# Robustness of Estimators in a Finitely Additive White Noise Model

by

Hans-Peter Hucke

Purdue University
Technical Report #86-2

Department of Statistics
Purdue University

February 1986

# Robustness of Estimators in a Finitely Additive White Noise Model

by

Hans-Peter Hucke
Purdue University

ABSTRACT

A finitely additive model is used for the non-linear estimation of a random signal in the presence of 'white' gaussian noise. The continuous dependence of the resulting estimators on the observed sample path and a priori parameters is investigated. It is shown that the estimators are robust in the sense that they are Lipschitz-continuous functions. Applications to the nonlinear filtering problem and discrete approximations are also given.

## 1. Introduction.

In this paper we consider the following estimation problem for a continuous time Markov process.

Let $\mathbf{T} := [0, T]$ be a finite time interval and $(X_t)_{t \epsilon \mathbf{T}}$ a progressively measurable Markov process on a countably additive probability space $(\Omega, \mathcal{A}, \Pi)$, taking values in a complete, separable metric space $S$.

We suppose that direct observation of $(X_t)$ is not possible, but that observations are available of the process $(y_t)_{t \epsilon \mathbf{T}}$, which is related to $(X_t)$ via

$$y_t = h_t(X_t) + e_t \tag{1.1}$$

where

$$h: \ \mathbf{T} \times S \to \mathbb{R}^m$$

is a known nonlinear function of its arguments with values in the Euclidean space $\mathbb{R}^m$ and $(e_t)$ is assumed to be an $\mathbb{R}^m$- valued 'gaussian white noise' process, independent of $(X_t)$.

If $t \epsilon \mathbf{T}$ is fixed and observations $Q_t y := \{y_u : \ 0 \leq u \leq t\}$ are given, then we consider the problem of finding estimators for functionals of the (signal-) process $(X_t)$ given the observations $Q_t y$ and the a priori information contained in the probability measure $\Pi$.

If we take the random variable

$$g: \ (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B})$$

equal to $f(X_s)$, for a fixed $s \epsilon \mathbf{T}$, and $f: (S, \mathcal{S}) \to (\mathbb{R}, \mathcal{B})$, i. e.

$$g(\omega) = f(X_s(\omega)) \qquad , \omega \epsilon \Omega$$

we encounter the nonlinear filtering, prediction or smoothing problems, according to whether $s = t, s > t$ or $s < t$. Here and in the sequel $\mathcal{B}^k$ and $\mathcal{S}$ will denote the Borel $\sigma$-fields in $\mathbb{R}^k$ and $S$ respectively; for $k = 1$ we omit the superscripts.

To make the model (1.1) mathematically rigorous we have to define what is meant by a gaussian white noise processes.

The usual approach is to treat $(e_t)$ formally as the time derivative of an $m$-dimensional standard Wiener process, i. e.

$$(e_t) = (\frac{dW_t}{dt})$$

and to consider the integrated model

$$(1.2) \qquad Y_t := \int_0^t y_u \, du = \int_0^t h_u(X_u) \, du + W_t \qquad t \epsilon \mathbf{T}.$$

Equation (1.2) gives now a well-defined model on a countably additive probability space. We will call (1.2) from now on the stochastic calculus model, since stochastic (Itô) calculus is generally used to solve the estimation problem in this context. Detailed accounts of the stochastic calculus model can be found in the books of Kallianpur [5] or Liptser and Shiryayev [13].

Recently, Kallianpur and Karandikar in a series of papers [6]–[9] have proposed a different approach for the study of the estimation problem, which they called white noise calculus.
Assuming that

$$(1.3) \qquad E \int_0^T | h_u(X_u) |^2 \, du < \infty$$

they write equation (1.1) as

$$(1.4) \qquad y = \xi + e$$

where $y, \xi$ and $e$ are now considered as elements of the Hilbert space $L^2 := \{f : [0, T] \to \mathbb{R}^m; f$ is measurable and $\int_0^T | f(u) |^2 \, du < \infty\}$, satisfying the following condition. $\xi : \Omega \to L^2$ is a random element in $L^2$ such that for all $\omega \epsilon \Omega, t \epsilon \mathbf{T}$

$$(1.5) \qquad \xi_t(\omega) := h_t(X_t(\omega))$$

and $e: L^2 \to L^2$ is a random element in $L^2$ defined by

$$(1.6) \qquad e(g) := g \qquad \text{for all } g \epsilon L^2$$

having a canonical gaussian distribution, i. e. with $C_e(g)$ denoting the characteristic functional of $e$

$$(1.7) \qquad C_e(g) = exp\{-\frac{1}{2}(g,g)\} \qquad \text{for all } g \epsilon L^2.$$

Here $(.,.)$ denotes the inner product in $L^2$, and we will denote the norm by $\|.\|$. $y: \Omega \times L^2 \to L^2$ is then defined as

$$(1.8) \qquad y(\omega, g) := \xi(\omega) + e(g) \qquad \text{for all } (\omega, g) \epsilon \Omega \times L^2$$

**Remark** (1.9): Since the measure induced by $e$ on $L_2$ is only finitely additive, the process $e$ is called finitely additive white noise. By the same token the observation process $y$ is not defined on a countably additive probability space, but on a so-called quasi cylindrical probability space. For a detailed account on the white noise model (1.4) we refer to the survey paper [9] of Kallianpur and Karandikar. We only quote here the following result from [9].

**THEOREM** (1.10): Let $g: (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B})$ be an integrable random variable and for $t \epsilon \mathbf{T}$ let $Q_t$ denote the orthogonal projection of $L^2$ onto $H_t := \{f \epsilon L^2 : \int_t^T | f(u) |^2 du = 0\}$. Then the conditional expectation $E[g \mid Q_t y]$ exists and has the representation

$$(1.11) \qquad E[g \mid Q_t y] = \frac{\sigma_t(g, Q_t y, \Pi)}{\sigma_t(1, Q_t y, \Pi)} \quad (Bayes formula)$$

where

$$(1.12) \qquad \sigma_t(g, Q_t y, \Pi) = \int_\Omega g(\omega) \cdot q_t(\omega, y) \Pi(d\omega)$$

with

$$(1.13) \qquad q_t(\omega, y) = exp\{(Q_t y, Q_t \xi(\omega)) - \frac{1}{2} \| Q_t \xi(\omega) \|^2\}.$$

$E[g \mid Q_t y]$ is the optimal estimator of g, with respect to mean square error, given the observations $Q_t y$ in the model (1.4).

It should be remarked here that despite its appearance, the quantity $E[g \mid Q_t y]$ is not the conditional expectation used in countably additive probability theory, but rather a 'weak' random variable on a quasi cylindrical probability space (cf. [9]).

The term $\sigma_t(y, Q_t y, \Pi)$ defined in (1.12) is called the unnormalized conditional expectation, and is finite for every $y \epsilon L^2$. Since $\sigma_t(y, Q_t y, \Pi)$ is proportional to $E[g \mid Q_t y]$ there will be no loss of generality if we state various results only for the unnormalized estimator.

The general estimation problem in the white noise model (1.4) being solved by the computation of the conditional expectation (1.11), we now turn to the investigation of its continuity properties.

## 2. Continuity.

As the terminology indicates $E[g \mid Q_t y]$ depends on the observed sample path $y \epsilon L^2$ and on the probability measure $\Pi$. We will now show that $E[g \mid Q_t y]$ depends continuously on $y$ and $\Pi$. The continuous dependence of the estimator on the observations is usually referred to as robustness of the estimator (cf. Clark [2], Picard [14].

The following theorem is the fundamental result of this section.

**THEOREM** (2.1): Let $g$ be a bounded random variable on $(\Omega, \mathcal{A}, \Pi)$ then $E[g \mid Q_t y]$ is a locally Lipschitz- continuous function of $y$, uniformly in $t \epsilon \mathbf{T}$, i. e.

(2.2)
$$\sup_{t \epsilon \mathbf{T}} \mid E[g \mid Q_t y] - E[g \mid Q_t z] \mid \leq K \cdot \parallel y - z \parallel$$

for all $y, z$ is a bounded subset $C$ of $L^2$ and some constant $K > 0$.

We will prove (2.1) by first establishing the corresponding result for the unnormalized expectation.

**THEOREM** (2.3): Under the conditions of Theorem (2.1) we have

(2.4)
$$\sup_{t \epsilon \mathbf{T}} \mid \sigma_t(g, Q_t y, \Pi) - \sigma_t(g, Q_t z, \Pi) \mid \leq K \cdot \parallel y - z \parallel .$$

**Proof**: From (1.12) we obtain

(2.5)
$$\mid \sigma_t(g, Q_t y, \Pi) - \sigma_t(g, Q_t z, \Pi) \mid \leq \int \mid g(\omega) \mid \cdot \mid q_t(y, \omega) - q_t(z, \omega) \mid \Pi(d\omega).$$

By applying the inequality

(2.6)
$$\mid e^a - e^b \mid \leq \mid a - b \mid \cdot (e^a + e^b)$$

which is valid for all $a, b \epsilon \mathbb{R}$ to (2.5), the right hand side can be bounded by

$$\int \mid g(\omega) \mid \cdot \mid (Q_t(y - z), Q_t \xi(\omega)) \mid \cdot [exp\{(Q_t y, Q_t \xi(\omega)) - \frac{1}{2} \parallel Q_t \xi(\omega) \parallel^2\}$$

(2.7)
$$+ exp\{(Q_t z, Q_t \xi(\omega)) - \frac{1}{2} \parallel Q_t \xi(\omega)) \parallel^2\}]\Pi(d\omega)$$

The triangle inequality now implies that

$$(2.8) \qquad (Q_t y, Q_t \xi(\omega)) - \frac{1}{2} \| Q_t \xi(\omega) \|^2 \leq \frac{1}{2} \| Q_t y \|^2 \leq \frac{1}{2} \| y \|^2$$

and from the Cauchy-Schwartz inequality we obtain

$$| (Q_t(y-z), Q_t \xi(\omega)) | \leq \| Q_t(y-z) \| \cdot \| Q_t \xi(\omega) \| \leq \| y - z \| \cdot \| \xi(\omega) \| .$$

Hence we can bound the right-hand side of (2.7) by

$$(2.9) \qquad \| y - z \| \cdot K(y;z) \int | g(\omega) | \cdot \| \xi(\omega) \| \Pi(d\omega)$$

with

$$K(y;z) = exp\{\frac{1}{2} \| y \|^2\} + exp\{\frac{1}{2} \| z \|^2\}.$$

By assumption (1.3) and the boundedness of $g$ the integral in (2.8) is finite, so that the boundedness of the set $C$ in $L^2$ implies

$$| \sigma_t(g, Q_t y, \Pi) - \sigma_t(g, Q_t z, \Pi) | \leq K \cdot \| y - z \| \qquad \text{for all } t \epsilon \mathbf{T}$$

with

$$(2.10) \qquad K := \sup_{v \epsilon C} \; exp\{\| v \|^2\} \cdot \int | g(\omega) | \cdot \| \xi(\omega) \| \Pi(d\omega) < \infty$$

which completes the proof. □

For the proof of (2.1) we observe that

$$E[g \mid Q_t y] - E[g \mid Q_t z] = \frac{1}{\sigma_t(1, Q_t y, \Pi)} \cdot \Big[ \sigma_t(y, Q_t y, \Pi) - \sigma_t(y, Q_t z, \Pi) - \{\sigma_t(1, Q_t y, \Pi) - \sigma_t(1, Q_t z, \Pi)\} \cdot E[g \mid Q_t y] \Big]$$

so that from (2.3) we have

$$| E[g \mid Q_t y] - E[g \mid Q_t z] | \leq 2 \cdot K \cdot \| y - z \| \cdot \frac{1}{\sigma_t(1, Q_t y, \Pi)}$$

(cf. Picard [14]).

Thus the proof will be complete it we can show that $\sigma_t(1, Q_t y, \Pi)$ is bounded away from zero for $y \epsilon C$. To this end we note that (1.3) implies the existence of a constant $M > 0$, such that

$$\delta := \Pi\{\| \xi(\omega) \| \leq M\} > 0.$$

Hence the following chain of inequalities is valid

$$\begin{aligned}\sigma_t(1,Q_ty,\Pi) &\geq \int exp\{-\parallel y\parallel\cdot\parallel\xi(\omega)\parallel-\frac{1}{2}\parallel\xi(\omega)\parallel^2\}\Pi(d\omega)\\ &\geq \int_{\{\parallel\xi(\omega)\parallel\leq M\}} exp\{-K_C\cdot\parallel\xi(\omega)\parallel-\frac{1}{2}\parallel\xi(\omega)\parallel^2\}\Pi(d\omega)\\ &\geq \delta\cdot exp\{-K_C\cdot M-\frac{1}{2}M^2\}>0\end{aligned}$$

with $K_C := \sup_{v\epsilon C}\parallel v\parallel$.

This establishes the strict positivity of $\sigma_t(1,Q_ty,\Pi)$ and thus completes the proof. □

Having thus proved the robustness of the estimator with respect to the observed sample path we now consider the continuity properties of $\sigma_t(g,Q_ty,\Pi)$ as a function of $\Pi$.

For this let $\mathcal{M}_b(\Omega)$ be the set of all finite signed measures on $(\Omega,\mathcal{A})$. Endowed with the total variation norm

$$\parallel\nu\parallel_{TV} := sup\sum_{n=1}^{\infty}\mid\nu(A_n)\mid \qquad \nu\epsilon\mathcal{M}_b(\Omega)$$

$\mathcal{M}_b(\Omega)$ is a Banach space (here the supremum is taken over all countable partitions $(A_n)_{n\geq 1}\subset\mathcal{A}$.

The following result now proves the strong continuity of the unnormalized estimator with respect to the apriori distribution $\Pi$.

**THEOREM** (2.11): Let $y\epsilon L^2$ be fixed and $g$ be a bounded random variable on $(\Omega,\mathcal{A})$. Then $\sigma_t(g,Q_ty,\Pi)$ is Lipschitz-continuous in $\Pi$, uniformly in $t$, i. e.

$$\sup_{t\epsilon\mathbf{T}}(\sigma_t(g,Q_ty,\Pi)-\sigma_t(g,Q_ty,\Pi')\mid\leq K\cdot\parallel\Pi-\Pi'\parallel_{TV} \tag{2.12}$$

for some constant $K$ and all probability measures $\Pi,\Pi'$ on $(\Omega,\mathcal{A})$.

**Proof:** From definition (1.12) we obtain

$$\begin{aligned}\mid\sigma_t(g,Q_ty,\Pi)-\sigma_t(g,Q_ty,\Pi')\mid=\mid\int g(\omega)\times exp\{(Q_ty,Q_t\xi(\omega))\\ -\frac{1}{2}\parallel Q_t\xi(\omega)\parallel^2\}(\Pi-\Pi')(d\omega)\mid.\end{aligned}$$

By the boundedness of $g$ and relation (2.8) we can bound the right-hand side by

$$C\cdot exp\{\frac{1}{2}\parallel y\parallel^2\}\cdot\int\mid\Pi-\Pi'\mid(d\omega)=C\cdot exp\{\frac{1}{2}\parallel y\parallel^2\}\parallel\Pi-\Pi'\parallel_{TV}$$

and since this bound is independent of $t$ (2.12) is proved. □

As a corollary we can now obtain the joint continuity of $\sigma_t(g, Q_t y, \Pi)$ in $(y, \Pi)$. For this we consider the product space $L^2 \times \mathcal{M}_b(\Omega)$ equipped with the metric

$$d((y,\Pi),(z,\Pi')) := \| y - z \| + \| \Pi - \Pi' \|_{TV} .$$

Then the following property follows from Theorems (2.3) and (2.11):

**Corollary (2.13):** Let $g$ be a bounded random variable, then the map

$$(y,\Pi) \mapsto \sigma_t(g, Q_t y, \Pi)$$

is a locally Lipschitz continuous function on the metric space $(L^2 \times \mathcal{M}_b(\Omega), d)$.

**Remarks:** Analogous robustness properties of estimators in the stochastic calculus approach to nonlinear filtering were first formulated by Clark [2] and were recently investigated by Picard [14].

The results of this section provide an extension of remarks by Kallianpur and/break Karandikav [9] (Lemma (5.4) regarding the continuity of the nonlinear filter.

A careful investigation of the proofs shows that the theorems will remain valid if the Euclidean space $\mathbb{R}^m$ in the white noise model is replaced by an infinite-dimensional Hilbert space $\mathbb{K}$. Thus the robustness of the estimator is retained even in a so-called infinite-dimensional white noise model, as considered by Kallianpur and Karandikar [7].

## 3. Properties of the nonlinear filter

As mentioned in the introduction the nonlinear filtering problem is encountered when the random variable $g$ is of the form $g = f(X_t)$ for some measurable function $f$ with finite expectation. Again $E[f(X_t) \mid Q_t y]$ is the optimal estimator in this case and thus for a bounded $f$ all results of Section 2 remain valid.

Rather than estimating functions of $X_t$ one may equivalently obtain the conditional distribution of $X_t$ given the observations, which can be defined as follows.

**Definition (3.11):**

i) For all $B \epsilon S$

(3.2)
$$\Gamma_t(B, Q_t y) := \int I_B(X_t(\omega)) \cdot q_t(\omega, y) \Pi(d\omega)$$

is called the unnormalized conditional distribution of $X_t$. $\Gamma_t(\cdot, Q_t y)$ is a countably additive measure in $\mathcal{M}_b(S)$ and

(3.3) $$\sigma_t(f(X_t), Q_t y, \Pi) = \int_S f(x)\Gamma_t(dx, Q_t y)$$

from all $f$ with $E \mid f(X_t) \mid < \infty$.

ii) The normalized conditional distribution of $X_t$ is defined for all $B \epsilon S$ by

(3.4) $$F_t(B, Q_t y) := \frac{1}{\Gamma_t(S, Q_t y)} \cdot \Gamma_t(B, Q_t y)$$

$F_t(\cdot, Q_t y)$ is a countably additive probability measure $S$ and

$$E[f(X_t) \mid Q_t y] = \int_S f(x) F_t(dx, Q_t y).$$

Considering $\mathcal{M}_b(S)$ endowed with the total variation norm we observe the following robustness property of the unnormalized conditional distribution.

**THEOREM** (3.6): $\Gamma_t(\cdot, Q_t y)$ is locally Lipschitz-continuous, uniformly with respect to $t$. This is to say that

(3.7) $$\sup_{t \epsilon \mathbf{T}} \| \Gamma_t(\cdot, Q_t y) - \Gamma_t(\cdot, Q_t z) \|_{TV} \leq K \cdot \| y - z \|$$

for all $y, z$ in a bounded subset $C$ of $L^2$.

**Proof:** Let $(A_n)_{n \geq 1} \subset S$ be a partition of $S$, then

$$\sum_{n=1}^{\infty} \mid \Gamma_t(A_n, Q_t y) - \Gamma_t(A_n, Q_t z) \mid$$
$$= \sum_{n=1}^{\infty} \mid \sigma_t(I_{A_n}(X_t), Q_t y, \Pi) - \sigma_t(I_{A_n}(X_t), Q_t z, \Pi) \mid.$$

Applying (2.3) to $g = I_{A_n}(X_t)$ we can bound the sum by

(3.8) $$\| y - z \| \cdot (exp\{\frac{1}{2} \| y \|^2\} + exp\{\frac{1}{2} \| z \|^2\}) \cdot \sum_{n=1}^{\infty} \int I_{A_n}(X_t(\omega)) \cdot \| \xi(\omega) \| \Pi(d\omega).$$

The last sum is now equal to $\int \| \xi(\omega) \| \Pi(d\omega)$, so that (3.8) is independent of the choice of the partition $(A_n)_{n \geq 1}$ and also of $t$.

Thus (3.7) holds with $K$ given by (2.10). □

Theorem (3.6) entails an important corollary in the special case when the measure $\Gamma_t(\cdot, Q_t y)$ admits a density with respect to some $\sigma$-finite measure $\mu$ on $(S, \mathcal{S})$. For this let $L^1(\mu) := \{f : (S, \mathcal{S}) \to (\mathbb{R}, \mathcal{B}) : \int | f(x) | \mu(dx) < \infty\}$ and denote by $\| \cdot \|_\mu$ the norm on this space.

Then it is well known (cf. Lang [12]) that for a measure $\nu \epsilon \mathcal{M}_b(S)$ which admits a $\mu$-density $\rho$ we have

$$\| \nu \|_{TV} = \| \rho \|_\mu$$

Thus the proof of the following result follows from Theorem (3.6).

**Corollary** (3.9): Suppose $\mu$ is a $\sigma$-finite measure on $(S, \mathcal{S})$ and that $\Gamma_t(\cdot, Q_t y)$ admits a density $p_t(x, Q_t y), x \epsilon S$ with respect to $\mu$.

Then the map

$$y \mapsto p_t(\cdot, Q_t y)$$

from $L^2$ into $L^1(\mu)$ is locally Lipschitz continuous.

**Remark:** Corollary (3.9) is particularly useful when $(X_t)$ takes values in a Euclidean space and the dominating measure $\mu$ can be taken as Lebesgue measure on this space. This situation is frequently encountered in the filtering of multidimensional diffusion processes. But the result also applies to the filtering of Markov processes with countable statespace. Here one may choose the counting measure as dominating measure. The statement of (3.9) may then be regarded as a robustness property of the conditional probability. This problem was previously addressed in the stochastic calculus model of nonlinear filtering by Clark [2] and Kushner [10].

Finally we will consider the behavior of the filter under approximations of the signal process.

Let $\mathcal{D}(\mathbf{T}, \mathbb{R}^d)$ be the (Skorohod-) space of all right continuous function on $\mathbf{T}$ with valued in $\mathbb{R}^d$, that have left limits everywhere. This space endowed with the Skorohod metric $\rho$ is a Polish space. (cf. Billingsley [1]).

By $\mathcal{C}(\mathbf{T}, \mathbb{R}^d)$ we denote the space of continuous $\mathbb{R}^d$-valued functions on $\mathbf{T}$, which becomes a Polish space with the supremum metric.

If $x$ is an element of $\mathcal{D}$ or $\mathcal{C}$ we denote its value at the point $t \epsilon \mathbf{T}$ by $X(t)$.

We make the following set of assumption (3.10):

i) For all $n \geq 1$ $(X_t^n)_{t \epsilon \mathbf{T}}$ is a Markov process with values in $\mathbb{R}^d$ and almost all sample paths are elements of $\mathcal{D}(\mathbf{T}, \mathbb{R}^d)$

ii) The Markov process $(X_t)_{t \epsilon \mathbf{T}}$ has sample Paths in $\mathcal{C}(\mathbf{T}, \mathbb{R}^d)$ with probability one.

iii) The map from $\mathcal{C}(\mathbf{T}, \mathbb{R}^d)$ into $L^2$ defined by

$$x \mapsto (h_t(x(t))_{t\epsilon\mathbf{T}}$$

is continuous.

If we denote by $P^n$ and $P$ the probability measures induced by $(X_t^n)$ (respectively $(X_t)$) on $\mathcal{D}(\mathbf{T}, \mathbb{R}^d)$ the following theorem shows the continuity of the filter under weak convergence.

**THEOREM** (3.11): Suppose conditions (3.10) are satisfied, and the function $f : \mathbb{R}^d \to \mathbb{R}$ is bounded and continuous. Then for all $y \epsilon L^2$

$$\sigma_t(f(X_t^n), Q_t y, \Pi) \to \sigma_t(f(X_t), Q_t y, \Pi)$$

as $(X_t^n) \xrightarrow{d} (X_t)$.

**Proof:** Setting

$$h(x) := f(x(t)) \cdot exp\{\int_0^t [< y_u, h_u(x(u)) > -\frac{1}{2} \mid h_u(x(u)) \mid^2] du\}$$

we observe that $h$ is continuous with $P$-probability one and

$$\sigma_t(f(X_t^n), Q_t y, \Pi) = \int_D h(x) P^n(dx).$$

Now if $P^n \xrightarrow{\omega} P$ (weak convergence) then

$$\int_D h(x) P^n(dx) \rightarrow \int_D h(x) P(dx) = \sigma_t(f(X_t), Q_t y, \Pi)$$

by Theorem 5.8 of [1] and thus completing the proof. □

In the stochastic calculus model Picard [14] proved a result similar to (3.11) under slightly stronger assumptions on the sequence of signal processes $(X_t^n)$.

Before indicating some applications of the last result we mention that the results of this section will apply to the problems of nonlinear prediction and smoothing as well with minor notational changes.

## 4. Applications

We now consider the so-called discrete approximation problem which presents an effective way to approximate the optimal filter for a diffusion signal process. For illustration purposes we will consider only one-dimensional diffusion processes, but point out that the technique is easily carried over to higher dimensions.

Suppose that the process $(X_t)$ is given by the unique solution of the stochastic differential equation

$$
\begin{aligned}
dX(t) &= b(t, X(t))dt + \sigma(t, X(t))dW_t \\
X(0) &= X_0
\end{aligned} \tag{4.1}
$$

where $b(t,x)$ and $\sigma(t,x)$ are real valued, measurable functions; $(W_t)$ is a one-dimensional standard Wiener process and $X_0$ a random variable independent of $(W_t)$. To formulate the next result we need to introduce the following additional notation.

Let $\mathcal{C}^{1,2}(\mathbf{T} \times \mathbb{R}) := \{f : \mathbf{T} \times \mathbb{R} \to \mathbb{R} : f$ is jointly continuous and has continuous derivatives up to order 1 in $t$ and up to order 2 in $x\}$ and $\mathcal{G}$ be the set of functions $g$ such that for some constants $K$ and $C$

$$
|\, g(t,x) \,| \leq K \cdot exp\{C \cdot |\, x \,|\} \quad \text{for all } t\epsilon\mathbf{T}, x\epsilon\mathbb{R} \tag{4.2}
$$

Further denote by $\mathcal{H}_0$ the space of all Lipschitz continuous functions from $\mathbf{T}$ to $\mathbb{R}$.

We can now state the following theorem which was first proved in [6].

**THEOREM** (4.3): Let $(X_t)$ be given as the solution to (4.1). Assume that with $a(t,x) = \sigma^2(t,x)$ we have that

$$
a, \;\; \frac{\partial}{\partial x}a, \;\; \frac{\partial^2}{\partial x^2}a, \;\; b, \;\; \frac{\partial}{\partial x}b \text{ and } h
$$

are bounded and Lipschitz continuous functions. Further suppose that $a(t,x) > 0$ and that $X_0$ admits a Lebesgue density $\phi\epsilon\mathcal{G}$. Then the unnormalized conditional distribution $\Gamma_t(\cdot, Q_t y)$ of the white noise model admits a Lebesgue density $p_t(x, Q_t y)$, which is the unique solution of the partial differential equation

$$
\begin{aligned}
\frac{\partial}{\partial t}u(t,x) &= a(t,x)\frac{\partial^2}{\partial x^2}u(t,x) + [2 \cdot \frac{\partial}{\partial x}a(t,x) - b(t,x)]\frac{\partial}{\partial x}u(t,x) \\
&\quad + [\frac{\partial^2}{\partial x^2}a(t,x) - \frac{\partial}{\partial x}b(t,x) + < y_t, h_t(x) > -\frac{1}{2} \,|\, h_t(x) \,|^2] \cdot u(t,x) \\
u(0,x) &= \phi(x)
\end{aligned} \tag{4.4}
$$

in the class $\mathcal{C}^{1,2}(\mathbf{T} \times \mathbb{R}) \cap \mathcal{G}$ for all $y\epsilon\mathcal{H}_0$.

**Remark:**
In the light of (3.9) the restriction of theorem (4.3) to observation functions in $\mathcal{H}_0$ provides no principal difficulty, since we can approximate an arbitrary element of $L^2$ by functions in $\mathcal{H}_0$. The denseness of $\mathcal{H}_0$ in $L^2$ and the continuity of the density thus make it possible to compute $p_t(x, Q_t y)$ for all $y\epsilon L^2$.

It is however not this approximation that we want to discuss here. Rather we want to point out that even for Lipschitz continuous observations $y$ the computation of $p_t(x, Q_t y)$ by solving the $PDE$ (4.4) is in all, but the very simple cases, a difficult task. To overcome this difficulty one possibility is to apply available numerical methods for the solution of parabolic $PDE$'s, a route which yet has to be explored.

We want to follow a different approach to this problem, one that has already been suggested in the stochastic calculus model by Clark [2], Kushner [10] and diMasi and Runggaldier [3].

This so-called discrete approximation method replaces the original filtering problem with one that is simpler in its solution and uses this as an approximation to the original problem.

The first step is to approximate the signal process $(X_t)$ by a suitable sequence of simple processes. We follow here the ideas of [3] and define $\mathbb{R}_h := \{z \epsilon \mathbb{R} : \; z = x_0 + n \cdot h, n$ integer $\}$ to be the grid of mesh width $h > 0$ centered at $x_0 \epsilon \mathbb{R}$.

Set

$$\begin{aligned} \lambda_+^h(t,x) &= b^+(t,x)/h + \sigma^2(t,x)/2 \cdot h^2 \\ \lambda_-^h(t,x) &= b^-(t,x)/h + \sigma^2(t,x)/2 \cdot h^2 \\ \lambda^h(t,x) &= - \mid b(t,x) \mid /h - \sigma^2(t,x)/h^2 \end{aligned} \tag{4.5}$$

with $b^+ = max\{b, 0\}$ and $b^- = max\{-b, 0\}$. Then we can define a continuous time birth and death process as follows.

**Definition** (4.6): Let $(X_t^h)$ be an $\mathbb{R}_h$-valued Markov process such that $\Pi\{X_0^h = x_0\} = 1$ and the transition probabilities are given by

$$\Pi\{X_{t+\varepsilon}^h = x + h \mid X_t^h = x\} = \int_t^{t+\varepsilon} \lambda_+^h(u,x)du$$

$$\Pi\{X_{t+\varepsilon}^h = x - h \mid X_t^h = x\} = \int_t^{t+\varepsilon} \lambda_-^h(u,x)du$$

$$\sum_{\substack{n \epsilon \mathbb{Z} \\ n \neq -1,0,+1}} \Pi\{X_{t+\varepsilon}^h = x + n \cdot h \mid X_t^h = x\} = o(\varepsilon)$$

$$\Pi\{X_{t+\varepsilon}^h = x \mid X_t^h = x\} = 1 + \int_t^{t+\varepsilon} \lambda^h(u,x)du - o(\varepsilon).$$

The Markov process thus defined is time inhomogeneous process with discrete state space and almost all paths in $\mathcal{D}(\mathbf{T}, \mathbb{R})$.

If we assume that the initial condition in (4.1) is degenerate, i. e. $X_0 = x_0$ w. p. 1, then as $h \to 0$ $(X^h) \xrightarrow{d} (X)$ in $\mathcal{D}(\mathbf{T}, \mathbb{R})$, cf. [3]. Since $(X_t^n)$ is $\mathbb{R}_h$-valued we can solve the nonlinear filtering problem by finding the unnormalized conditional probabilities

$$\pi_t^h(x, Q_t y) = \sigma_t(I_{\{x\}}(X_t^h), Q_t y) \tag{4.7}$$

for all $x \epsilon \mathbb{R}_h$. These can now be found by solving a system of linear ordinary differential equations.

**THEOREM** (4.8): Under the same condition as (4.3) the unnormalized filtering distribution
$(\pi_t^h(x, Q_t y))_{x \epsilon \mathbb{R}_h}$ is the unique solution (within the class of summable sequences) of the system of equations

$$(4.9) \qquad \frac{d}{dt}u(t,x) = \lambda_+^h(t, x-h) \cdot u(t, x-h) + \lambda^h(t,x) \cdot u(t,x) + \lambda_-^h(t, x+h) \cdot u(t, x+h) + \{ < y_t, h_t(x) > -\frac{1}{2} \mid h_t(x) \mid^2 \} \cdot u(t,x)$$

$$u(0,x) = I_{\{x_o\}}(x).$$

The proof of this theorem follows from a more general result proved in [4].

**Remark:** The filtering problem for the signal process $(X_t^h)$ thus leads only to ordinary differential equation for the conditional probabilities. Although we cannot use $\pi_t(x, Q_t y)$ directly as an approximation to $p_t(x, Q_t y)$, Theorem (3.11) provides the link between the two filtering problems.

**THEOREM** (2.1): Suppose that the function $h_t(x)$ satisfies the condition (3.10) and that $f$ is bounded. Then with

$$\sigma_t(f(X_t^h), Q_t y) = \sum_{x \epsilon \mathbb{R}_h} f(x) \pi_t^h(x, Q_t y)$$

and

$$\sigma_t(f(X_t), Q_t y) = \int f(x) \cdot p_t(x, Q_t y) dx$$

$$(4.11) \qquad \sigma_t(f(X_t^h), Q_t y) \;\to\; \sigma_t(f(X_t), Q_t y)$$

as $h \to 0$.

The proof is a straightforward application of (3.11) to the approximating sequence $(X_t^h)$.

Discrete state approximations for multi-dimensional diffusions are also discussed in [10] and [11]. For these cases the results of this section still remain valid with minor notational changes.

## 5. References


[1] BILLINGSLEY, P. (1968). Convergence of Probability Measures. John Wiley, New York.

[2] CLARK, J. M. C. (1978). The design of robust approximations to the stochastic differential equations of nonlinear filtering. In Communication Systems and Random Process Theory, ed. J. K. Skwirzynski, Sijthoff and Nordhoff, Alphen aan den Rijn.

[3] DI MASI, G. B. and RUNGGALDIER, W. J. (1981). Continuous time approximations for the nonlinear filtering problem. *Appl. Math. and Optim.* (7), pp. 233–245.

[4] HUCKE, H.-P. (1985). Estimation of continuous time Markov processes in a finitely additive white noise model. CSP Technical Report No. 121, Dept. of Statistics, Univ. of North Carolina.

[5] KALLIANPUR, G. (1980). Stochastic Filtering Theory. Springer, Berlin.

[6] KALLIANPUR, G. and KARANDIKAR, R. L. (1983). A finitely additive white noise approach to nonlinear filtering. *Appl. Math. and Optimization* (10), pp. 159–185.

[7] KALLIANPUR, G. and KARANDIKAR, R. L. (1984). Measure valued equations for the optimal filter in finitely additive nonlinear filtering theory. Z. W. (66), pp. 1–17.

[8] KALLIANPUR, G. and KARANDIKAR, R. L. (1984). The non-linear filtering problem for the unbounded case. *Stoch. Proc. and Applications* (18), pp. 57–66.

[9] KALLIANPUR, G. and KARANDIKAR, R. L. (1985). White noise calculus and nonlinear filtering theory. *Annals of Probability* (13), pp. 1033–1107.

[10] KUSHNER, H. J. (1979). A robust discrete state approximation to the optimal nonlinear filter for a diffusion. *Stochastics* (3), pp. 75–83.

[11] KUSHNER, H. J. and DI MASI, G. (1978). Approximation of Functionals and optimal control problems on jump diffusion processes. *J. Math. Anal. and Appl.* (63), pp. 772–800.

[12] LANG, S. (1983). Real Analysis. 2nd ed., Addison-Wesley, Reading.

[13] LIPTSER, R. S. and SHIRYAYEV, A. N. (1977, 1978). Statistics of Random Processes. Vols. I and II, Springer, New York.

[14] PICARD, J. (1984). Robustness de la solution des Problemes de filtrage avec limit blanc independant. *Stochastics* (13), pp. 229–245.