalman problem.

If it is assumed that $\mathcal{E}\{x^o\} = 0$ then u(k) is zero mean and its covariance matrix $P_u(k) = \mathcal{E}\left\{u(k)u(k)^T\right\}$ is

$$P_u(k) = \mathcal{E}\left\{x^o x^{oT}\right\} + \sum_{i=1}^{k} \phi(o,i)B(i-1)\mathcal{E}\left\{w(i-1)w(i-1)^T\right\}B(i-1)^T\phi(o,i)^T \quad . \tag{4.24}$$

Thus, the covariance matrix $P_u(k)$ of the wide-sense martingale u(k) is completely known given the initial covariance matrix $\mathcal{E}\left\{x^o x^{oT}\right\}$ and the system noise covariance matrix $\mathcal{E}\left\{w(k)w(k)^T\right\} = P_w(k)$.

Therefore, using equations (4.20), (4.21), (4.23) and (4.24) it is shown that the Kalman problem can be written in the form $x(k) = \phi(k)u(k)$ where u(k) is a wide-sense martingale with known covariance matrix $P_u(k)$ and the martingale increment uncorrelated with the observation noise. Hence, the Kalman problem can be solved alternatively by using the algorithm (4.12), ..., (4.17) of the wide-sense martingale approach in conjunction with the equations (4.9). Whether the direct approach using the Kalman filter or the alternative one is more efficient is discussed in chapter 5.

4.2.3 White observation noise correlated with the martingale increment

Let the observation noise be zero mean white noise, $\mathcal{E}\left\{v(k)v(\ell)^T\right\} = P_v(k)\delta_{k\ell}$ and correlated with the martingale increment, i.e.

$$\mathcal{E}\left\{\left(u_{k+1}-u_k\right)v(\ell)^T\right\} = S(k)\delta_{k\ell} \quad \text{for all } \ell,k \quad . \tag{4.25}$$

It follows that for $k>\ell+1$ $\left(u(k)-u(\ell)\,|\,L_2^n(y;\ell)\right) = 0$ and the predicted estimate $\hat{u}(\ell+i|\ell)$ for $i>1$ satisfies

$$\begin{aligned}\hat{u}(\ell+i|\ell) &= \left(u(\ell+i)-u(\ell+1)\,|\,L_2^n(y;\ell)\right) + \left(u(\ell+1)\,|\,L_2^n(y;\ell)\right) \\ &= \hat{u}(\ell+1|\ell) \qquad i>1\end{aligned} \tag{4.26}$$

i.e., the i-step predicted estimate (i>1) is equal to the one step predicted estimate.

Due to the correlation between the martingale increment and the observation noise, it is simplest for this case to give a recursion for the one step predicted estimate $\hat{u}(k+1|k)$ and to compute the filtered estimate $\hat{u}(k+1|k+1)$ from $\hat{u}(k+1|k)$ separately. The recursion for $\hat{u}(k+1|k)$ is

$$\hat{u}(k+1|k) = \hat{u}(k|k-1) + K(k+1|k)\Big[y(k) - M(k)\hat{u}(k|k-1)\Big] \tag{4.27}$$

$$K(k+1|k) = \Big[P_{\tilde{u}}(k|k-1)H(k)^T + S(k)\Big]\Big[H(k)P_{\tilde{u}}(k|k-1)H(k)^T + P_v(k)\Big]^{-1} \tag{4.28}$$

$$\begin{aligned} P_{\tilde{u}}(k+1|k) = P_{\tilde{u}}(k|k-1) - \Big[P_{\tilde{u}}(k|k-1)H(k)^T + S(k)\Big] * \\ \Big[H(k)P_{\tilde{u}}(k|k-1)H(k)^T + P_v(k)\Big]^{-1}\Big[H(k)P_{\tilde{u}}(k|k-1)H(k)^T + S(k)^T\Big] + \\ + P_u(k+1) - P_u(k) \end{aligned} \tag{4.29}$$

where the initial condition is

$$\hat{u}(1|o) = 0 \quad , \quad P_{\tilde{u}}(1|o) = P_u(1) \quad . \tag{4.30}$$

The process $\hat{u}(k+1|k)$ $k = 0,1,2,\ldots$ is a zero mean wide-sense martingale.

The filtered estimate $\hat{u}(k|k)$ is related to the predicted estimate $\hat{u}(k|k-1)$ and the observation $y(k)$ by

$$\hat{u}(k|k) = \hat{u}(k|k-1) + K(k)\Big[y(k) - H(k)\hat{u}(k|k-1)\Big] \tag{4.31}$$

where

$$K(k) = P_{\tilde{u}}(k|k-1)H(k)^T \Big[H(k)P_{\tilde{u}}(k|k-1)H(k)^T + P_v(k)\Big]^{-1} \tag{4.32}$$

$$\begin{aligned} P_{\tilde{u}}(k|k) = P_{\tilde{u}}(k|k-1) - P_{\tilde{u}}(k|k-1)H(k)^T\Big[H(k)P_{\tilde{u}}(k|k-1)H(k)^T + \\ + P_v(k)\Big]^{-1} H(k)P_{\tilde{u}}(k|k-1) \quad . \end{aligned} \tag{4.33}$$

The structure of the above algorithm is the same as that of the Kalman filter for a state $x(k)$ satisfying $x(k+1) = x(k) + B(k)w(k)$ where the system and observation noises $w(k)$ and $v(k)$ are correlated.

As for the Kalman problem (2.8) and (2.9), it can be seen from equation (4.23) that correlated martingale increment and observation noise corresponds with correlated system noise and observation noise, i.e.

$$\mathcal{E}\Big\{w(k)v(j)^T\Big\} = S(k)\delta_{kj} \quad . \tag{4.34}$$

The direct algorithm for this version of the Kalman problem consists of a recursion for the predicted estimate $\hat{x}(k+1|k)$ and a separate algorithm for computing the filtered estimate $\hat{x}(k|k)$ from $\hat{x}(k|k-1)$ and $y(k)$ like equations (4.31), (4.32) and (4.33). (See for example Ref. 11.)

4.2.4 Coloured observation noise correlated with the martingale increment

Let the observation noise be zero mean coloured noise. A usual approach is to assume that the coloured noise $v(k)$ is generated by a linear difference equation driven by white noise $n(k)$, that is

$$\begin{aligned} v(k+1) &= a(k+1,k)v(k) + n(k) \\ v(o) &= v^{o} \end{aligned} \tag{4.35}$$

where v^{o} and the white-noise sequence $n(k)$ are independent. But the solution $v(k)$ of equation (4.35) can be written as

$$v(k) = \Lambda(k,o)\left\{v^{o} + \sum_{i=1}^{k} \Lambda(o,i)n(i-1)\right\} \tag{4.36}$$

where

$$\Lambda(k,o) = \prod_{i=1}^{k} a(i,i-1) \ , \ \Lambda(o,i) = \Lambda(i,o)^{-1} \ . \tag{4.37}$$

Now define the process $m(k)$ by

$$\begin{aligned} m(k) &= v^{o} + \sum_{i=1}^{k} \Lambda(o,i)n(i-1) \\ m(o) &= v^{o} \ . \end{aligned} \tag{4.38}$$

The coloured noise $v(k)$ can be written then as $v(k) = \Lambda(k,o)m(k)$ where $m(k)$ satisfies the linear-difference equation driven by white noise $n(k)$

$$m(k+1) = m(k) + \Lambda(o,k+1)n(k) \ . \tag{4.39}$$

It follows that the process $m(k)$ is a zero mean wide-sense martingale (compare the Kalman problem at the end of section 4.2.2) with covariance matrix $P_m(k) = \mathcal{E}\left\{m(k)m(k)^T\right\}$ where

$$\begin{aligned} P_m(k) &= \mathcal{E}\left\{v^{o}v^{oT}\right\} + \sum_{i=1}^{k} \Lambda(o,i)\mathcal{E}\left\{n(i-1)n(i-1)^T\right\}\Lambda(o,i)^T \\ &= P_m(k-1) + \Lambda(o,k)\mathcal{E}\left\{n(k-1)n(k-1)^T\right\}\Lambda(o,k)^T \ . \end{aligned} \tag{4.40}$$

Using this result, the observations $y(k)$ can be written in the form

$$y(k) = H(k)u(k) + \Lambda(k,o)m(k) \tag{4.41}$$

where both $u(k)$ and $m(k)$ are wide-sense martingales. It follows from the above analysis that the case of coloured observation noise $v(k)$ can be more generally described by $v(k) = \psi(k)m(k)$ where $m(k)$ is a zero mean wide-sense martingale with known covariance matrix then by a linear difference equation for $v(k)$.

Assume further that the initial condition v^o is independent of the martingale $u(k)$ for every k and that the martingale increment $u(k+1) - u(k)$ is correlated with the white noise $n(k)$ as follows:

$$\mathcal{E}\left\{\left(u(k+1) - u(k)\right)n(j)^T\right\} = S(k)\delta_{kj} \quad . \tag{4.42}$$

Then, for $j \leq k$

$$\mathcal{E}\left\{\left(u(k+1) - u(k)\right)m(j)^T\right\} = \mathcal{E}\left\{\left(u(k+1) - u(k)\right)v^{oT}\right\} + $$
$$+ \sum_{i=1}^{j} \mathcal{E}\left\{\left(u(k+1) - u(k)\right)n(i-1)^T\right\}\Lambda(o,i)^T = 0+0 = 0 \tag{4.43}$$

and therefore $(j \leq k)$

$$\mathcal{E}\left\{\left(u(k+1) - u(k)\right)y(j)^T\right\} = \mathcal{E}\left\{\left(u(k+1) - u(k)\right)u(j)^T\right\}H(j)^T + $$
$$+ \mathcal{E}\left\{\left(u(k+1) - u(k)\right)m(j)^T\right\}\Lambda(j,o)^T = 0+0 = 0 \tag{4.44}$$

because $u(k)$ is a wide-sense martingale. Equation (4.44) implies that for $k>\ell$, $u(k) - u(\ell)$ is orthogonal to $L_2^n(y;\ell)$. Therefore,

$$\hat{u}(k|\ell) = \left(u(k)|L_2^n(y;\ell)\right) = \left(u(k) - u(\ell)|L_2^n(y;\ell)\right) + $$
$$+ \left(u(\ell)|L_2^n(y;\ell)\right) = \hat{u}(\ell|\ell) \tag{4.45}$$

that is, the predicted estimate $\hat{u}(k|\ell)$ is equal to the filtered estimate $\hat{u}(\ell|\ell)$ for all $k>\ell$.

It remains now to give an equation for the filtered estimate $\hat{u}(\ell|\ell)$ of the wide-sense martingale $u(\ell)$ given observations $y(i)$, $i \leq \ell$ in coloured martingale noise $m(i)$ as in equation 4.41. The filtering equations are

$$\hat{u}(k|k) = \hat{u}(k|k-1) + K(k)\left[y(k) - a(k,k-1)y(k-1) - \tilde{H}(k)\hat{u}(k|k-1)\right] \tag{4.46}$$

$$K(k) = \left[P_{\tilde{u}}(k|k-1)H(k)^T + S(k-1)\right]P_{\tilde{y}}^{-1}(k|k-1) \tag{4.47}$$

$$P_{\tilde{y}}(k|k-1) = \tilde{H}(k)P_{\tilde{u}}(k|k-1)\tilde{H}(k)^T + H(k)Q(k-1)H(k-1)^T *$$
$$a(k,k-1)^T + a(k,k-1)H(k-1)Q(k-1)H(k)^T +$$
$$+ H(k)S(k-1) + S(k-1)^TH(k)^T + P_n(k-1) +$$
$$- a(k,k-1)H(k-1)Q(k-1)H(k-1)^Ta(k,k-1)^T \tag{4.48}$$

where

$$\tilde{H}(k) = H(k) - a(k,k-1)H(k-1) \qquad (4.49)$$

$$Q(k-1) = P_u(k) - P_u(k-1) \qquad (4.50)$$

and finally

$$P_{\tilde{u}}(k|k) = P_{\tilde{u}}(k|k-1) - K(k)\left[\tilde{H}(k)P_{\tilde{u}}(k|k-1) + S(k)^T\right] \quad . \qquad (4.50^a)$$

The predicted estimate $\hat{u}(k|k-1)$ and covariance matrix satisfy

$$\hat{u}(k|k-1) = \hat{u}(k-1|k-1) \qquad (4.51)$$

$$P_{\tilde{u}}(k|k-1) = P_{\tilde{u}}(k-1|k-1) + P_u(k) - P_u(k-1) \qquad (4.52)$$

whereas the initial conditions are

$$\hat{u}(1|o) = 0 \quad , \quad P_{\tilde{u}}(1|o) = P_u(1) \quad . \qquad (4.53)$$

The process $\hat{u}(k|k)$, $k = 0,1,2,\ldots$ is a zero mean wide-sense martingale.

Notice that in the measurement update, equation (4.46) for the filtered estimate two observations are processed simultaneously. The algorithm is of the type developed by Bryson and Henrikson (Ref. 6) for coloured noise, avoiding the augmented state procedure which, as is well known, often gives rise to ill conditioned problems.

In this approach the first observation is still processed by using the augmented state procedure.

The type of Kalman problem covered by the wide-sense martingale approach considered in this section consists of observations in coloured noise $v(k)$ (equations (4.7) and (4.35)) where the driving noise $n(k)$ is correlated with the system noise $w(k)$,

$$\mathcal{E}\left\{\Big(u(k+1) - u(k)\Big)n(j)^T\right\} = \Lambda(o,k+1)B(k)\mathcal{E}\left\{w(k)n(j)^T\right\} \qquad (4.54)$$

i.e. $S(k) = \Lambda(o,k+1)B(k)$ in (4.42). Therefore, it is an extension of the already cited approach of Bryson and Henrikson where it was assumed that the driving noise $n(k)$ and the system noise $w(k)$ were independent.

4.3 Wide-sense martingale approach; continuous time

In this section the continuous time version of the estimation problem using the wide-sense martingale approach is discussed, however, only for the simplest case of uncorrelated martingale increment and observation noise. For $0 \leq t \leq 1$, say, let the state process $x(t)$ ($(n\times1)$-vector) be given in the form

$$x(t) = \phi(t)u(t) \tag{4.55}$$

where $\phi(t)$ is a known matrix time function and $u(t)$ ((nx1)-vector) is a continuous time zero mean wide-sense martingale with covariance matrix $P_u(t) = \mathcal{E}\left\{u(t)u(t)^T\right\}$. The state $x(t)$ is a wide-sense Markov process then. Suppose that the observation process is an (mx1)-vector process $Y(t)$ where $m \leqslant n$ and such that the martingale increment $u(t)-u(s)$ is uncorrelated with the past of the observations, i.e.

$$\mathcal{E}\left\{\Big(u(t) - u(s)\Big)Y(\tau)^T\right\} = 0 \qquad \tau \leqslant s \leqslant t \tag{4.56}$$

The linear least-squares estimates $\hat{x}(t|s)$ and $\hat{u}(t|s)$ of $x(t)$ and $u(t)$ given the observations $Y(\tau)$, $\tau \leqslant s$ are related to each other exactly as in the discrete time case, that is

$$\begin{aligned} \hat{x}(t|s) &= \phi(t)\hat{u}(t|s) \\ \tilde{x}(t|s) &= \phi(t)\tilde{u}(t|s) \\ P_{\tilde{x}}(t|s) &= \phi(t)P_{\tilde{u}}(t|s)\phi(t)^T \quad . \end{aligned} \tag{4.57}$$

Also the following prediction result is completely analogous to the discrete time case:

$$\hat{u}(t|s) = \hat{u}(s|s) \qquad t>s \tag{4.58}$$

$$P_{\tilde{u}}(t|s) = P_{\tilde{u}}(s|s) + P_u(t) - P_u(s) \quad . \tag{4.59}$$

To solve the filtering problem for $\hat{u}(s|s)$, let the observations be given by

$$Y(t) = \int_0^t C(s)x(s)ds + \int_0^t D(s)dW(s) \tag{4.60}$$

or, using $x(s) = \phi(s)u(s)$

$$Y(t) = \int_0^t H(s)u(s)ds + \int_0^t D(s)dW(s) \tag{4.61}$$

where $H(s) = C(s)\phi(s)$ and $W(t)$ is a standard Wiener process independent of the wide-sense martingale $u(t)$. (This implies that condition (4.56) holds.) Moreover, let $C(s)$ and $D(s)$ be continuous functions of time and let $R(t) = D(t)D(t)^T > 0$. It is proved in reference 24 that the filtered estimate $\hat{u}(s|s)$ is a wide-sense martingale and that it can be written as

$$\hat{u}(t|t) = \int_0^t K(s)dM(s) \tag{4.62}$$

where the kernel K(s) is an $n \times m$ matrix function, independent of t and M(s) is the innovation process of the observations defined by

$$M(t) = Y(t) - \int_0^t H(s)\hat{u}(s|s)ds \tag{4.63}$$

and such that $L_2^n(M;t) = L_2^n(Y;t)$ (compare equations (4.62) and (4.63) with equations (3.52) and (3.55) of Balakrishnan). The recursive filtering equations for u(t) are now readily obtained. Combining (4.62) and (4.63) and differentiating with respect to t yields

$$\frac{d}{dt}\,\hat{u}(t|t) = K(t)\,\frac{dM}{dt} \tag{4.64}$$

where it can be shown that

$$K(t) = P_{\tilde{u}}(t|t)H(t)^T R^{-1}(t) \tag{4.65}$$

and

$$\frac{d}{dt}\,P_{\tilde{u}}(t|t) = \frac{d}{dt}\,P_u(t) - P_{\tilde{u}}(t|t)H(t)^T R^{-1}(t)H(t)P_{\tilde{u}}(t|t) \quad . \tag{4.66}$$

Equations (4.58), (4.59), (4.64), ..., (4.66) determine completely the prediction and filtering problem for u(t), whereas the solution for x(t) can be found using the equations (4.57).

Consider finally the Kalman-Bucy problem, where the state x(t) is generated by a linear differential equation

$$\begin{aligned} \frac{d}{dt}\,x(t) &= A(t)x(t) + B(t)w(t) \\ x(o) &= x^o \end{aligned} \tag{2.15}$$

driven by Gaussian white noise w(t) independent of the observation noise. As remarked before, x(t) is then a wide-sense Markov process and can be written as

$$x(t) = \phi(t,o)\left\{x^o + \int_0^t \phi^{-1}(s,o)B(s)w(s)ds\right\} \tag{4.67}$$

where the state transition matrix $\phi(t,o)$ satisfies the matrix differential equation

$$\frac{d}{dt}\,\phi(t,o) = A(t)\phi(t,o) \quad , \quad \phi(o,o) = I \quad . \tag{4.68}$$

If the process u(t) is defined by

$$u(t) = x^o + \int_o^t \phi^{-1}(s,o)B(s)w(s)ds$$
$$u(o) = x^o \qquad (4.69)$$

then the state x(t) can be written in the form $x(t) = \phi(t,o)u(t)$ where u(t) is a (continuous time) wide-sense martingale.

It satisfies the differential equation

$$\frac{d}{dt} u(t) = \phi^{-1}(t,o)B(t)w(t)$$
$$u(o) \quad = x^o \qquad (4.70)$$

driven by white noise w(t) and is therefore, in the absence of system noise simply a constant, i.e. it is equal then for every $t, o \leqslant t \leqslant 1$ to the initial condition x^o.

Hence, to find the linear least-squares filtered estimate $\hat{x}(t|t)$ of x(t) given by (2.15) based on the observations one can either use the Kalman-Bucy filter or the wide-sense martingale approach. It follows from the definition of u(t) for the Kalman-Bucy problem that the derivative of the covariance matrix $P_u(t)$ in the differential equation for the error covariance matrix $P_{\tilde{u}}(t|t)$ is given by

$$\frac{d}{dt} P_u(t) = \phi^{-1}(t,o)B(t)P_w(t)B(t)^T\phi^{-1}(t,o)^T \quad . \qquad (4.71)$$

Hence, the u(t) filtering equations of the wide-sense martingale approach for the Kalman-Bucy problem are:

$$\frac{d}{dt} \hat{u}(t|t) = K(t) \left\{\frac{d}{dt} Y(t) - C(t)\hat{u}(t|t)\right\} \qquad (4.72)$$

$$\frac{d}{dt} P_{\tilde{u}}(t|t) = \phi^{-1}(t,o)B(t)P_w(t)B(t)^T\phi^{-1}(t,o)^T +$$
$$- P_{\tilde{u}}(t|t)H(t)^TR^{-1}(t)H(t)P_{\tilde{u}}(t|t) \qquad (4.73)$$

$$K(t) = P_{\tilde{u}}(t|t)H(t)^TR^{-1}(t) \quad . \qquad (4.74)$$

These have to be supplemented with equations (4.57) to yield estimates for the state x(t). A comparison of the u(t) filter with the Kalman-Bucy filter as given in chapter 2.2 clearly reveals the effect of the transformation $x(t) = \phi(t,o)u(t)$. The original system dynamics is brought into a form with $A(t) \equiv 0$ which causes some terms in the Kalman-Bucy filter to disappear. On the other hand, the remaining terms are made more complicated.

5 APPLICATION OF WIDE-SENSE MARTINGALE APPROACH TO SOME MODEL PROBLEMS

In the preceeding chapter the wide-sense martingale approach has been described. It has also been shown that the Kalman and Kalman-Bucy problem can be brought into the form considered there by solving the state equation and writing $x(t) = \phi(t,o)u(t)$ where $\phi(t,o)$ is the state transition matrix. Therefore, the Kalman and Kalman-Bucy problem can be solved by estimating the state $x(t)$ directly (in this chapter also called $x(t)$ filter) or by estimating $u(t)$ and transforming these estimates into estimates for $x(t)$ (in this chapter called $u(t)$ filter). To compare both approaches a couple of model problems has been simulated and solved with each algorithm. These model problems together with some elucidating comment on the computational aspects are discussed in this chapter.

5.1 Model problem one

Consider the linear least-squares estimation of the state $x(t)$ of a system based on observations $y(k)$ where the state is given by

$$\begin{aligned} \frac{d}{dt}\, x(t) &= A(t)x(t) + B(t)w(t) \\ x(o) &= x^o \end{aligned} \tag{5.1}$$

$$x(t) = \begin{pmatrix} x_1(t) \\ \\ x_2(t) \end{pmatrix} \quad , \quad x^o = \begin{pmatrix} x_1^o \\ \\ x_2^o \end{pmatrix} \quad , \quad w(t) = \begin{pmatrix} w_1(t) \\ \\ w_2(t) \end{pmatrix}$$

$$A(t) = \begin{pmatrix} o & -\omega_o \\ \\ \omega_o & o \end{pmatrix} \quad , \quad B(t) = \begin{pmatrix} b_{11}(t) & b_{12}(t) \\ \\ b_{21}(t) & b_{22}(t) \end{pmatrix} . \tag{5.2}$$

The state equation (5.1) is driven by Gaussian white noise $w(t)$. The state transition matrix $\phi(t,o)$ of the homogeneous system

$\frac{d}{dt}\, x(t) = A(t)x(t)$ is

$$\phi(t,o) = \begin{pmatrix} \cos \omega_o t & -\sin \omega_o t \\ \\ \sin \omega_o t & \cos \omega_o t \end{pmatrix} \tag{5.3}$$

and the state equation (5.1) describes the motion of a second-order oscillator perturbed by noise. The state $x(t)$ can be written then as

$$x(t) = \phi(t,o)\left\{x^o + \int_o^t \phi^{-1}(s,o)B(s)w(s)ds\right\} \tag{5.4}$$

or

$$x(t) = \phi(t,o)u(t) \tag{5.5}$$

where

$$\begin{aligned} u(t) &= x^o + \int_o^t \phi^{-1}(s,o)B(s)w(s)ds \\ u(o) &= x^o \quad . \end{aligned} \tag{5.6}$$

It follows that $u(t)$ is a wide-sense martingale if x^o is uncorrelated with the white noise $w(t)$ for every t. In differential form $u(t)$ satisfies

$$\frac{d}{dt}\, u(t) = \phi^{-1}(t,o)B(t)w(t) \quad . \tag{5.7}$$

Assume that scalar observations $y(k)$ are given at discrete, not necessarily equidistant, points of time by

$$y(k) = C(k)x(k) + v(k) \tag{5.8}$$

or

$$y(k) = H(k)u(k) + v(k) \tag{5.9}$$

where

$$C(k) = \Big(c_1(k) \;\; c_2(k)\Big) \quad , \; H(k) = C(k)\phi(k,o) \tag{5.10}$$

and $v(k)$ is zero mean Gaussian white noise with $\mathcal{E}\{v(k)v(\ell)\} = P_v(k)\delta_{k\ell}$ and uncorrelated with x^o and the system noise $w(t)$. Notice that $x(k)$ and $u(k)$ are used to denote x and u at the time t_k of the k-th observation.

As described in chapter four, the wide-sense martingale approach consists of two steps:

1) Processing of the observations $y(k)$ $k = 1,2,\ldots$ to obtain estimates $\hat{u}(k|k)$ of the signal $u(k)$

2) Computation of estimates $\hat{x}(k|k)$ of the state $x(k)$ using the relations

$$\hat{x}(k|k) = \phi(k,o)\hat{u}(k|k) \tag{5.11}$$

$$P_{\tilde{x}}(k|k) = \phi(k,o)P_{\tilde{u}}(k|k)\phi(k,o)^T \quad . \tag{5.12}$$

The Kalman filter yields $\hat{x}(k|k)$ directly. Notice that in the absence of system noise the filtered estimate $\hat{u}(k|k)$ is exactly the least-squares estimate of the initial state x^o given the observations $y(i)$, $i = 1,2,\ldots, k$. Hence, the wide-sense martingale gives the smoothed least-squares estimate of the initial condition x^o from which the filtered estimate $\hat{x}(k|k)$ is computed using (5.11).

The filtering problem for the continuous time signal $x(t)$ with observations available at discrete points of time is embedded in a discrete time filtering problem by integrating the system dynamics between successive observations. Hence,

$$x(k+1) = \phi(k+1,k)x(k) + n(k) \tag{5.13}$$

where

$$n(k) = \int_{t_k}^{t_{k+1}} \phi(t_{k+1},s)B(s)w(s)ds \quad . \tag{5.14}$$

Assuming that the system noise $w(t)$ is constant, $w(t) \equiv w(k)$, between successive observations, it follows that

$$x(k+1) = \phi(k+1,k)x(k) + E(k)w(k) \tag{5.15}$$

and

$$u(k+1) = u(k) + F(k)w(k) \tag{5.16}$$

where

$$E(k) = \int_{t_k}^{t_{k+1}} \phi(t_{k+1}s)B(s)ds \tag{5.17}$$

$$F(k) = \int_{t_k}^{t_{k+1}} \phi^{-1}(s,o)B(s)ds \quad . \tag{5.18}$$

The Kalman filter described in chapter 2.2 can be applied to the system described by equation (5.15) and the observations given in (5.8) if $P_w(k) = \mathcal{E}\left\{w(k)w(k)^T\right\}$ is given. The wide-sense martingale algorithm (4.12), ..., (4.16) can be applied to the system described by equation (5.16) and the observations given in (5.9). Equation (4.16) for the prediction of the covariance matrix of the estimation error may be replaced by

$$P_{\tilde{u}}(k+1|k) = P_{\tilde{u}}(k|k) + F(k)P_w(k)F(k)^T \quad . \qquad (5.19)$$

At this point some remarks on the difference, from a computational point of view between the Kalman and the wide-sense martingale approach can be made already. It is clear, for the simplest case in which there is no system noise, that the prediction from t_k to t_{k+1} for the wide-sense martingale u is trivial. The prediction step for the Kalman filter requires computation of the state transition matrix $\phi(k+1,k)$ and matrix multiplications for the estimate and its covariance matrix. Notice that for this particular model problem the matrix $\phi(k+1,k)$ depends only on the time difference between observations. In case of equidistant observation times it needs therefore to be computed only once. The same holds for the matrix E(k) for the case with system noise provided that the system matrix B(t) is constant. However, even in the latter case some form of computation is needed for the matrix F(k) (5.18) in the wide-sense martingale approach. Moreover, the inverse of the state transition matrix is required there. As for the measurement update, the algorithm of the wide-sense martingale approach needs the modified observation matrix $H(k) = C(k)\phi(k,o)$ instead of C(k). Thus, the wide-sense martingale approach requires the computation of the transition matrix $\phi(k,o)$ for the measurement update whereas the Kalman approach needs $\phi(k+1,k)$ in the prediction step.

Clearly, the wide-sense martingale approach for the computation of the filtered estimates $\hat{u}(k|k)$ needs fewer matrix multiplications than the Kalman approach. However, it must be realized that the former yields estimates of the wide-sense martingale u and that the transformation by way of (5.11) and/or (5.12) is required to yield the corresponding results for the state x. The use of this transformation after each measurement update will decrease the possible gain in computation time of the wide-sense martingale approach considerably. Apparently, one may expect the wide-sense martingale approach to be more efficient if the transformation (5.11) and in particular (5.12) is not required after each measurement update. The need of the transformation (5.11) and/or (5.12) will generally depend on the problem at hand. For example, in case of a relatively high measurement rate it might not be necessary to use (5.11) and/or (5.12) more than, say, once per ten measurement updates.

Both algorithms have been tested on a Cyber 72 digital computer. The dynamical system (5.15) has been simulated, with as well as without system noise, together with the observations (5.8). Numerical parameters used are:

time between observations	$t_{k+1}-t_k = 0.1$ s
	$\omega_o = 0.628$ rad/s
observation matrix	$C(k) = (1.0 \quad 2.0)$
observation noise variance	$P_v(k) = (0.9)^2$
initial state	$x^o = \begin{pmatrix} 2.0 \\ 1.0 \end{pmatrix}$

Additional values used in case of system noise are:

system noise covariance matrix	$P_w(k) = \begin{pmatrix} (0.02)^2 & 0 \\ 0 & (0.02)^2 \end{pmatrix}$
system noise coefficient matrix	$B(t) = \begin{pmatrix} 1.0 & 1.0 \\ 1.0 & 1.0 \end{pmatrix}$
observation noise variance	$P_v(k) = (0.3)^2$

In all cases 1000 observations have been processed and the following initial estimates have been used:

$$\hat{u}(o|o) = \hat{x}(o|o) = \begin{pmatrix} 1.5 \\ 0.5 \end{pmatrix}, \; P_{\tilde{u}}(o|o) = P_{\tilde{x}}(o|o) = \begin{pmatrix} (0.8)^2 & 0 \\ 0 & (0.8)^2 \end{pmatrix}.$$

The results with regard to computation time are presented in tables 1 and 2 where x(t) filter denotes the Kalman filter and u(t) filter the wide-sense martingale algorithm inclusive of the transformation (5.11) and (5.12) after the specified number of measurement updates. It can be read from table 1 that without system noise the u(t) filter needs 7.2 % less computation time than the x(t) filter if both the state estimate and the corresponding error covariance matrix are computed from the martingale estimates after every ten observations. The gain in computation time rises to 14.5 % if the transformation of the error covariance matrix is omitted completely. For the latter case an improvement to 16.0 % is obtained if the transformation of the state is done only after every 20 observations. For the x(t) filter the state transition matrix $\phi(k+1,k)$ has been calculated for each measurement update as it would be required for the case of non-equidistant measurement times. For the present example it was found that the x(t) filter needed 6.100 seconds execution time if $\phi(k+1,k)$ was computed only once.

It is interesting to compare the filtering results. Figure 1 gives for the x(t) filter the estimates $\hat{x}_1(k|k)$ and $\hat{x}_2(k|k)$ which, apart from the first few hundred observations coincide with the simulated states $x_1(k)$ and $x_2(k)$. The oscillatory character of the state is clear. The standard deviation $\sigma_{\tilde{x}_1}(k|k)$ of the estimation error $\tilde{x}_1(k|k) = x_1(k) - \hat{x}_1(k|k)$ is presented in figure 2 and shows the expected relatively fast decay during the first part of the filtering process. Figure 3 presents the correlation coefficient between $\hat{x}_1(k|k)$ and $\hat{x}_2(k|k)$. In figure 4 the filtered estimates $\hat{u}_1(k|k)$ and $\hat{u}_2(k|k)$ are shown. From this figure the convergence of the filtered estimates to the exact initial condition is clear. The standard deviation $\sigma_{\tilde{u}_1}(k|k)$ of the estimation error $\tilde{u}_1(k|k) = u_1(k) - \hat{u}_1(k|k)$ is given in figure 5 and the correlation coefficient between $\hat{u}_1(k|k)$ and $\hat{u}_2(k|k)$ in figure 6. In figure 7 the dotted lines correspond with $\hat{x}_1(k|k)$ and $\hat{x}_2(k|k)$ as they have been computed from $\hat{u}_1(k|k)$ and $\hat{u}_2(k|k)$ using 5.11. The result of the transformation has been plotted after every 30 observations. These estimates were found to coincide exactly with the x(t) filter results. Finally, figure 8 shows their correlation coefficient.

Consider now the results for the case with system noise. For both the u(t) and x(t) filter the contribution of the system noise to the predicted covariance matrices $P(k+1|k)$ has been computed for each observation. Thus, for the x(t) filter the execution time found is representative for the case of non-equidistant measurement times. It is seen from table 2 that the gain in computation time for the u(t) filter has decreased to 2.5 % or 8.7 % depending on the transformation of the error covariance matrix. The filtering results of this case are shown in figures 9 to 19. In figure 9 the simulated $x_1(t)$ is presented; as a reference also the $x_1(t)$ without system noise is shown there. Figure 10 gives the filtering results of the x(t) filter whereas figure 11 shows the corresponding standard deviation. It is seen that this decreases faster, initially, then for the case without system noise.

This is caused by the fact that the measurement noise standard deviation has been taken much lower here. This overcompensates the uncertainty added to the state by the system noise. The correlation coefficient is shown in figure 12. In figure 13 the filtering results $\hat{u}_1(t|t)$ and $\hat{u}_2(t|t)$ are shown globally and in figures 14 and 15 in more detail. It is clear that due to the system noise the exact u(t) signal is no longer constant. The figures 16 and 17 give a comparison of the exact and estimated u(t) values. Figure 18 shows the course of the standard deviation $\sigma_{\tilde{u}_1}(k|k)$. The figure indicates that this standard deviation, due to the effect of system noise does not decrease monotonically. This is shown even more clearly in the enlargement of figure 19.

5.2 Model problem two

The preceeding problem was rather simple in that the state transition matrix was known analytically and the state vector consisted of only two elements. Consider now the more complicated case where the state vector has four elements, $x(t) = \left(x_1(t)x_2(t)x_3(t)x_4(t)\right)^T$ and satisfies the differential equation

$$\begin{aligned} \frac{d}{dt}\, x(t) &= Ax(t) \\ x(o) &= x^o \end{aligned} \tag{5.20}$$

where

$$A = \begin{pmatrix} -0.40 & 0.10 & 0.20 & 0.50 \\ 0.10 & -1.00 & 0.40 & 0.25 \\ 0.20 & 0.40 & -1.00 & -0.10 \\ 0.50 & 0.25 & -0.10 & -1.00 \end{pmatrix}, \quad x^o = \begin{pmatrix} 1.5 \\ 1.8 \\ 2.0 \\ 2.5 \end{pmatrix} \tag{5.21}$$

This model problem has no direct physical interpretation. Because the eigenvalues of the matrix A are all real and less than zero, the solution of the system (5.20) and (5.21) tends to zero as t increases. The state transition matrix $\phi(t,o)$ of the system 5.20 satisfies the differential equation

$$\begin{aligned} \frac{d}{dt}\, \phi(t,o) &= A\phi(t,o) \\ \phi(o,o) &= I \text{ (unit matrix)} \end{aligned} \tag{5.22}$$

and can be written as

$$\phi(t,o) = \sum_{i=0}^{\infty} \frac{A^i}{i!}(t-o)^i \quad . \tag{5.23}$$

For small values of t it can be approximated by

$$\phi(t,o) = I + At + \frac{A^2t^2}{2!} \tag{5.24}$$

or

$$\phi(t,o) = I + At + \frac{A^2t^2}{2!} + \frac{A^3t^3}{3!} + \frac{A^4t^4}{4!} \quad . \tag{5.25}$$

For larger t values the system (5.22) can be integrated numerically to yield $\phi(t,o)$. Alternatively, $\phi(t)$ can be computed as a product of state transition matrices $\phi(i,i-1)$ where $t_i - t_{i-1}$ is small enough to use (5.24) or (5.25) for $\phi(i,i-1)$.

It is wanted again to find the linear least-squares estimate of $x(t)$ based on the observations

$$y(k) = C(k)x(k) + v(k) \tag{5.8}$$

where

$$C(k) = (1.0 \quad 1.0 \quad 1.0 \quad 1.0)$$

and $v(k)$ is zero mean Gaussian white noise uncorrelated with the initial state x^o. Both the Kalman approach and the wide-sense martingale approach are used to solve this problem and a comparison of their results is made. In the Kalman approach the state transition matrix $\phi(k+1,k)$ is required for the prediction of the state at the time t_{k+1} of the next observation. In the wide-sense martingale approach the transition matrix $\phi(t,o)$ is used again to transform the state $x(t)$ into a wide-sense martingale $u(t)$, $x(t) = \phi(t,o)u(t)$. As a consequence, the observations take the form

$$y(k) = H(k)u(k) + v(k) \tag{5.9}$$

where $H(k) = C(k)\phi(k,o)$. One thousand observations have been simulated at time intervals of 0.05 seconds. For the Kalman approach $\phi(k+1,k)$ has been computed for each measurement update using equation (5.25). It has been assumed therefore that the measurement times could be non-equidistant but the matrix products $A^i (1 \leq i \leq 4)$ have been computed only once. For the wide-sense martingale approach $\phi(k+1,o) = \prod_{i=0}^{k} \phi(i+1,i)$ has been used with $\phi(i+1,i)$ again according to equation (5.25).

Table 3 gives a summary of the computation time required by both algorithms. It is clear that due to the increased number of elements in the state vector the total execution time has approximately doubled. The wide-sense martingale approach is 7.7 % faster than the Kalman approach if the transformation (5.11) and (5.12) is applied after every 10 observations for both the state estimate and the error covariance matrix. If the latter is omitted completely, a gain in computation time of 13.8 % is obtained. These results are approximately the same as for the first model problem.

Consider finally the estimation results. In figures 20 up to 23 inclusive, the four simulated elements of the state vector are plotted together with their estimated values using the Kalman approach. The similarity is good. The corresponding standard deviations are shown in figures 24 and 25. Figure 26 shows the estimates $\hat{u}_1$ and $\hat{u}_2$ converging to the exact initial values 1.5 and 1.8 respectively. In figure 27 the estimates $\hat{u}_3$ and $\hat{u}_4$ are shown, together with the exact values $u_3 = 2.0$ and $u_4 = 2.5$. The corresponding standard deviations are shown in figures 28 and 29. Although the achieved accuracy of the estimates of the wide-sense martingale u is not very high, the filtering results are clearly consistent. Finally, in figures 30 and 31 the exact values and the estimation results for x_1 and x_4 using the wide-sense martingale approach and the transformation (5.11) are shown. The drawn curves represent the exact values whereas the top of each peak corresponds with an estimated value. The result of the transformation has been plotted for every 30 observations in figures 30 and 31. From the figures it seems that there is no difference in the Kalman results and the wide-sense martingale approach results for the state x(t). A careful analysis of the numerical results confirmed this. The transformed standard deviations were also found to be equal to the Kalman filter results.

6 CONCLUSIONS

A study of literature on martingales in estimation theory has been carried out, guided by the intent to consider the implications for the present day practical filtering work at NLR. Attention has been directed at the recursive estimation equations of Kalman and Kalman-Bucy. Some experimental work concerning the efficiency of an estimation algorithm based on martingales is also described.

Martingales are stochastic processes with a specific structure defined in terms of given increasing information fields. It follows from the definition that the increments of a martingale have a property which is intermediate between the properties of independence and uncorrelatedness. With respect to estimation, a number of theoretical results concerning martingales and linear as well as non-linear estimation are summarized in the form of an innovation and representation theorem. These are useful in that they provide insight in the structure of the estimation problem. Using them, a simple derivation of the Kalman and Kalman-Bucy filter has been sketched. The knowledge and insight gained during the study are also important because they offer the possibility to study more general estimation problems of the types described in some of the references (Refs 5, 31, 32, 34, 37) and, probably, to attack practical estimation problems at NLR which at present cannot be handled. It is therefore of importance both to study the more general estimation problems and to make up an inventory of them at NLR.

Finally, "a wide-sense martingale approach" has been analyzed. This approach yields recursive linear estimation equations for the class of state processes $x(t)$ modelled by $x(t) = \phi(t)u(t)$ where $u(t)$ is a wide-sense martingale with known covariance and $\phi(t)$ is a known matrix function of time. The observations may be corrupted by either white or coloured noise. The estimation equations hold for the wide-sense martingale $u(t)$ but, due to the linearity of the state model, the estimates of the actual state $x(t)$ can be directly computed. The new estimation equations are more general than the Kalman and Kalman-Bucy filters in the sense that the state models on which the latter are based can be transformed into the above form. It was claimed in the literature that this approach applied to the Kalman and Kalman-Bucy problem would be more efficient. This claim has been investigated in this report by means of numerical simulation of some model problems. The wide-sense martingale approach has been found to be more efficient indeed, but the amount of gain is highly dependent on the number of times the estimates of the actual state $x(t)$ are computed from the estimates of the wide-sense martingale $u(t)$.

7 REFERENCES

1 Bagchi, A. — A martingale approach to continuous time linear smoothing. SIAM J. Appl. Math. Vol.10, No.2, 1975, pp 276-281.

2 Bagchi, A. — A new martingale approach to Kalman filtering. Information Sciences, Vol.10, 1976, pp 187-192.

3 Balakrishnan, A.V. — A martingale approach to linear recursive state estimation. SIAM J. Control., Vol.10, No.4, November 1972, pp 754-766.

4 Beutler, F.J. — Multivariate wide-sense Markov processes and their prediction theory. Ann. Math. Statist., Vol.34, 1963, pp 424-438.

5 Boel, R. — Jump processes in filtering and detection. In: Communication systems and random process theory. Skwirzynski, J.K. (ed.), pp 777-787, 1978. Sijthoff and Noordhoff, Alphen aan de Rijn, Netherlands.

6 Bryson, A.E., Henrikson, L.J. — Estimation using sampled data containing sequentially correlated noise. J. of Spacecraft, Vol.5, No.6, 1968, pp 662-665.

7 Clark, J.M.C. — The representation of functionals of Brownian motion by stochastic integrals. Ann. Math. Statist., Vol.41, No.4, 1970, pp 1282-1295.

8 Doob, J.L. — Stochastic processes. John Wiley and Sons, Seventh printing 1967.

9 Frost, P.A., Kailath, T. — An innovations approach to least-squares estimation. Part III: Nonlinear estimation in white Gaussian noise. IEEE Trans. on Automatic Control, Vol.AC-16, No.13, June 1971, pp 217-226.

10 Fujisaki, M., Kallianpur, G., Kunita, H. — Stochastic differential equations for the nonlinear filtering problem. Osaka J. Math., Vol.9, No.2, 1972, pp 19-40.

11 Jazwinski, A.H. — Stochastic processes and filtering theory. Academic Press, 1970.

12 Kailath, T. — An innovations approach to least squares estimation.
Part I: Linear filtering in additive white noise.
IEEE Trans. on Automatic Control, Vol.AC-13, No.6, December 1968, pp 646-655.

13 Kailath, T., Frost, P.A. — An innovations approach to least squares estimation.
Part II: Linear smoothing in additive white noise.
IEEE Trans. on Automatic Control, Vol.AC-13, No.6, December 1968, pp 655-660.

14 Kailath, T. — A general likelihood ratio formula for random signals in Gaussian noise.
IEEE Trans. Inform. Theory, Vol.IT-15, No.3, May 1969, pp 350-361.

15 Kailath, T. — Some extensions of the innovations theorem.
The Bell System Technical Journal, Vol.50, No.4, April 1971, pp 1487-1494.

16 Kailath, T., Geesey, R.A. — An innovations approach to least squares estimation.
Part IV: Recursive estimation given lumped covariance functions.
IEEE Trans. on Autom. Control, Vol.AC-16, No.6, December 1971, pp 720-726

17 Kailath, T. — A note on least squares estimation by the innovations method.
SIAM J. Control, Vol.10, No.3, August 1972, pp 477-486.

18 Kailath, T., Geesey, R.A. — An innovations approach to least squares estimation.
Part V: Innovations representations and recursive estimation in coloured noise.
IEEE Trans. on Automatic Control, Vol.AC-18, No.5, October 1973, pp 435-453.

19 Kalman, R.E. — A new approach to linear filtering and prediction problems.
Journal of Basic Engineering, Vol.82, March 1960, pp 35-45.

20 Kalman, R.E., Bucy, R.S. — New results in linear filtering and prediction theory.
Journal of basic engineering, Vol.83, March 1961, pp 95-107.

21 Kara, H. — Wide sense martingale approach to linear discrete time optimal estimation.
Ph.D. dissertation, Michigan State University, East Lansing, 1971.

22 Kara, H.
Mandrekar, V.
Park, G.L.
Wide sense martingale approach to linear optimal estimation.
SIAM J. Appl. Math., Vol.27, No.2, September 1974, pp 293-302.

23 Kunita, H.
Watenabe, S.
On square integrable martingales.
Nagoya J. Math., Vol.30, 1967, pp 209-245.

24 Lee, K.Y.L.
Mandrekar, V.
Multiplicity and martingale approach to linear state estimation.
Proceedings of the seventh annual Princeton conference on information sciences and systems.

25 Lipster, R.S.
Shiryayev, A.N.
Statistics of random processes.
I. General theory.
Springer Verlag, Berlin, 1978.

26 Mandrekar, V.
On multivariate wide-sense Markov processes.
Nagoya Math. J., Vol.33, 1968, pp 7-19.

27 Nelson, E.
Dynamical theories of Brownian motion.
Princeton University Press, Princeton, 1967.

28 Ouvrard, J.Y.
Martingale projection and linear filtering in Hilbert spaces.
I: The theory.
SIAM J. Control and Opt., Vol.16, No.6, November 1978, pp 912-937.

29 Ouvrard, J.Y.
Linear filtering in Hilbert spaces.
II: An application to the smoothing theory for hereditary systems with observation delays.
SIAM J. Control and Opt. Vol.16, No.6, November 1978, pp 938-953.

30 Prasad, T.
Sinha, N.K.
Martingales and their application to optimal state estimation.
Int. J. Systems Sci., Vol.5, No.11, 1974, pp 1039-1054.

31 Segall, A.
Kailath, T.
The modelling of randomly modulated jump processes.
IEEE Trans. on Information Theory, Vol.IT-21, No.2, March 1975, pp 135-143.

32 Segall, A.
Davis, M.H.A.
Kailath, T.
Nonlinear filtering with counting observations.
IEEE Trans. on Information Theory, Vol.IT-21, No.2, March 1975, pp 143-149.

33 Segall, A.
Stochastic processes in estimation theory.
IEEE Trans. on Information Theory, Vol.IT-22, No.3, May 1976, pp 275-286.

34 Segall, A. — Recursive estimation from discrete-time point processes.
IEEE Trans. on Information Theory, Vol.IT-22, No.4, July 1976, pp 422-431.

35 Traas, C.R. — Estimation of modelling errors in satellite attitude determination.
IAF-76-026; 27-th congress of IAF, Anaheim, October 1976.

36 Traas, C.R.
Moek, G. — Processing of Meteosat attitude measurement flight data and evaluation of the performance of an adaptive filtering technique.
NLR TR 78146 U, 1978.
National Aerospace Lab., Amsterdam.

37 Varaiya, P. — The martingale theory of jump processes.
IEEE Trans on Automatic Control, Vol.AC-20, No.1, February 1975, pp 34-42.

38 Wong, E. — Recent progress in stochastic processes- a survey.
IEEE Trans. Inform. Theory, Vol.IT-19, No.3, May 1973, pp 262-275.

39 - — Are Ito calculus and martingale theory useful in practice?
In: Communication systems and random process theory. Skwirzynski, J.K. (ed.) pp 777-787, 1978.
Sijthoff and Noordhoff, Alphen aan de Rijn, Netherlands.

TABLE 1

Comparison of execution time for x(t) and u(t) filter for second-order oscillator (model problem one) without system noise

	seconds execution time	transformation 5.11/5.12 used after every N observations	
		state vector (5.11) N	covariance matrix (5.12) N
x(t) filter	6.381		
u(t) filter	5.920	10	10
u(t) filter	5.428	10	-
u(t) filter	5.362	20	-

TABLE 2

Comparison of execution time for x(t) and u(t) filter for second-order oscillator (model problem one) with system noise

	seconds execution time	transformation 5.11/5.12 used after every N observations	
		state vector (5.11) N	covariance matrix (5.12) N
x(t) filter	8.254		
u(t) filter	8.046	10	10
u(t) filter	7.535	10	-

TABLE 3

Comparison of execution time for x(t) filter and u(t) filter for model problem two

	seconds execution time	transformation 5.11/5.12 used after every N observations	
		state vector (5.11) N	covariance matrix (5.12) N
x(t) filter	14.650		
u(t) filter	13.528	10	10
u(t) filter	12.631	10	-

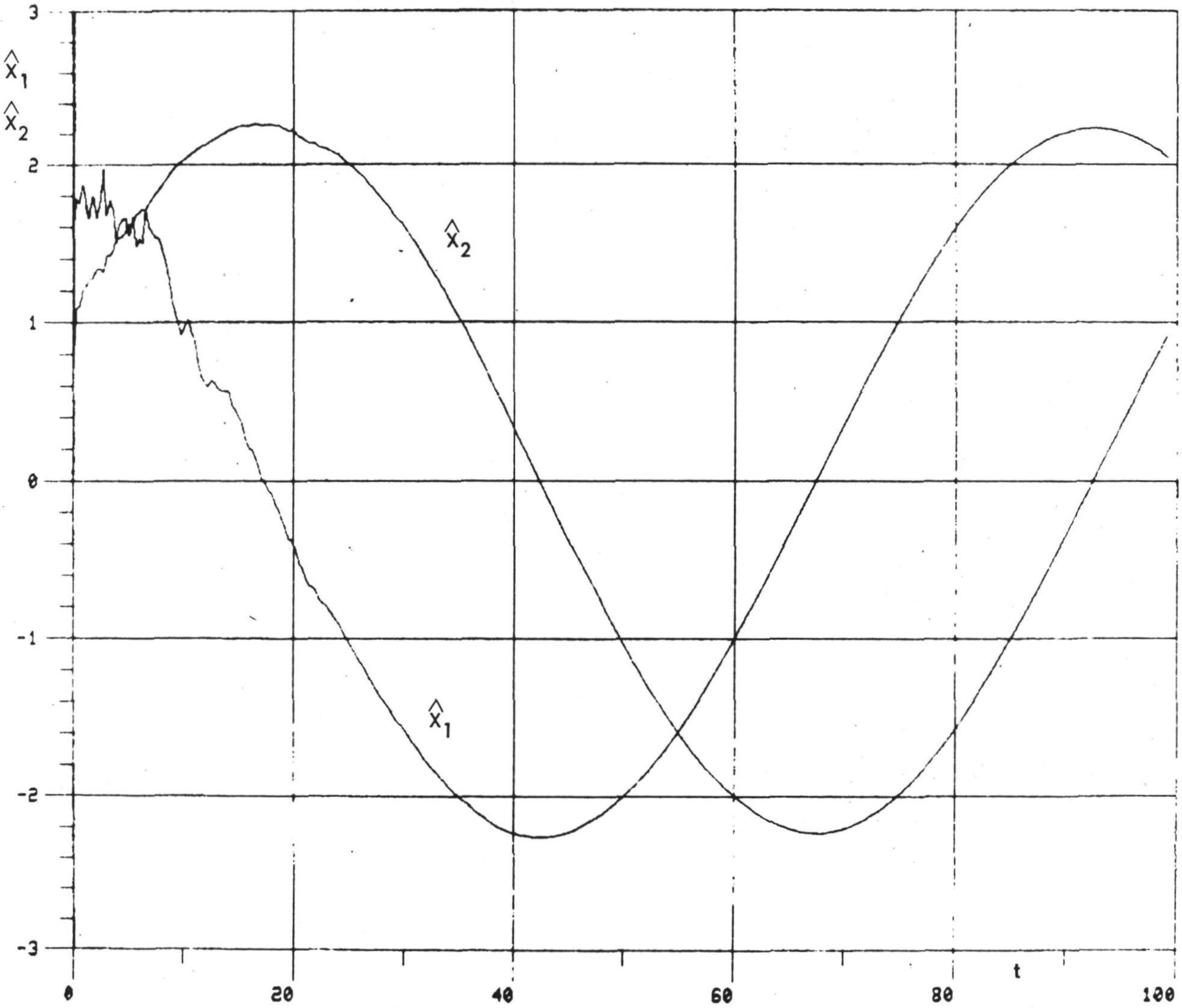

Fig. 1 Estimated components of state vector x using Kalman filter (model problem 1; no system noise)

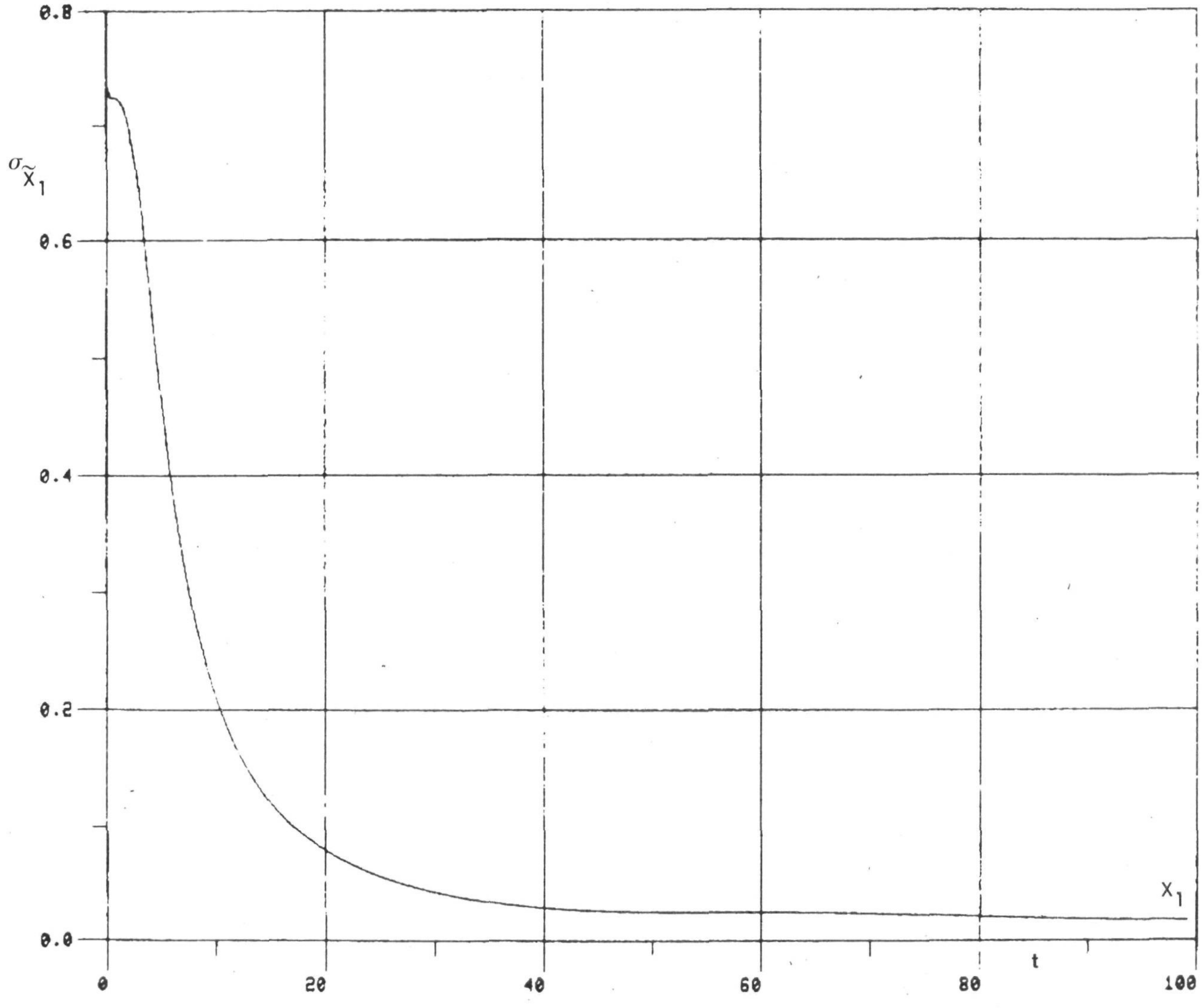

Fig. 2 Estimated standard deviation of state vector element x_1 using Kalman filter (model problem 1; no system noise)

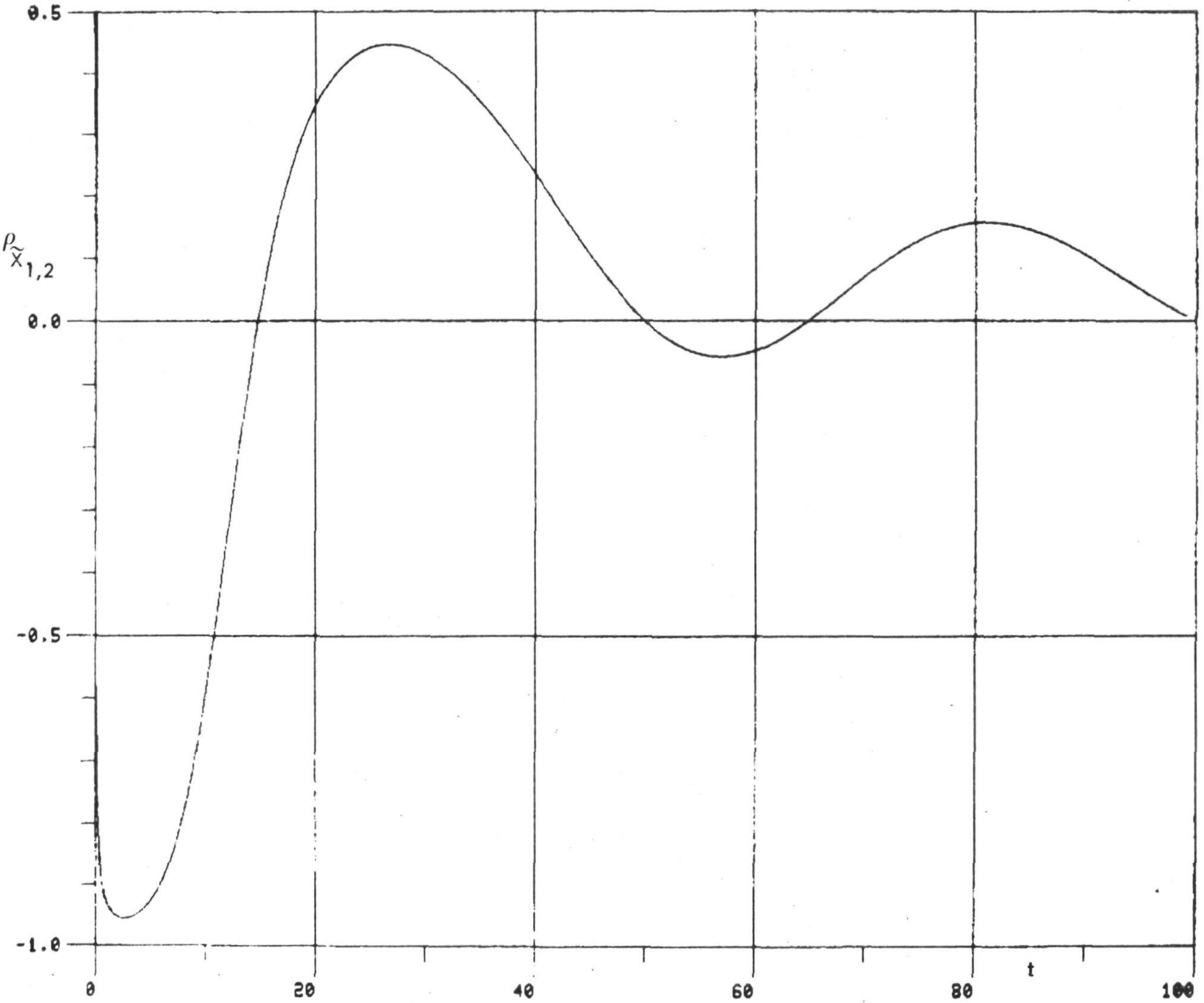

Fig. 3 Estimated correlation coefficient between state vector elements using Kalman filter (model problem 1; no system noise)

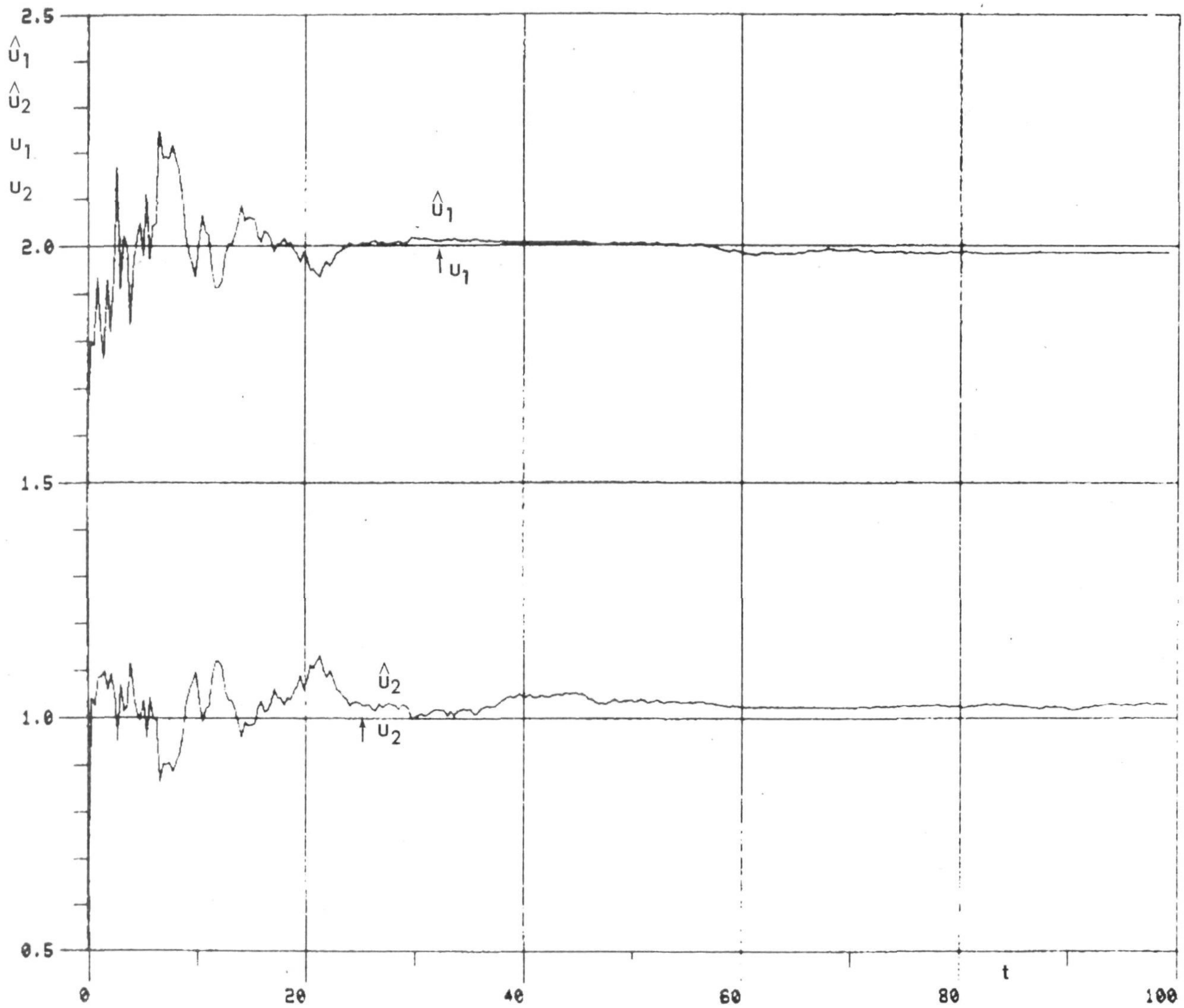

Fig. 4 Exact and estimated components of wide-sense martingale u (model problem 1; no system noise)

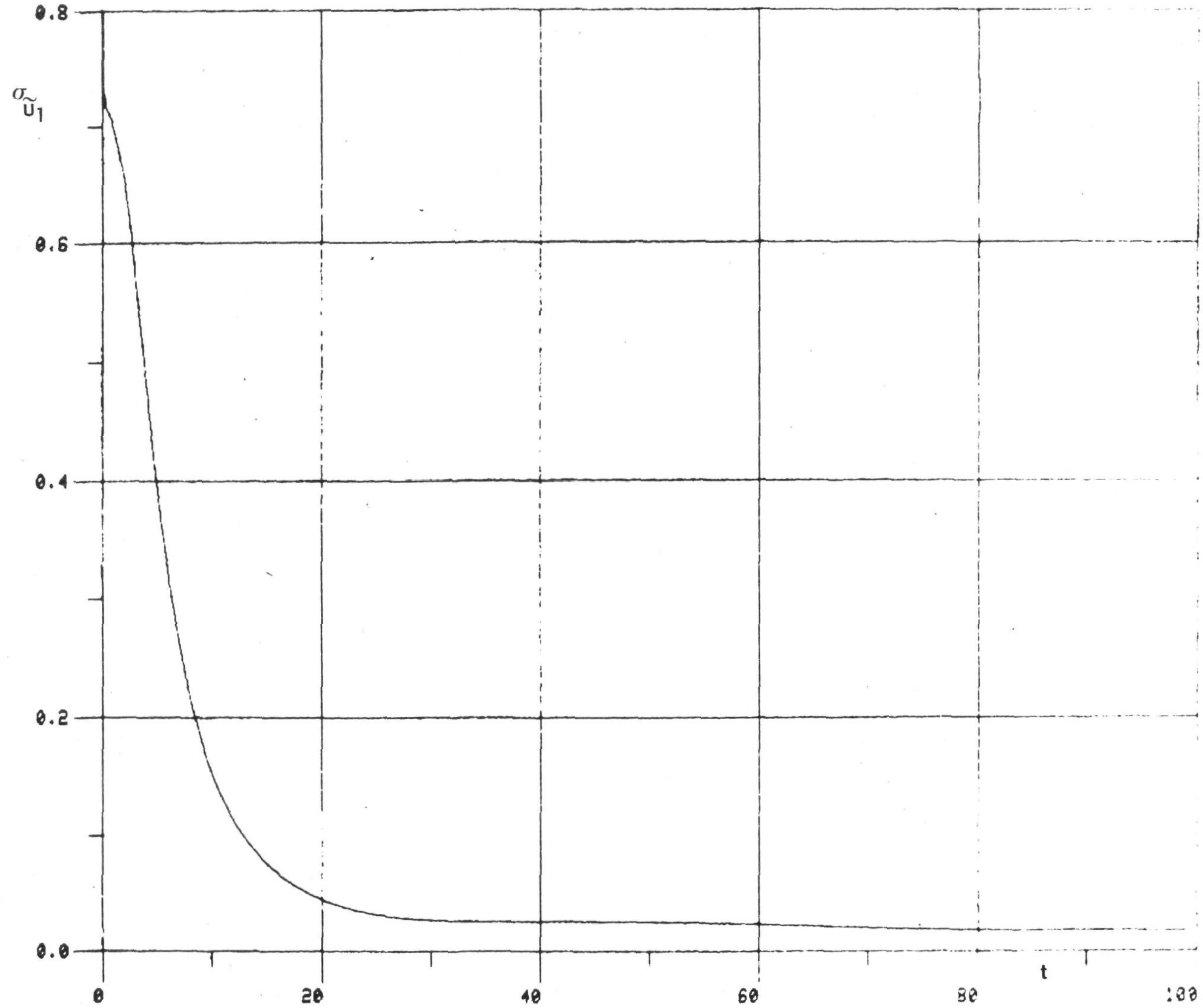

Fig. 5 Estimated standard deviation of first component of wide-sense martingale u (model problem 1; no system noise)

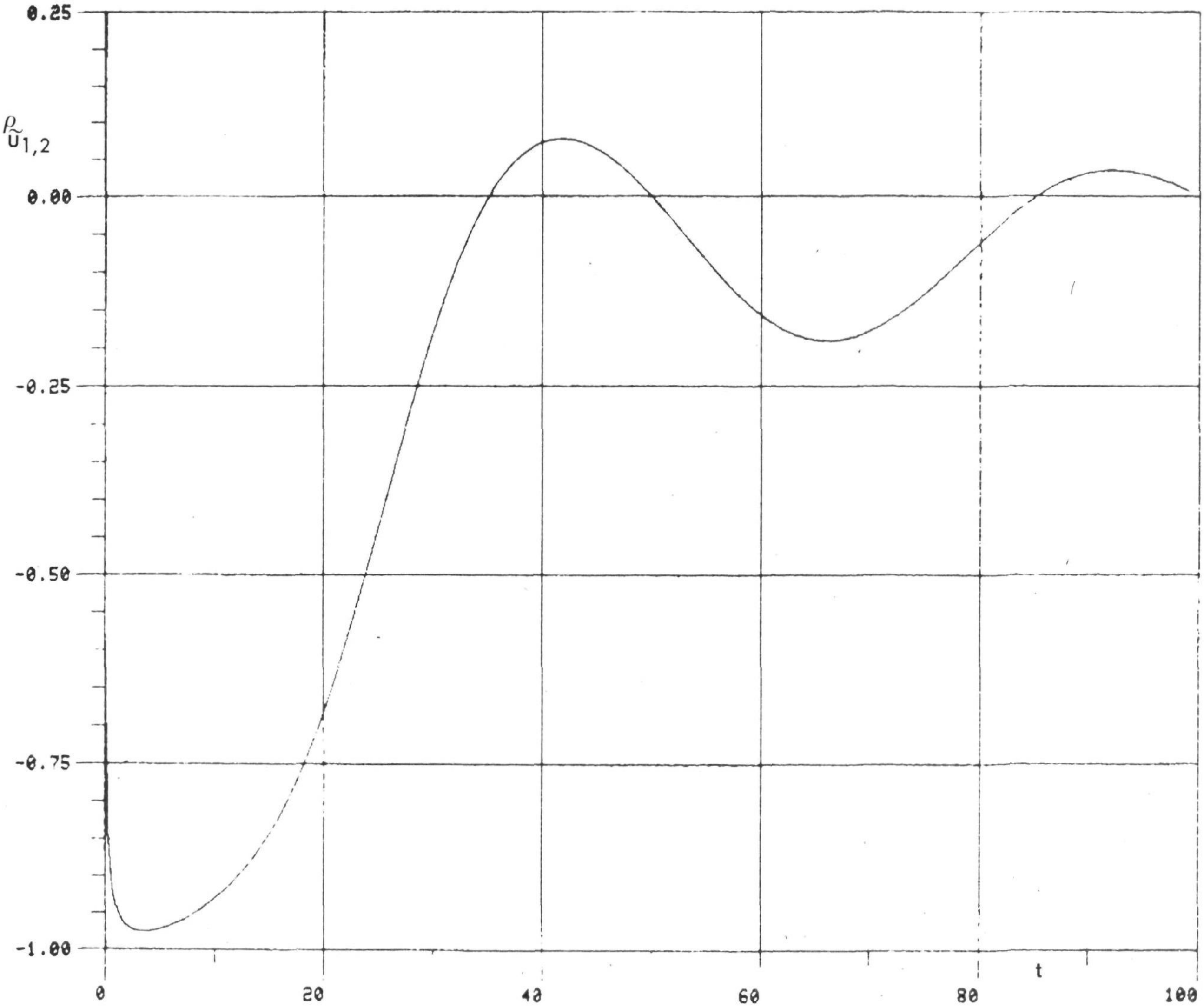

Fig. 6 Estimated correlation coefficient between wide-sense martingale u components (model problem 1; no system noise)

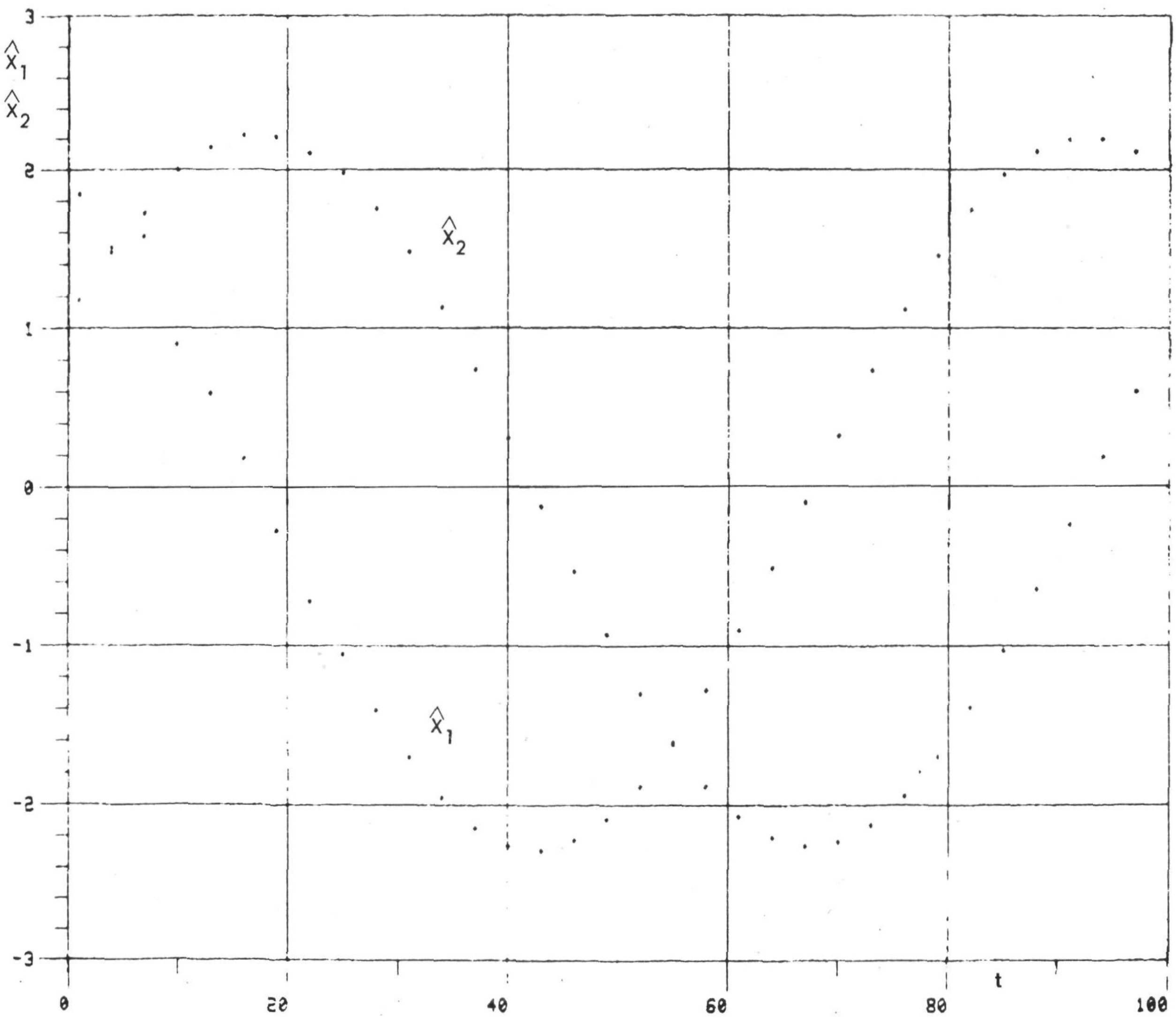

Fig. 7 Estimated components of state vector x using wide-sense martingale approach (model problem 1; no system noise)

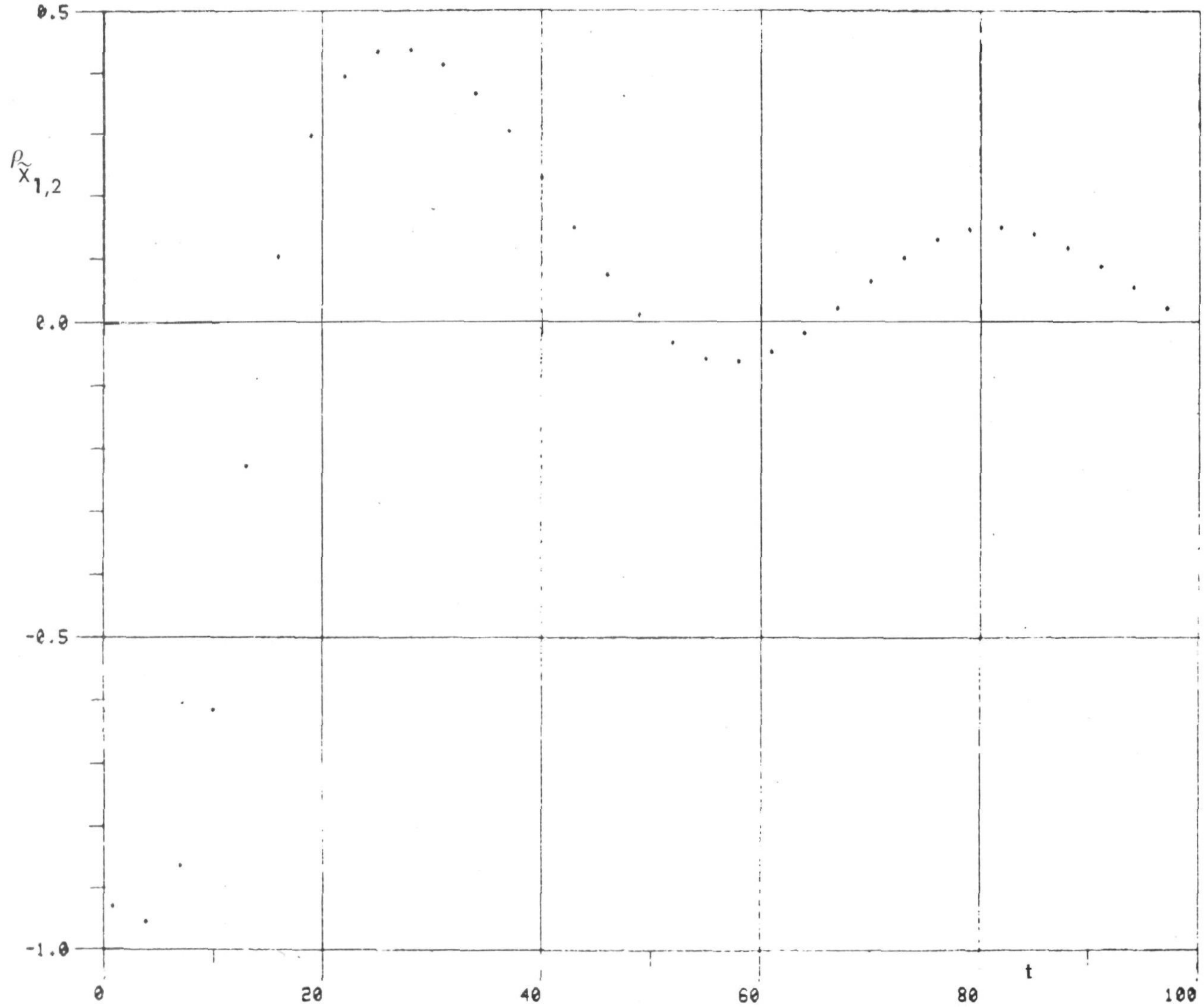

Fig. 8 Estimated correlation coefficient between state-vector elements x_1 and x_2 using wide-sense martingale approach (model problem 1; no system noise)

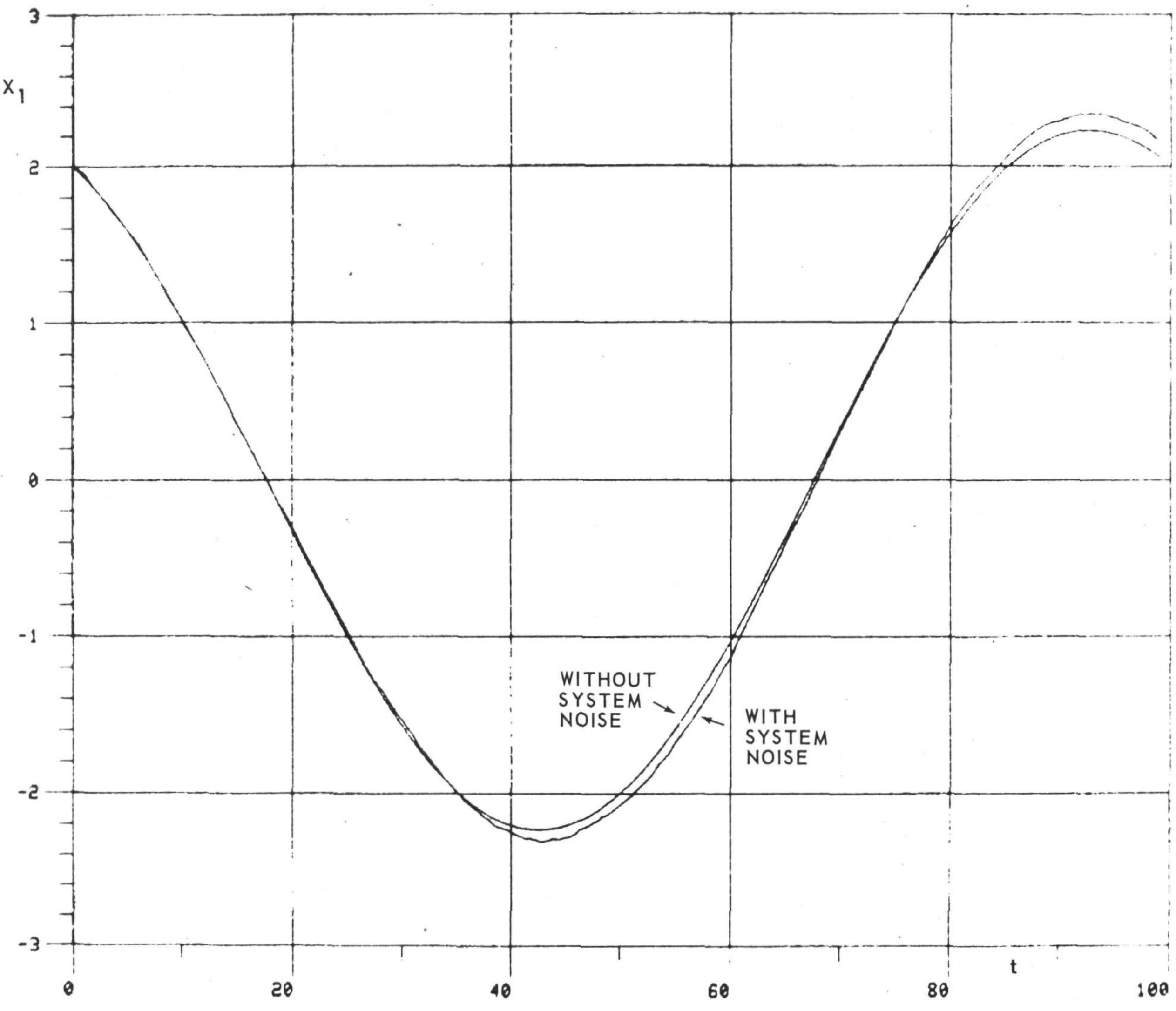

Fig. 9 Simulated state vector element x_1 with and without system noise (model problem 1)

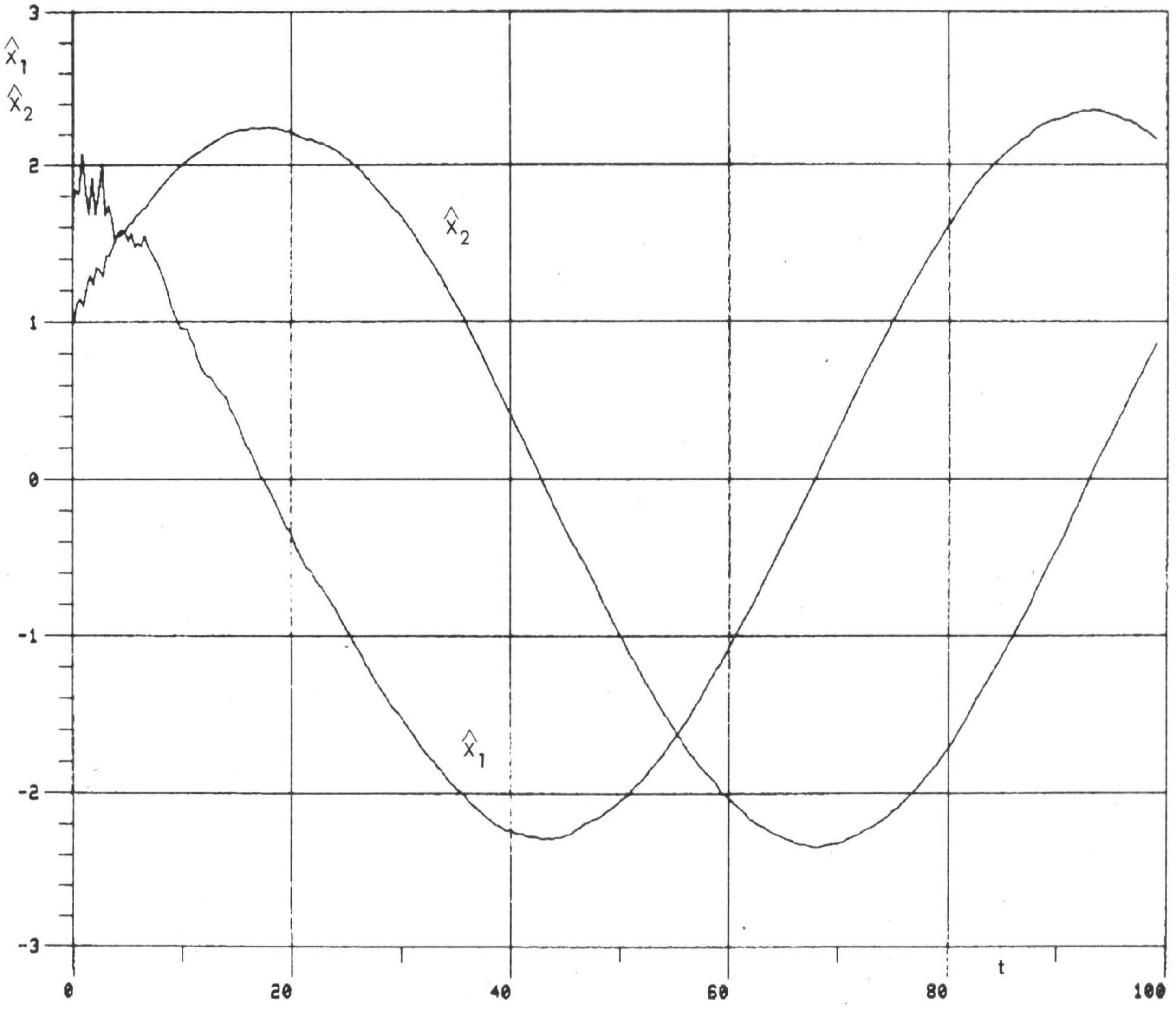

Fig. 10 Estimated components of state vector x using Kalman filter (model problem 1; system noise)

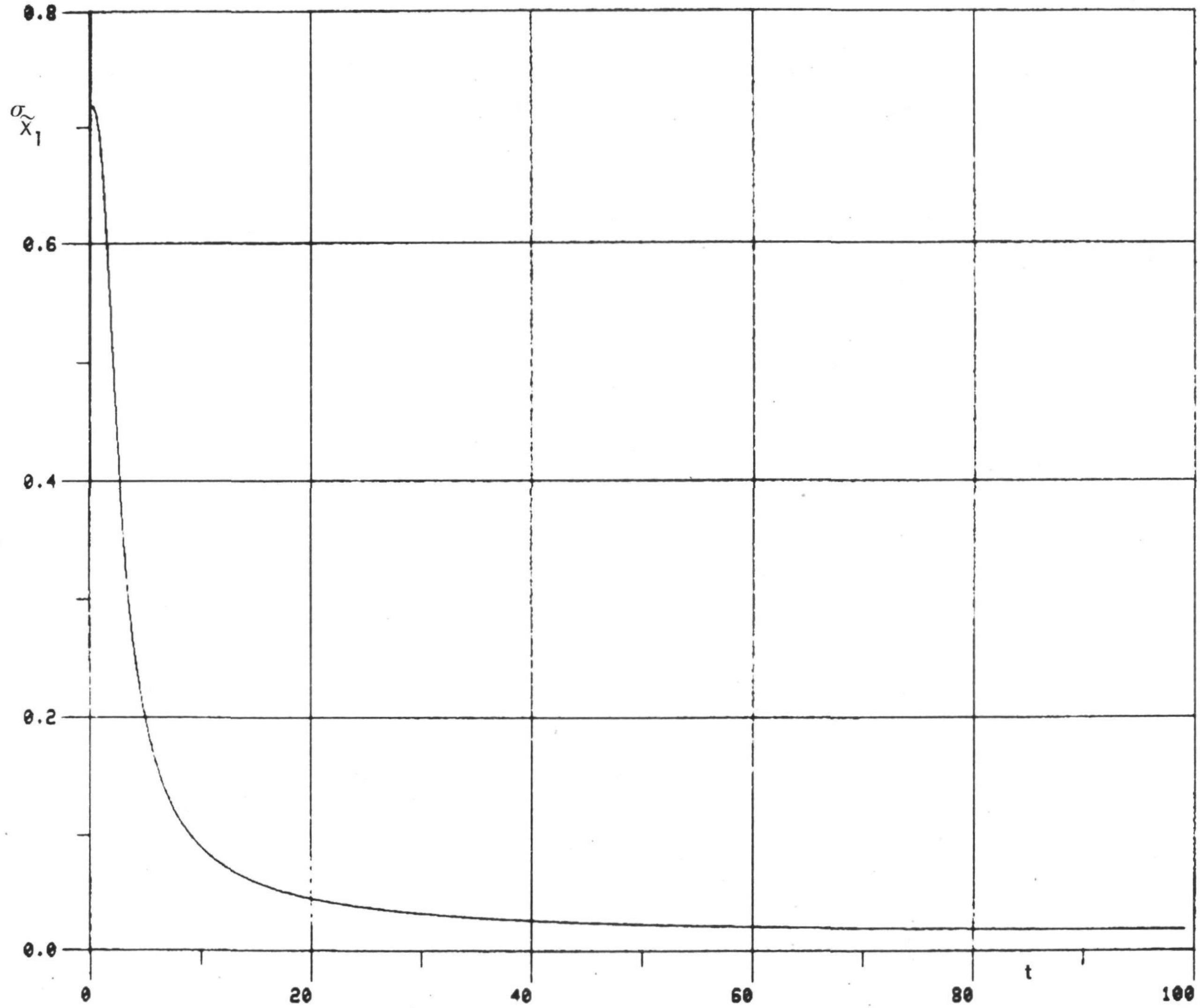

Fig. 11 Estimated standard deviation of state vector element x_1 using Kalman filter (model problem 1; system noise)

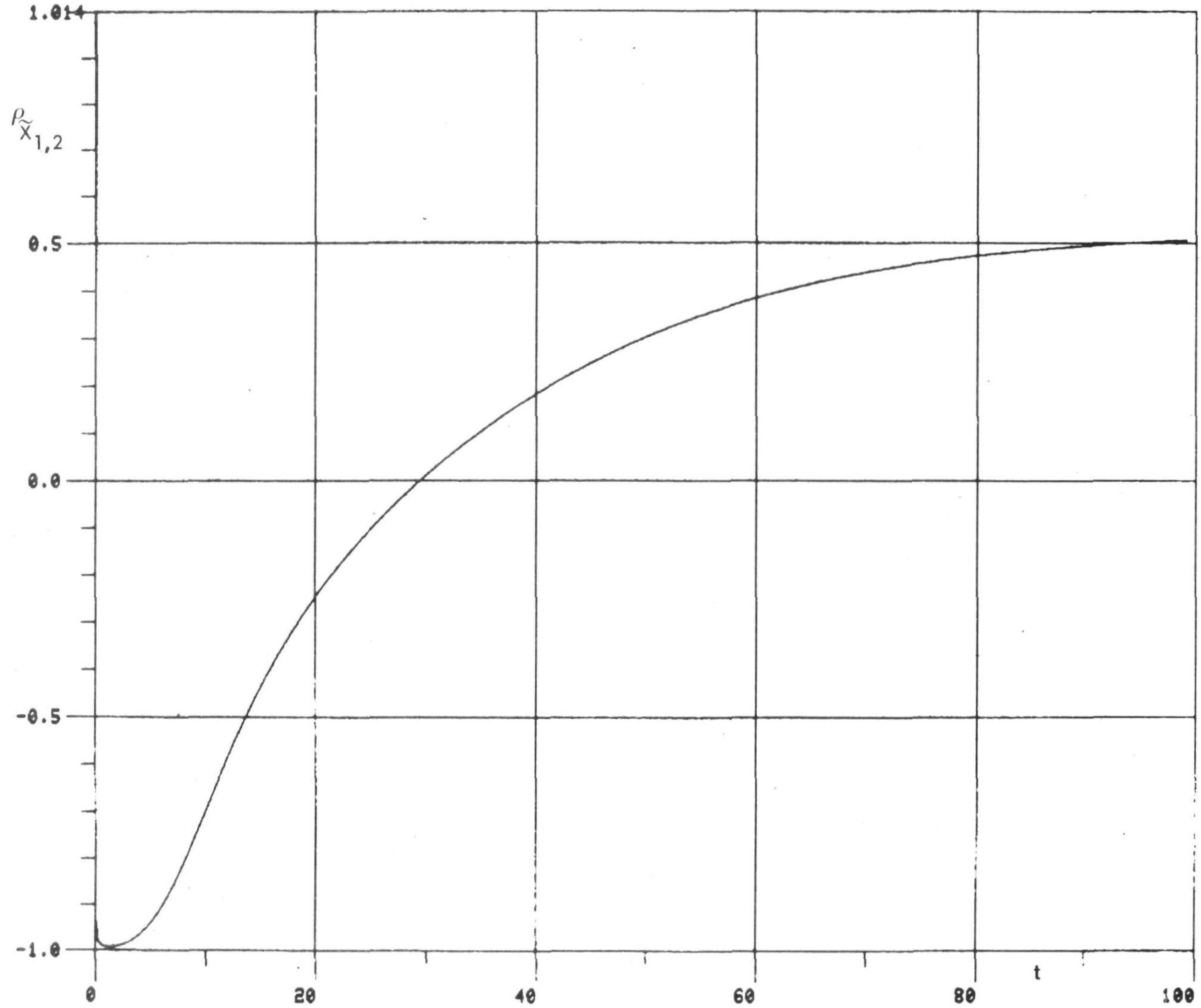

Fig. 12 Estimated correlation coefficient between state vector elements x_1 and x_2 using Kalman filter (model problem 1; system noise)

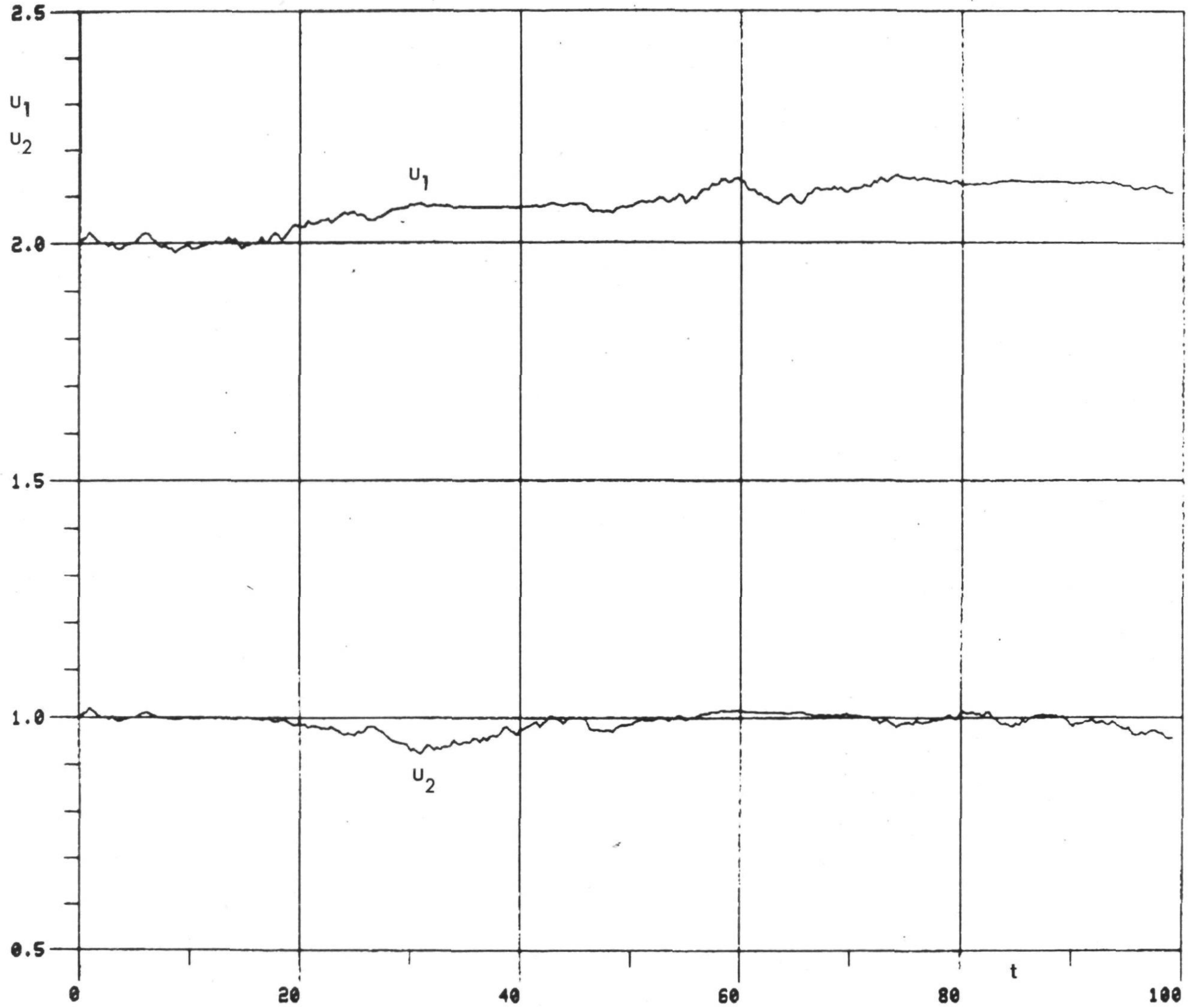

Fig. 13 Exact components of wide-sense martingale u
(model problem 1 ; system noise)

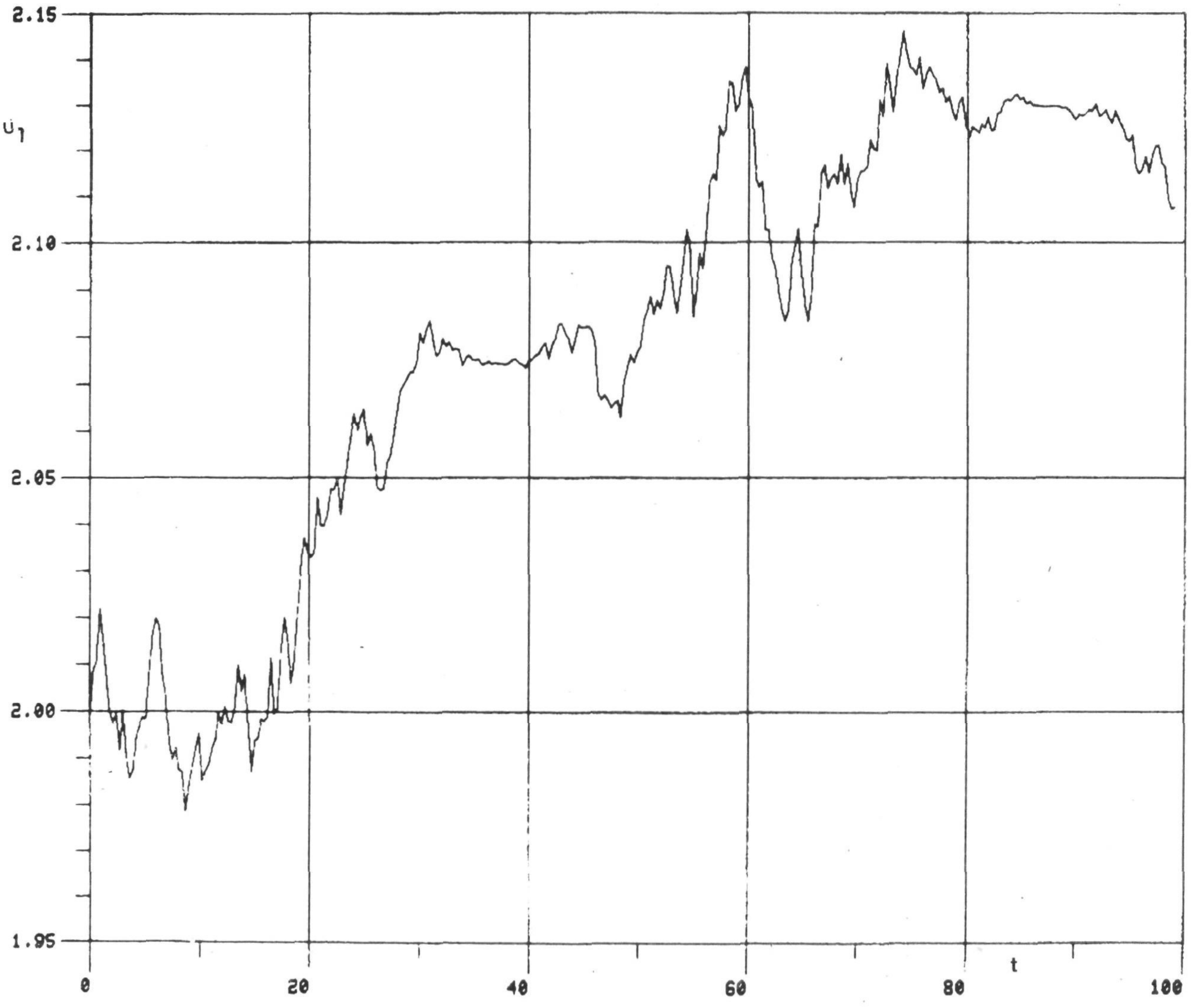

Fig. 14 First element of wide-sense martingale u (enlargement)
(model problem 1; system noise)

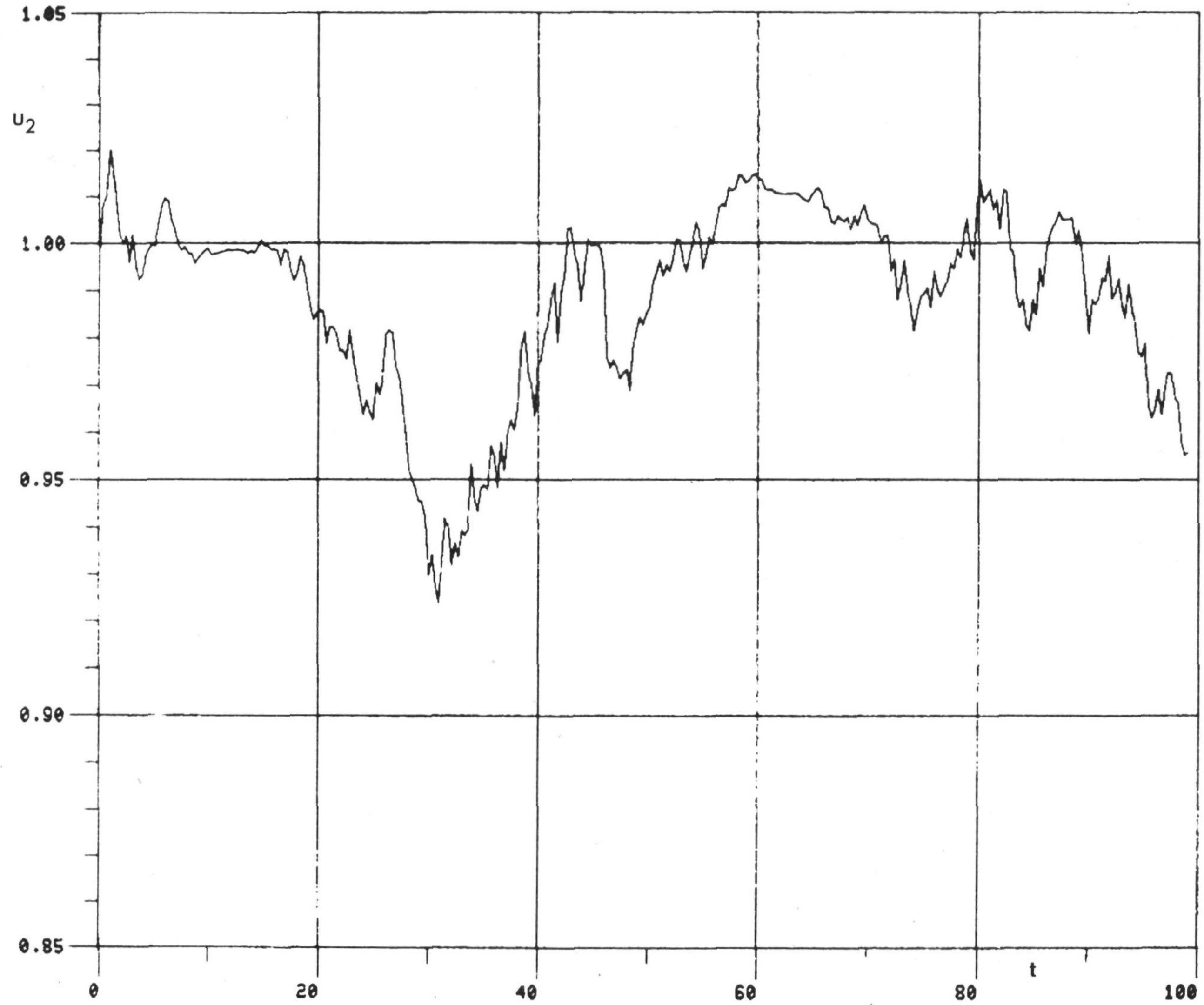

Fig. 15 Second element of wide-sense martingale u (enlargement) (model problem 1; system noise)

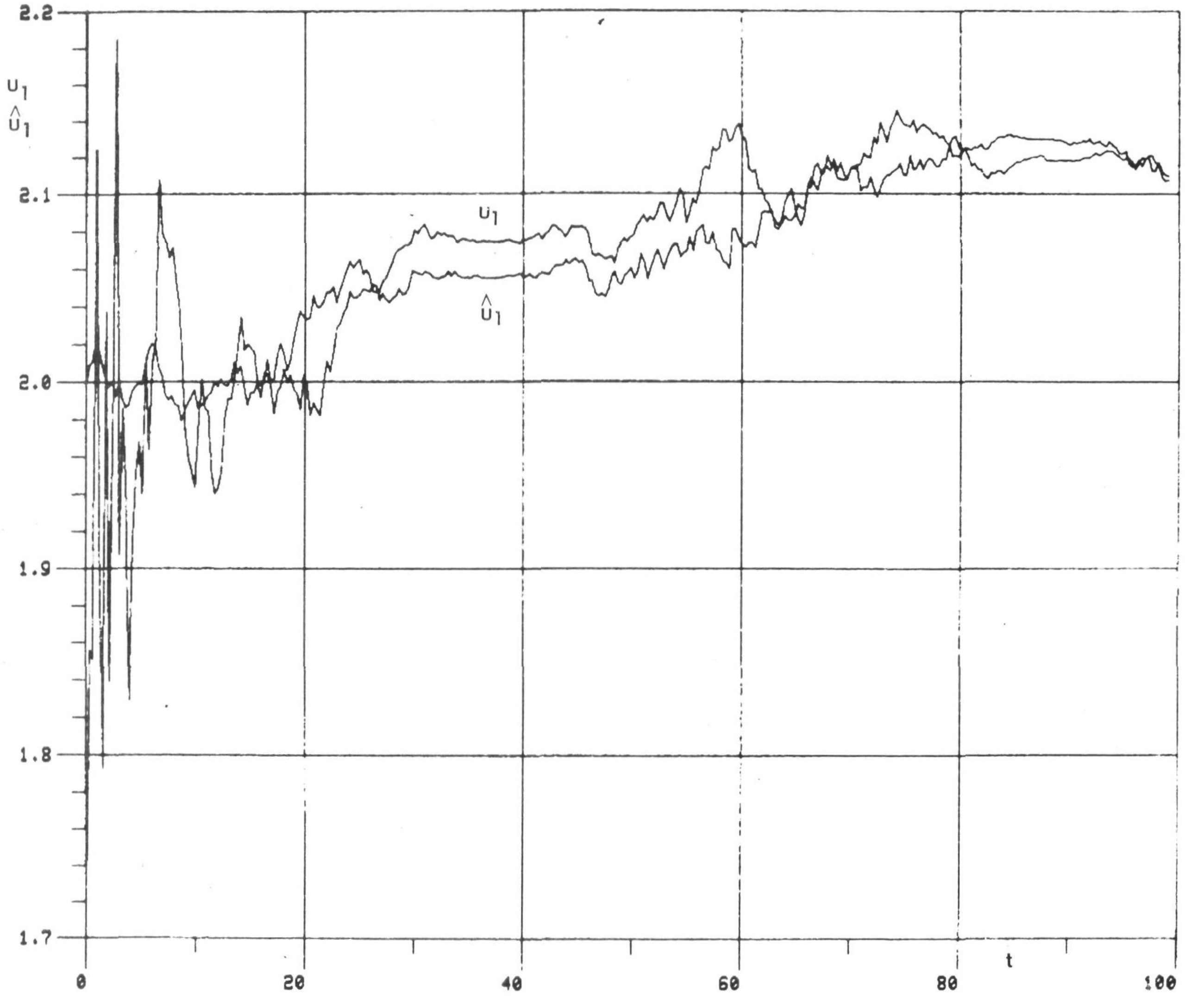

Fig. 16 Exact and estimated first component of wide-sense martingale u (model problem 1; system noise)

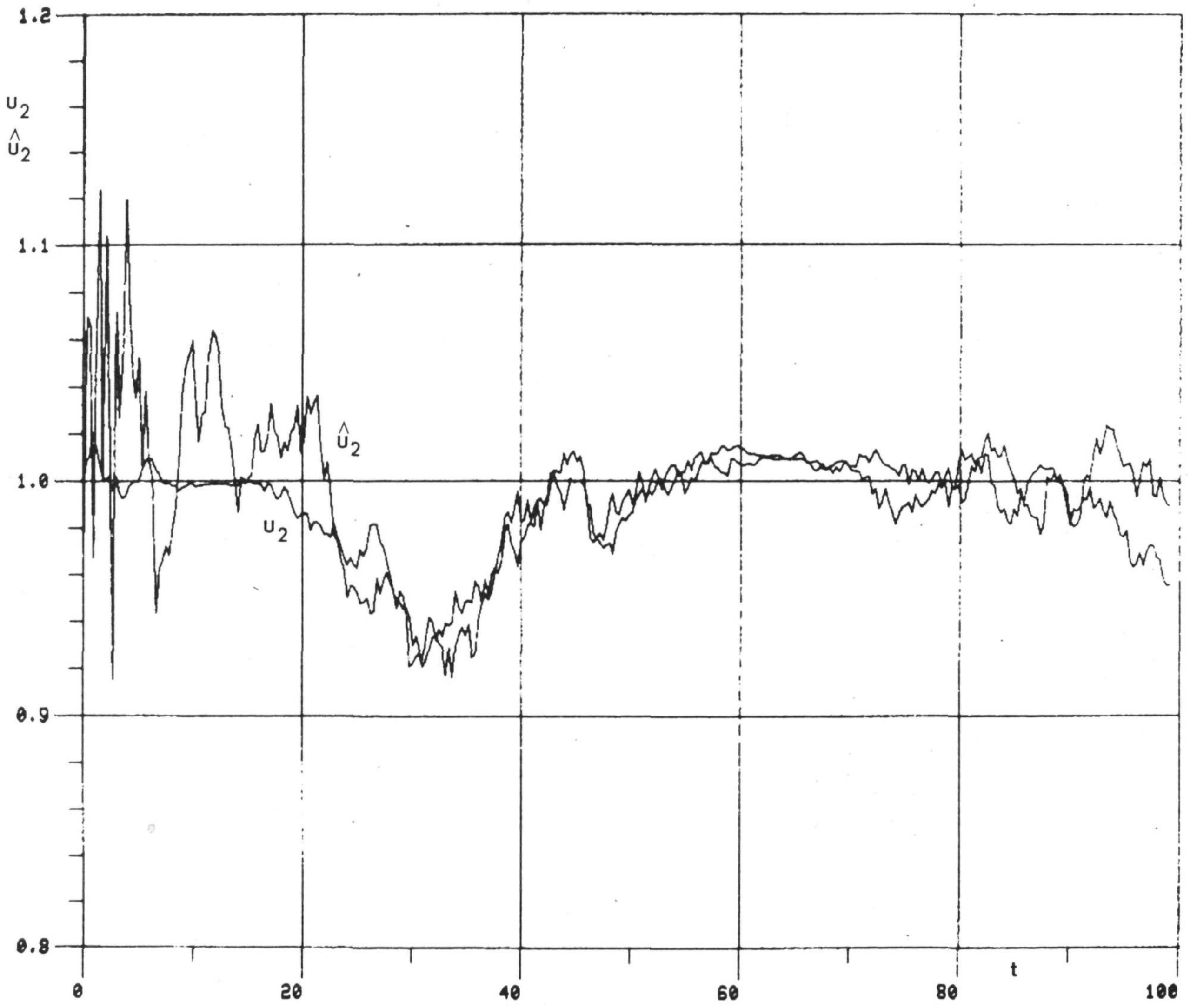

Fig. 17 Exact and estimated second component of wide-sense martingale u (model problem 1; system noise)

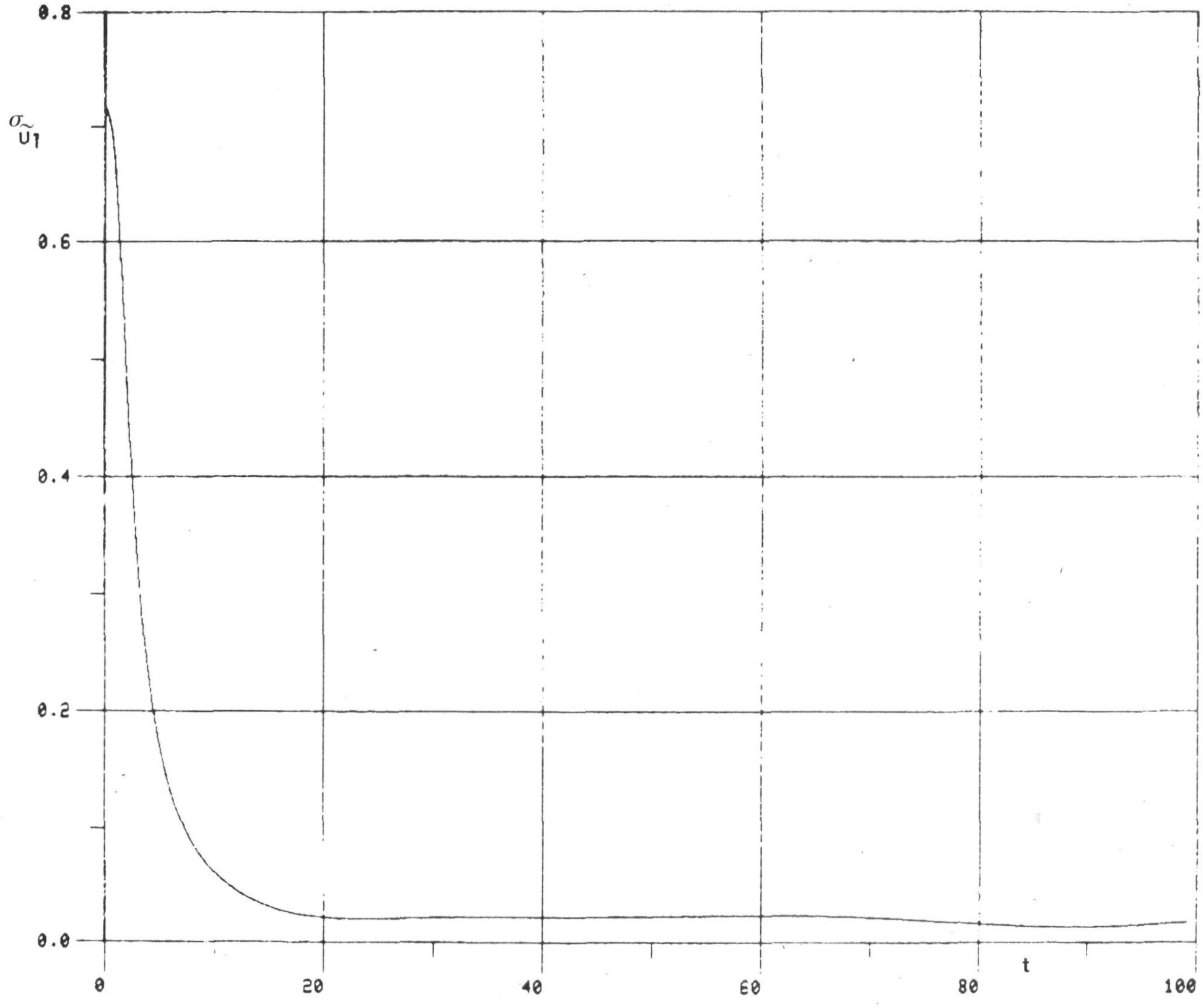

Fig. 18 Estimated standard deviation of first component of wide-sense martingale u (model problem 1; system noise)

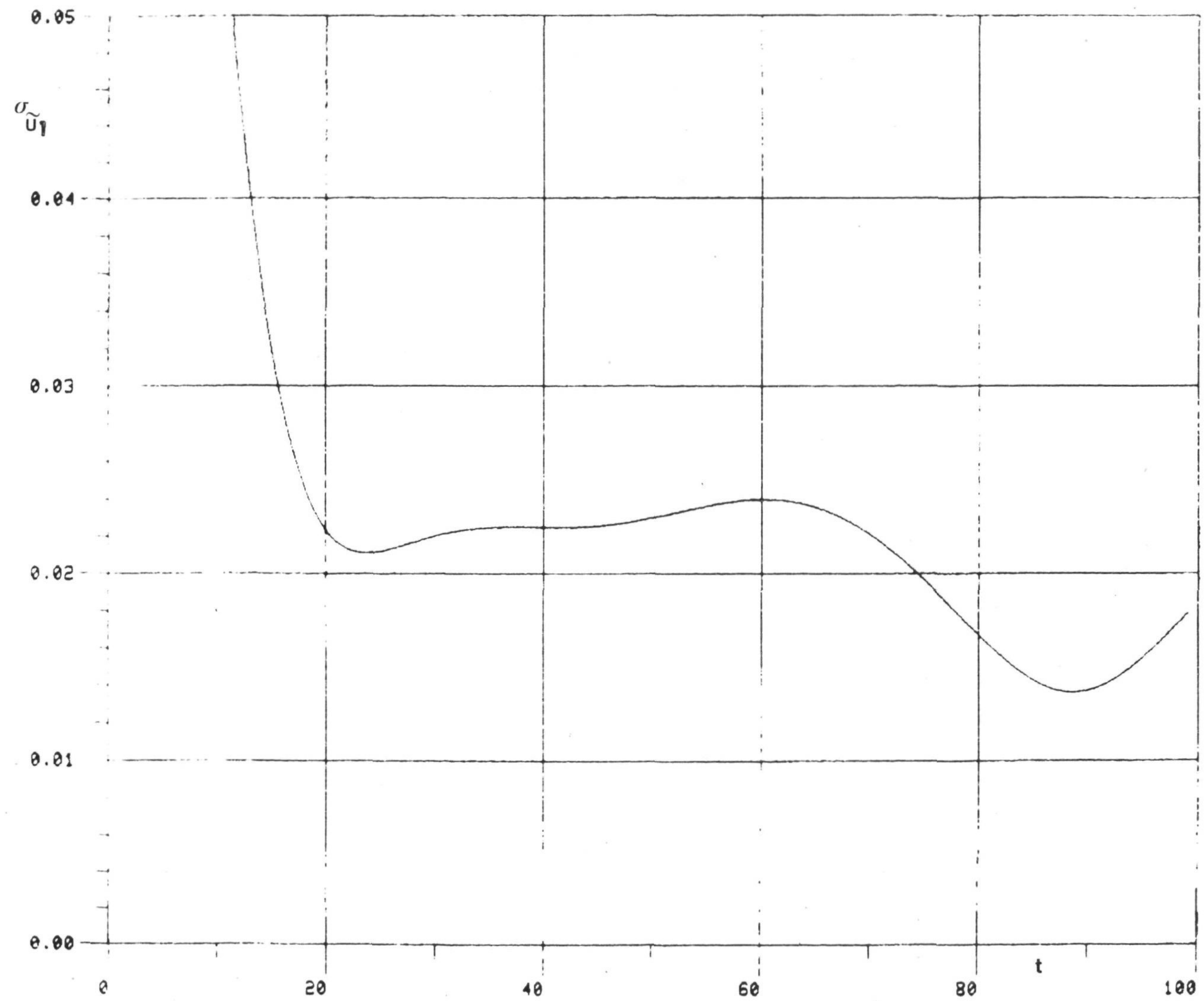

Fig. 19 Estimated standard deviation of first component of wide-sense martingale u (enlargement) (model problem 1; system noise)

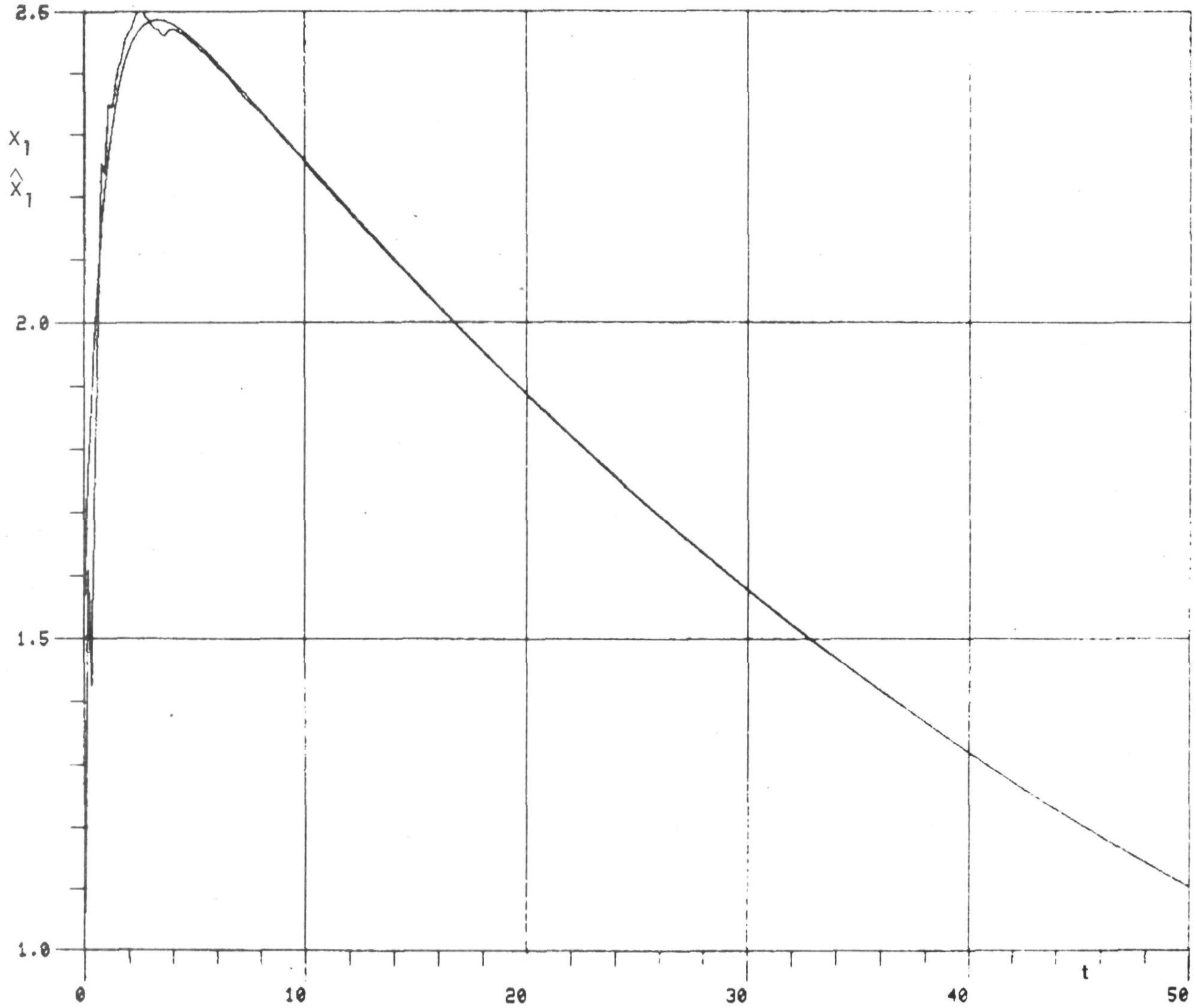

Fig. 20 Exact and estimated first component of state vector x using Kalman filter (model problem 2)

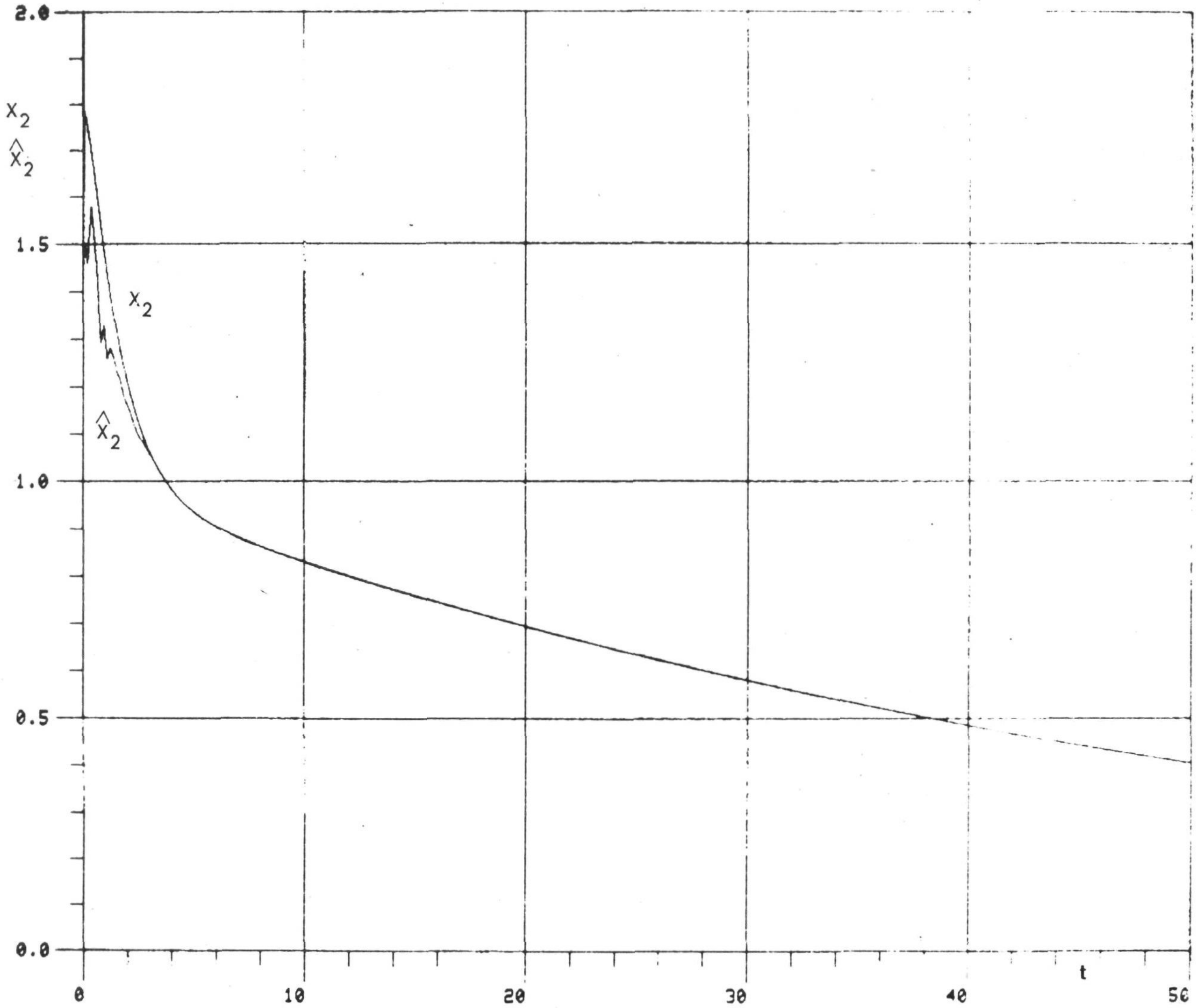

Fig. 21 Exact and estimated second component of state vector x using Kalman filter (model problem 2)

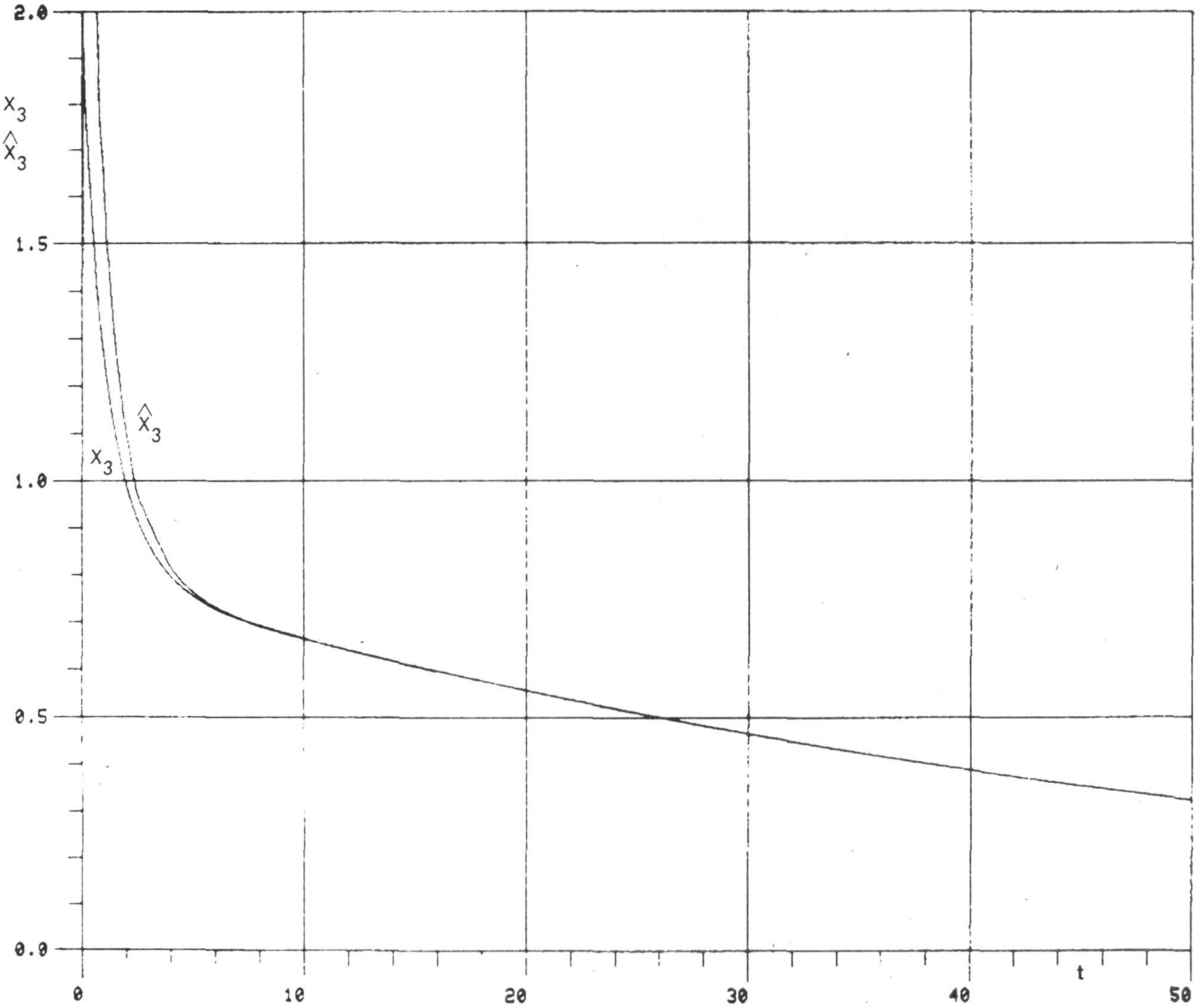

Fig. 22 Exact and estimated third component of state vector x using Kalman filter (model problem 2)

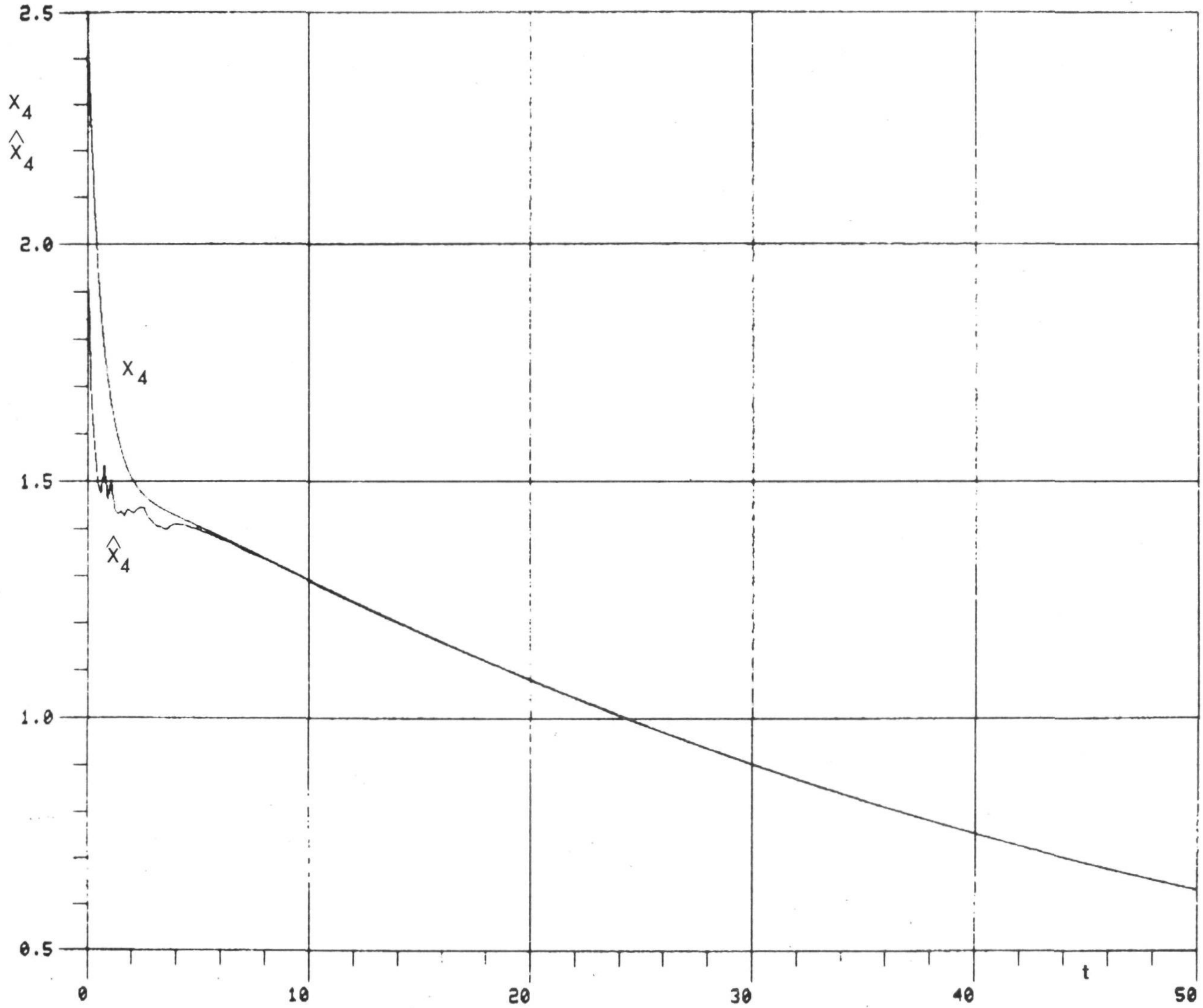

Fig. 23 Exact and estimated fourth component of state vector x using Kalman filter (model problem 2)

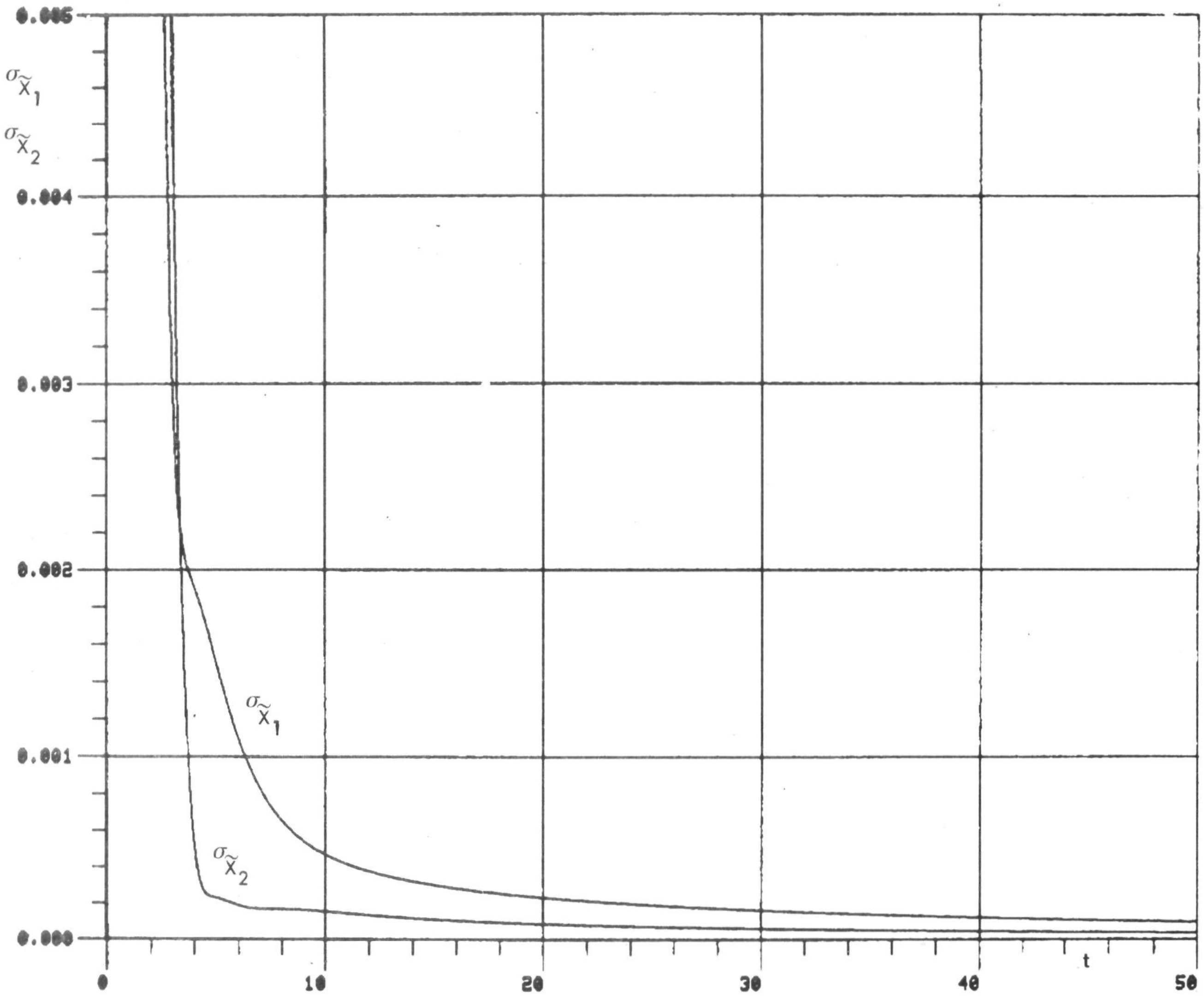

Fig. 24 Estimated standard deviation of first and second component of state vector x using Kalman filter (model problem 2)

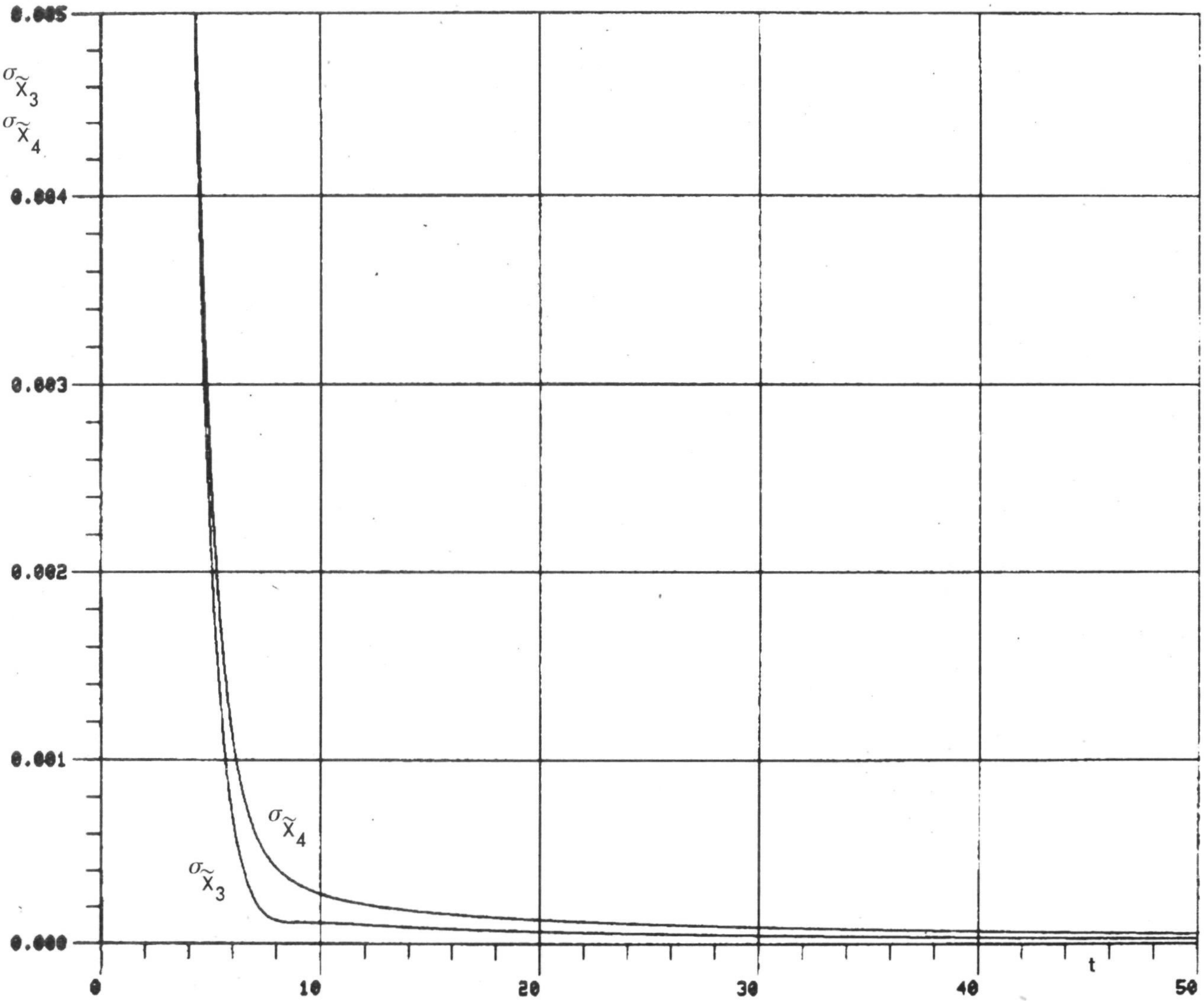

Fig. 25 Estimated standard deviation of third and fourth component of state vector x using Kalman filter (model problem 2)

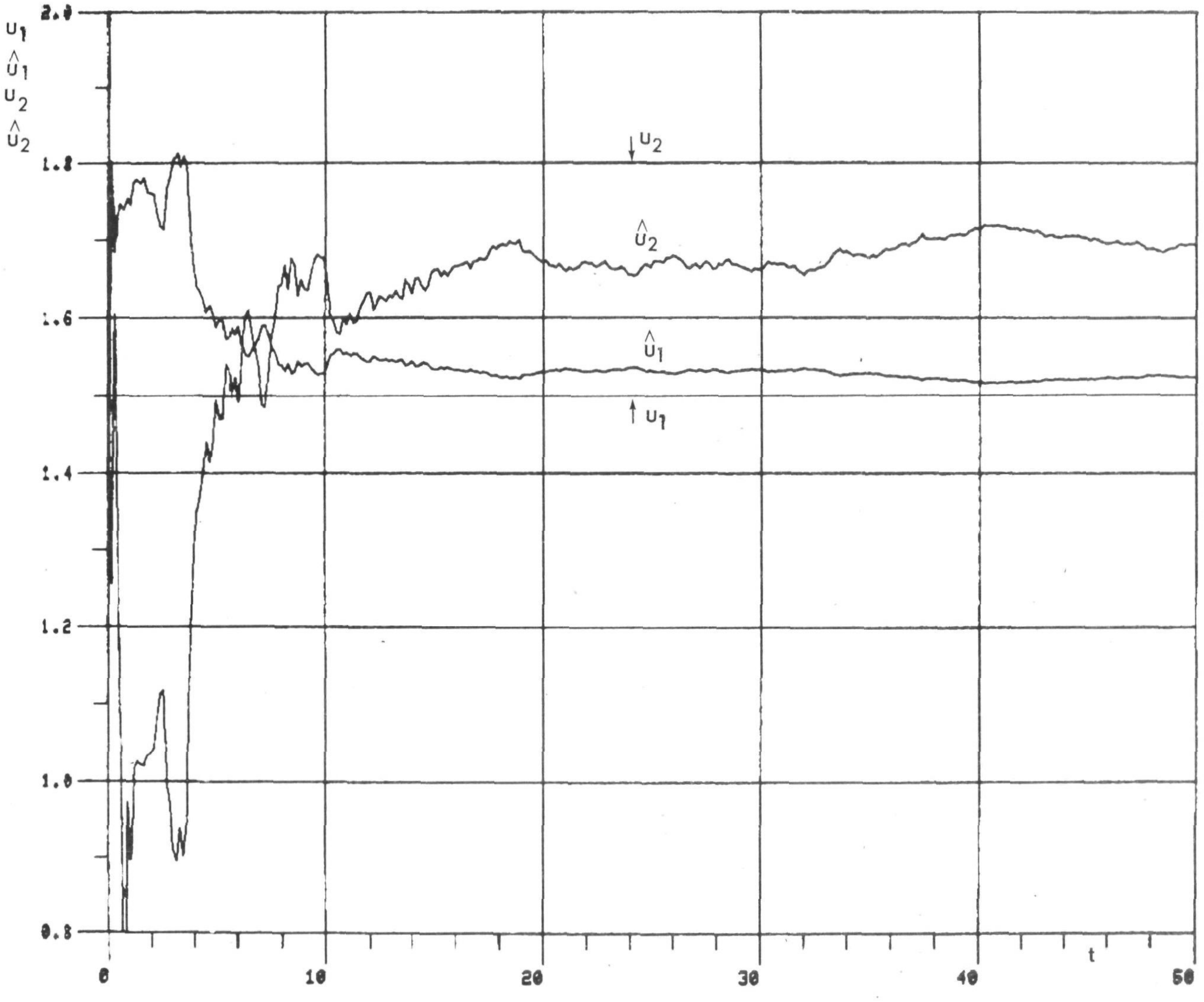

Fig. 26 Exact and estimated first and second component of wide-sense martingale u (model problem 2)

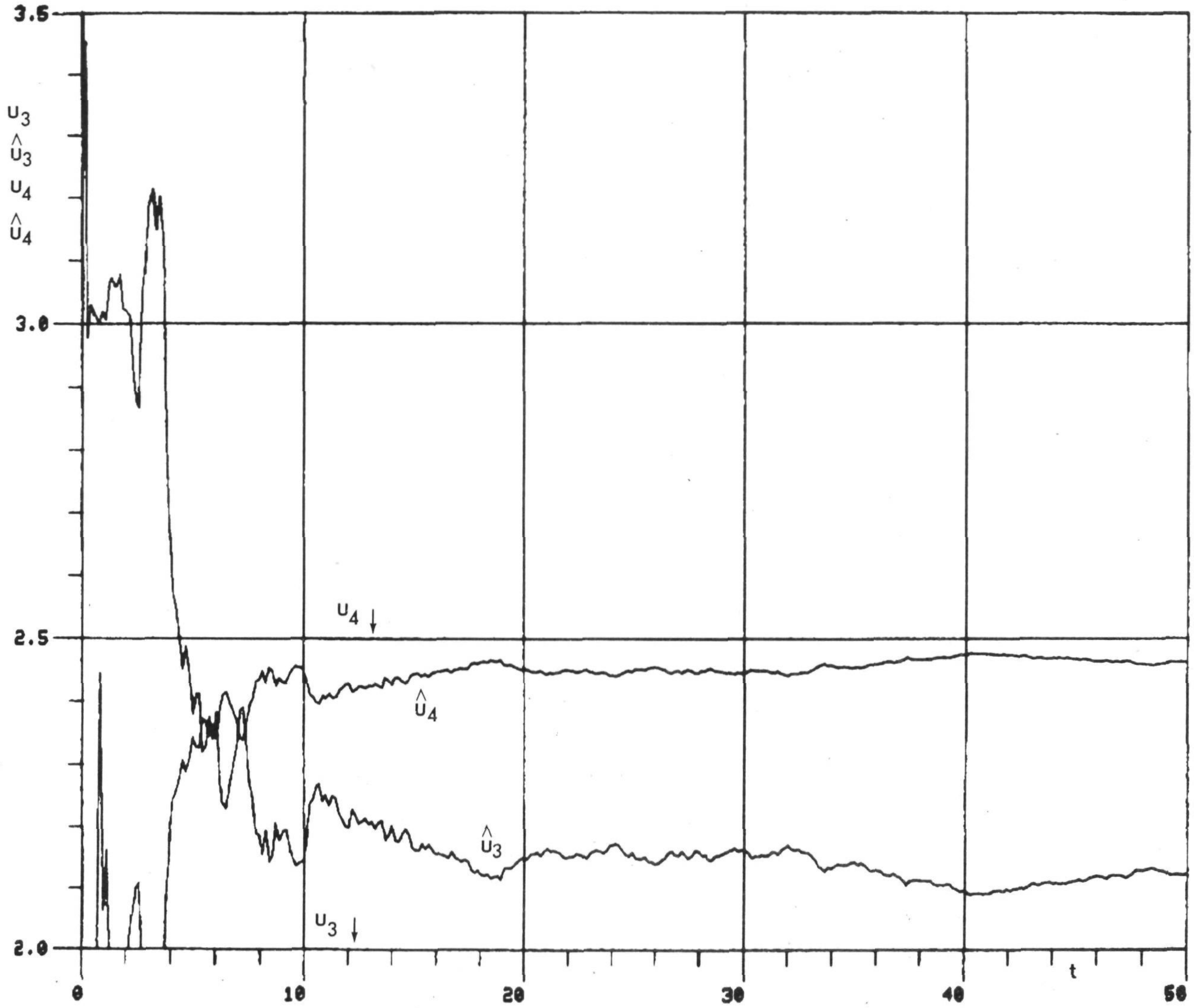

Fig. 27 Exact and estimated third and fourth component of wide-sense martingale u (model problem 2)

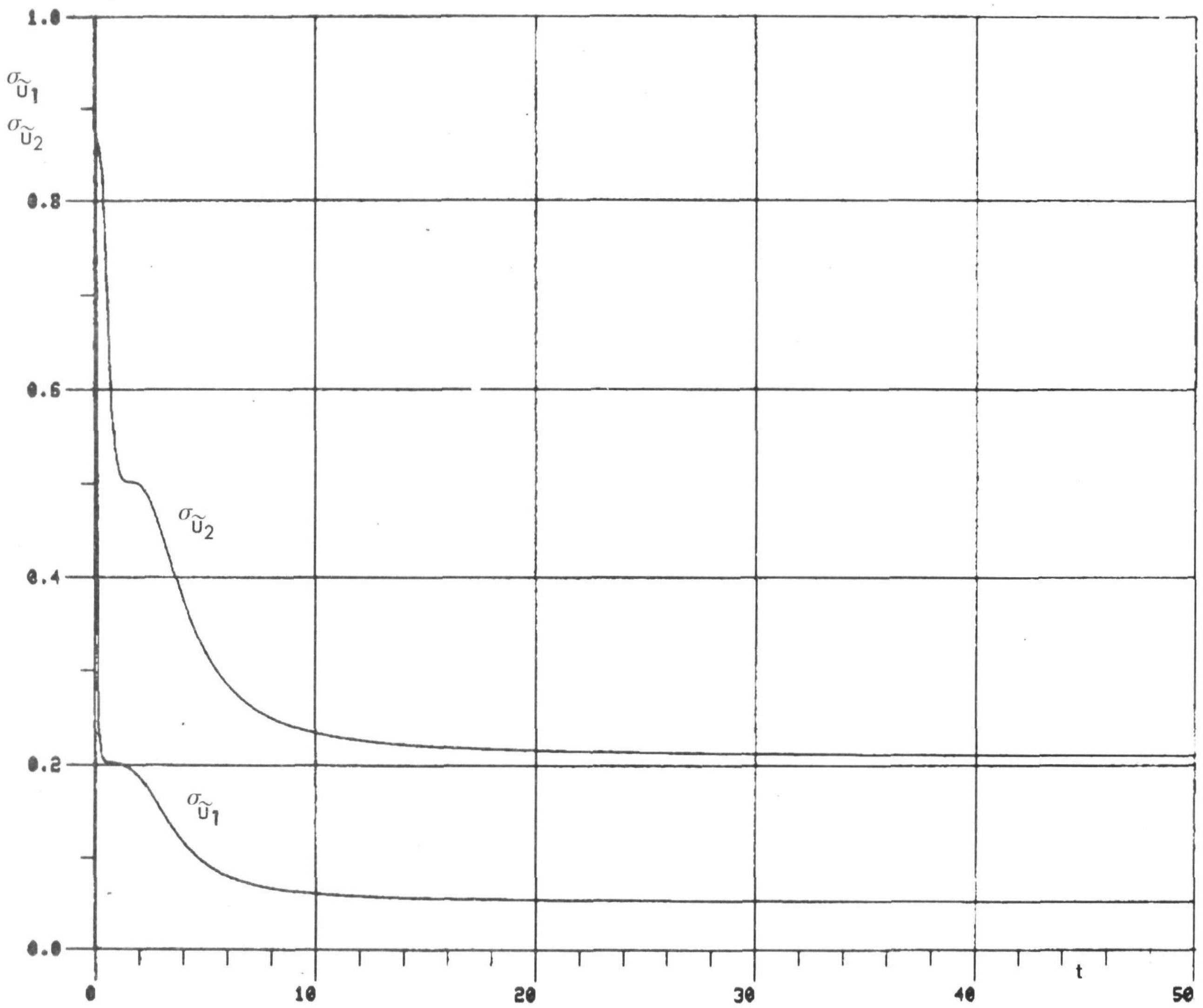

Fig. 28 Estimated standard deviation of first and second component of wide-sense martingale u (model problem 2)

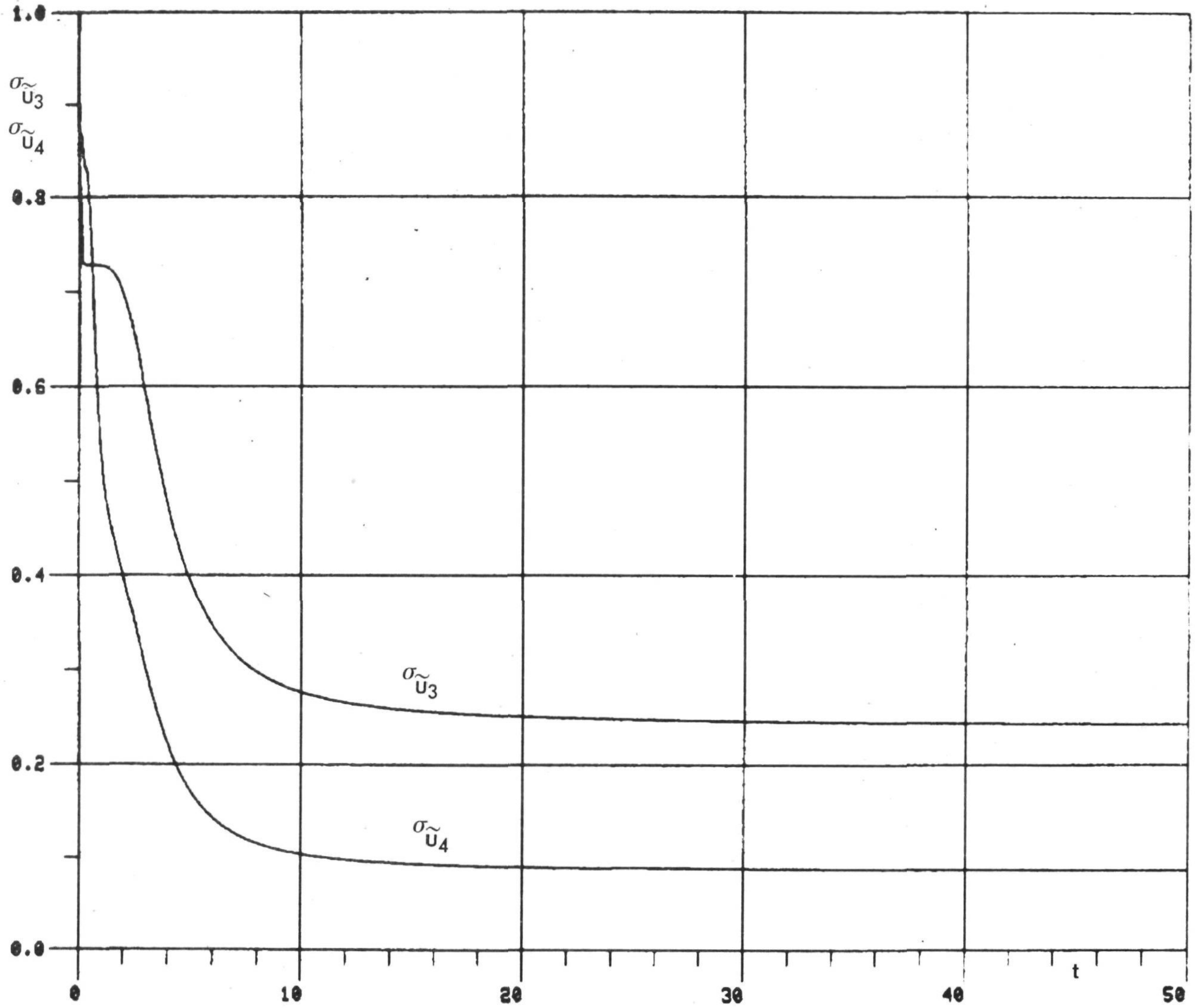

Fig. 29 Estimated standard deviation of third and fourth component of wide-sense martingale u (model problem 2)

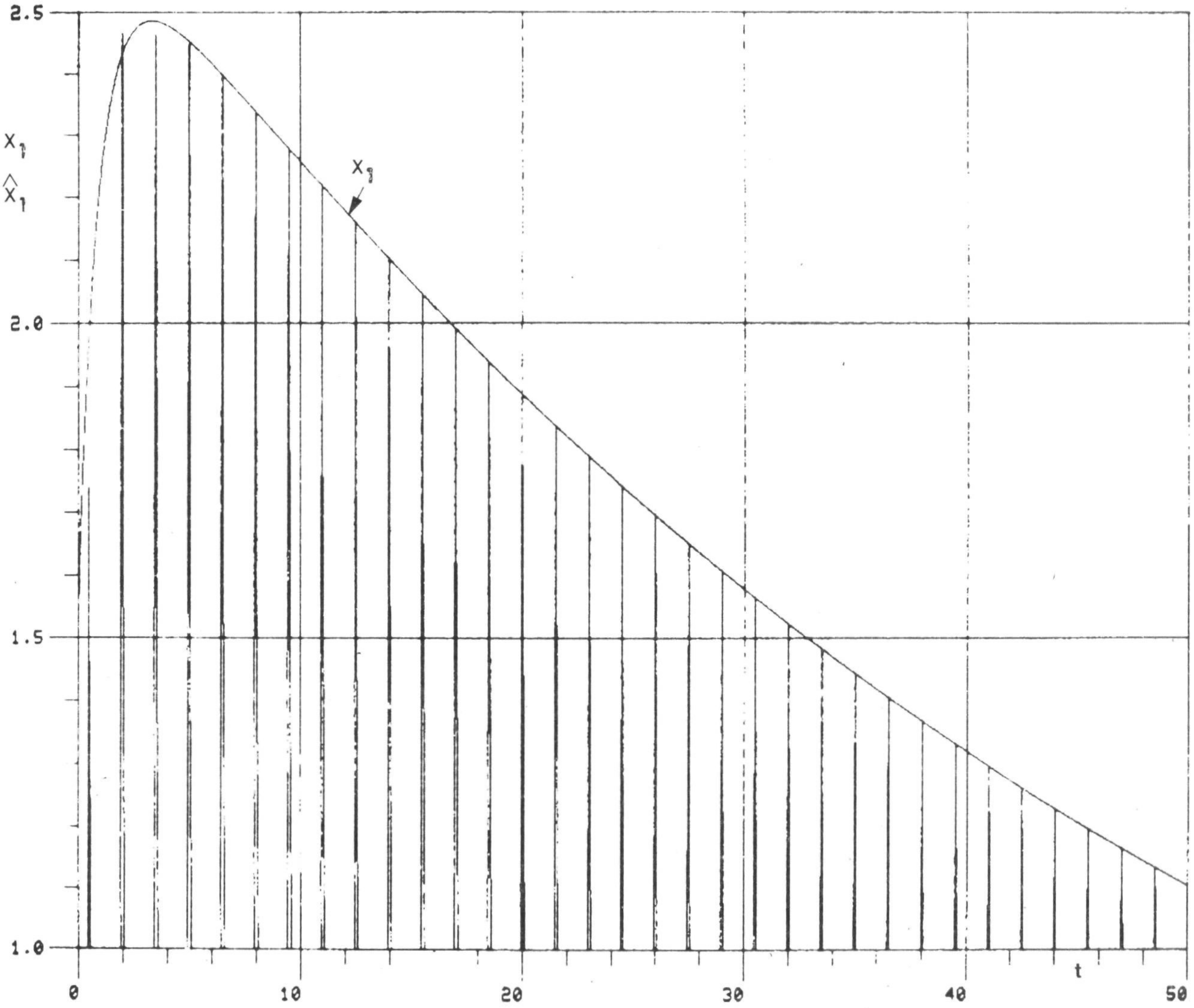

Fig. 30 Exact and estimated first component of state vector x using wide-sense martingale approach (model problem 2)

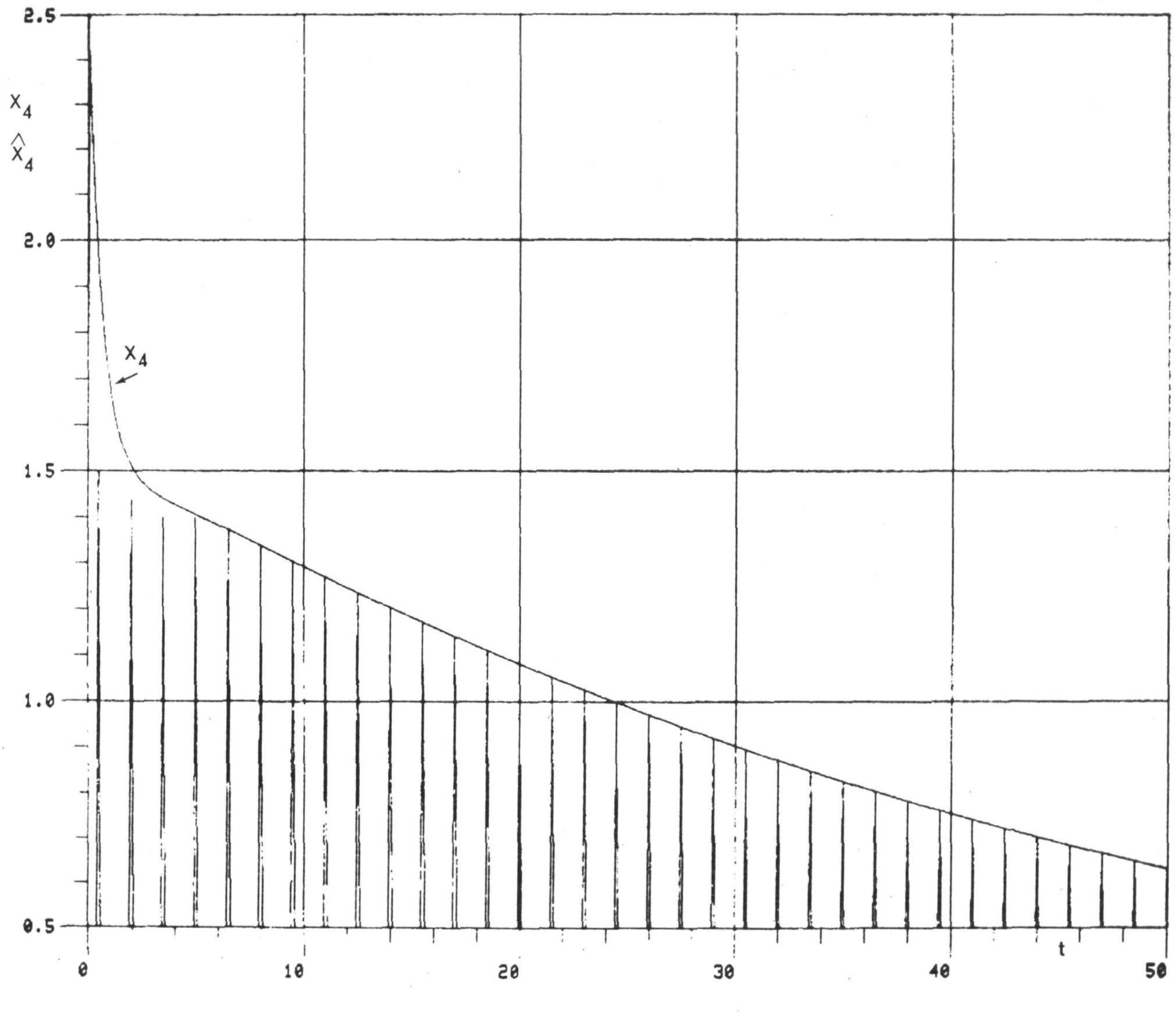

Fig. 31 Exact and estimated fourth component of state vector x using wide-sense martingale approach (model problem 2)