SIAM J. CONTROL AND OPTIMIZATION
Vol. 20, No. 4, July 1982

0363-0129/82/2004-0004 $01.00/0



# ON SOME PROPERTIES OF CONDITIONAL MOMENTS IN NONLINEAR FILTERING*

JACOB HAMMER†

**Abstract.** The nonlinear filtering problem of diffusion processes embedded in additive white noise is considered. It is shown that the paths of all conditional moments of the measurement function can be causally calculated when the path of its first conditional moment is known. The formulas involved in this calculation are independent of the specific parameters of the information process.

In addition, asymptotic properties of the nonlinear filter, as the signal to noise ratio approaches infinity, are also considered. It is shown that, asymptotically, the deviation from Gaussian properties is of the order of the noise to signal ratio at most.

**1. Introduction.** The optimal filtering problem of diffusion processes embedded in additive white noise is stated as follows. Let $(\Omega, \mathcal{F}, P)$ be a probability space on the set $\Omega$, with complete $\sigma$-field $\mathcal{F}$, and probability measure $P$. Further, let $R$ denote the real numbers, and let $[0, T]$ be the set of all points $t \in R$ satisfying $0 \leqq t \leqq T$. We assume that there exists a Brownian motion $B:(\Omega, \mathcal{F}, P)\times[0, T]\rightarrow R$. Next, let $x:(\Omega, \mathcal{F}, P)\times[0, T]\rightarrow R$ be a diffusion process, given by the following stochastic differential equation:

$$\begin{aligned} dx_t &= m(x_t, t)\,dt+\sigma(x_t, t)\,dB_t, \qquad t\in[0, T], \\ x_{t=0} &= x_0, \end{aligned} \tag{1.1}$$

where $m, \sigma: R\times[0, T]\rightarrow R$. We assume that the functions $m, \sigma$ satisfy, for every $(x, t)\in R\times[0, T]$, the uniform Lipschitz (or linear growth) condition

$$m^2(x, t)+\sigma^2(x, t)\leqq K(1+x^2) \tag{1.2}$$

for a suitable constant $K\geqq 0$. We also assume that the initial condition $x_0$ is stochastically independent of $B_t$ for every $t\in[0, T]$ and satisfies, for all integers $k\geqq 0$, $E\{x_0^{2k}\}<\infty$ (where $E\{\cdot\}$ denotes the expectation). Under these conditions (see Gikhman and Skorokhod [1972]), the solution $x_t$ of (1.1) is almost surely unique, almost surely continuous in $t$, and, for every integer $k\geqq 0$ and every $t\in[0, T]$, we have that $E\{x_t^{2k}\}<\infty$. We shall refer to $x$ as the *information process*.

Next, let $W$ be a standard Brownian motion on $(\Omega, \mathcal{F}, P)\times[0, T]$, and assume that, for all $t\in[0, T]$, $\{W_t\}$ is stochastically independent of $\{B_t\}$. Also, let $g: R\rightarrow R$ be a twice continuously differentiable function satisfying, for every $t\in[0, T]$, the following conditions: (i) $E\{g^{2k}(x_t)\}<\infty$ for every integer $k\geqq 0$; (ii) $E\{[g'(x_t)]^2\}<\infty$, and (iii) $E\{[g''(x_t)]^2\}<\infty$, where $g'$ (resp. $g''$) denotes the first (resp. second) derivative of $g$. Finally, let $\lambda$ be a positive real number. Then, the *measurement process* $y$ is defined as

$$\begin{aligned} dy_t &= \lambda g(x_t)\,dt+dW_t, \qquad t\in[0, T], \\ y_{t=0} &= y_0, \end{aligned} \tag{1.3}$$

* Received by the editors March 30, 1981, and in revised form August 5, 1981.

† Center for Mathematical System Theory, University of Florida, Gainesville, Florida 32611. This work was done while the author was with the Department of Electrical Engineering, Technion-Israel Institute of Technology, Haifa, Israel. The present revision was supported in part by U.S. Army research grant DAAG29-80 C0050 and US Air Force grant AFOSR76-3034D through the Center for Mathematical System Theory, University of Florida, Gainesville.

where the initial condition $y_0$ is stochastically independent of both $B_t$ and $W_t$ for all $t \in [0, T]$. The function $g$ will be called the *measurement function*. If $g$ is the identity function (that is, $g(x) = x$), then we say that (1.3) is a *linear measurement* process.

Having defined the information process $x$ and the measurement process $y$, we can now state the classical nonlinear filtering problem. Let $f: R \to R$ be a twice continuously differentiable function, and assume that, for every $t \in [0, T]$, we have that $E\{[f(x_t)]^2\} < \infty$, $E\{[f'(x_t)]^2\} < \infty$ and $E\{[f''(x_t)]^2\} < \infty$. Also, let $y_0^t := \{y_\theta \mid \theta \in [0, t]\}$ denote a sample of the measurement process during the time interval $[0, t]$. Then, given $y_0^t$, an estimate $\hat{f}(x_t)$ for $f(x_t)$ is sought such that, for any other estimate $\tilde{f}(x_t)$, the following holds: $E\{[f(x_t) - \hat{f}(x_t)]^2\} \leqq E\{[f(x_t) - \tilde{f}(x_t)]^2\}$.

The solution to the nonlinear filtering problem is well known, and is given by the conditional expectation

$$\hat{f}(x_t) = E\{f(x_t) \mid y_0^t\},$$

conditioned on the $\sigma$-field generated by $y_0^t$. This solution can also be represented in the following form (see Fujisaki, Kallianpur and Kunita [1972]):

$$\begin{aligned} \hat{f}(x_t) - \hat{f}(x_0) = \int_0^t [\, \widehat{m(x_u, u) f'(x_u)} + \widehat{\tfrac{1}{2}\sigma^2(x_u, u) f''(x_u)}\,]\, du \\ + \lambda \int_0^t [\widehat{f(x_u) g(x_u)} - \hat{f}(x_u)\hat{g}(x_u)]\, d\nu_u, \end{aligned} \tag{1.4}$$

where $\nu$ is the *innovation process* and is given by

$$d\nu_t = dy_t - \lambda \hat{g}(x_t)\, dt. \tag{1.5}$$

It is also known (see op. cit.) that the *innovation process $\nu$ is a standard Brownian motion*.

Various aspects of the nonlinear filtering problem are considered in Stratonovich [1960], Kushner [1967], Zakai [1969] and [1975], Kailath [1969], Jazwinski [1970], Fujisaki, Kalliapur and Kunita [1972] and others. In the present paper we examine the nonlinear filtering problem using a linear derivative-type operator, which we call the *martingale derivative*. We summarize below the main results.

Let $\widehat{g_t^k} := E\{g^k(x_t) \mid y_0^t\}$, where $k$ is a positive integer, be the $k$th conditional moment of the measurement function. Also, assume that the path $\hat{g}_0^t$ of a specific sample of the first conditional moment is known. Then, we show that the paths of all other conditional moments $(\widehat{g^k})_0^t$, $k = 2, 3, \cdots$, (related to the same sample $y_0^t$) can be causally calculated using the path $\hat{g}_0^t$ only. Moreover, the formulas relating $(\widehat{g^k})_0^t$ and $\hat{g}_0^t$ are independent of the functions $m(x, t)$ and $\sigma(x, t)$ determining the information process $x$ in (1.1). Thus, this calculation can be performed even in cases where a detailed description of the information process is not available.

As an additional application of the martingale derivative, we consider the asymptotic behavior of the nonlinear filter as the constant $\lambda$ in (1.3) approaches infinity. Informally, this is equivalent to the consideration of filters under conditions of high signal to noise ratio. We show that, as $\lambda \to \infty$, the conditional probability measure of $g(x_t)$, conditioned on the $\sigma$-field generated by $y_0^t$, approaches a Gaussian measure at the "rate" of $1/\lambda$ at least.

The paper is organized as follows. In § 2 we define the martingale derivative, and in § 3 we apply it to obtain a representation of all conditional moments in terms of martingale derivatives of the first one. The paper is concluded in § 4 with a consideration of some asymptotic properties as the signal to noise ratio approaches infinity.

**2. The martingale derivative.** The martingale derivative is a linear operator defined on a certain class of functions of unbounded variation. Its properties are similar to the properties of the usual derivative. First we establish our notation. Let $(\Omega, \mathcal{F}, P)$ be a complete probability space. Also, let $\{\mathcal{F}_t\}$, $t \in [0, T]$, be an increasing family of complete $\sigma$-fields, which is continuous from the right and satisfies $\mathcal{F}_T \subset \mathcal{F}$. We denote by $W$ a standard Brownian motion on $(\Omega, \mathcal{F}, P; \mathcal{F}_t)$, $0 \leqq t \leqq T$. Further, let $f_t$ and $h_t$ be square integrable semimartingales of the form

$$f_t = a_t + \int_0^t \phi_s \, dW_s, \qquad h_t = b_t + \int_0^t \psi_s \, dW_s, \tag{2.1}$$

where $\phi_t$ and $\psi_t$ are well measurable with respect to the family $\{\mathcal{F}_t\}$, $\int_0^T E\{\phi_s^2\}\, ds < \infty$ and $\int_0^T E\{\psi_s^2\}\, ds < \infty$ and, finally, $a_t$ and $b_t$ are adapted to $\mathcal{F}_t$ and differentiable on $[0, T]$. Denote by $[f, h]_t := \int_0^t \phi_s \psi_s \, ds$ the cross quadratic variation of $f$ and $h$ (e.g., Meyer [1975, Ch. 3]). We next state a formal definition of the martingale derivative, and, immediately afterwards, we give an interpretation of this definition in the particular case which is of main interest to us. (We note that $[h, h]_t$ is almost surely strictly increasing in $t$ if and only if $\psi_t \neq 0$ almost surely for all $t \in [0, T]$.)

DEFINITION 2.2. Let $f_t$ and $h_t$ be as in (2.1), and assume that the quadratic variation $[h, h]_t$ is almost surely strictly increasing in $t$. Then, the *martingale derivative* $f_t^h$ of $f_t$ with respect to $h_t$ is

$$f_t^h := \lim_{\substack{\Delta \to 0 \\ \Delta > 0}} \frac{[f, h]_t - [f, h]_{t-\Delta}}{[h, h]_t - [h, h]_{t-\Delta}},$$

on the optional $\sigma$-field on $\Omega \times [0, T]$.

Assume now in (2.1) that, for all $t \in [0, T]$, the processes $\phi_t$ and $\psi_t$ are almost surely continuous and $\psi_t \neq 0$ almost surely. Then, we have that

$$f_t^h = \frac{\phi_t}{\psi_t}, \quad \text{and} \quad f_t^W = \phi_t.$$

In cases where it is possible to iterate the martingale derivative, we shall adopt the following notation:

$$f_t^{\{k+1\}h} := (f^{\{k\}h})_t^h, \qquad f_t^{\{0\}h} := f_t, \tag{2.3}$$

where $k \geqq 0$ is an integer.

Next, we list a series of simple properties of the martingale derivative, showing that it obeys the usual differentiation rules. To this end, we let $f_{1_t}, \cdots, f_{n_t}, h_t, l_t$ be a set of semimartingales of the form (2.1), and assume that both of the quadratic variations $[h, h]_t$ and $[l, l]_t$ are almost surely strictly increasing in $t$. Further, we let $c_t$ be a stochastic process adapted to $\{\mathcal{F}_t\}$ and almost surely differentiable with respect to $t$, for every $t \in [0, T]$. Finally, we let $H: R^n \times [0, T] \to R$ be a function twice continuously differentiable in its first $n$ arguments, and differentiable in the last one. Then, using the Ito differential formula (e.g., Gikhman and Skorokhod [1972]), the following (2.4) to (2.8) can readily verified:

$$(f_1 + cf_2)_t^h = f_{1_t}^h + c_t f_{2_t}^h, \tag{2.4}$$

$$(f_1 f_2)_t^h = f_{1_t}^h f_{2_t} + f_{1_t} f_{2_t}^h, \tag{2.5}$$

(2.6) if $f_{2_t} \neq 0$ almost surely for every $t \in [0, T]$, then $\left(\frac{f_1}{f_2}\right)^h_t = \frac{f^h_{1_t} f_{2_t} - f_{1_t} f^h_{2_t}}{(f^2_{2_t})}$,

$$f^l_{1_t} = f^h_{1_t} h^l_t, \tag{2.7}$$

$$[H(f_{1_t}, f_{2_t}, \cdots, f_{n_t}, t)]^h = \sum_{i=1}^n \left[\frac{\partial H(f_{1_t}, \cdots, f_{n_t}, t)}{\partial f_{i_t}}\right] f^h_{i_t}. \tag{2.8}$$

Thus, the usual rules of calculus apply to the martingale derivative.

**3. A representation result for conditional moments.** We consider the nonlinear filtering problem with information process (1.1) and measurement process (1.3). In this section we let $\lambda = 1$ in (1.3). As before, we denote by $g(x)$ the measurement function and, for every integer $k \geqq 0$, $\widehat{g^k_t} := E\{g^k(x_t) \mid y^t_0\}$ is its $k$th conditional moment. We examine first the relation between the conditional moments of $g(x_t)$ and the martingale derivatives, with respect to the innovation process $\nu$, of $\hat{g}_t$. It turns out that $\widehat{g^k_t}$ is a polynomial function of the martingale derivatives $\hat{g}_t, \hat{g}^\nu_t, \cdots, \hat{g}^{\{k-1\}\nu}_t$. Moreover, the $k$th martingale derivative $\hat{g}^{\{k\}\nu}_t$ can also be expressed as a polynomial function of the conditional moments $\hat{g}_t, \cdots, \widehat{g^{k+1}_t}$. Thus, we encounter the interesting situation where an infinite set of polynomial equations has a polynomial solution.

We next define two families of multivariable polynomials, $P_k(x_1, \cdots, x_k)$ and $G_k(x_1, \cdots, x_{k+1})$, where $k = 0, 1, \cdots$, by the following recursive formulas:

$$\begin{aligned} &P_0 = 1, \\ &P_{k+1}(x_1, \cdots, x_{k+1}) = x_1 P_k(x_1, \cdots, x_k) + \sum_{i=1}^k \left[\frac{\partial P_k(x_1, \cdots, x_k)}{\partial x_i}\right] x_{i+1}, \end{aligned} \tag{3.1}$$

and

$$\begin{aligned} &G_0(x_1) = x_1, \\ &G_{k+1}(x_1, \cdots, x_{k+2}) = \sum_{i=1}^{k+1} \left[\frac{\partial G_k(x_1, \cdots, x_{k+1})}{\partial x_i}\right] (x_{i+1} - x_i x_1). \end{aligned} \tag{3.2}$$

Now, we can state the following:

THEOREM 3.3. *Given a nonlinear filtering problem with information process* (1.1) *and measurement process* (1.3) (*where* $\lambda = 1$), *the following hold*:

(i) $\widehat{g^k_t} = P_k(\hat{g}_t, \cdots, \hat{g}^{\{k-1\}\nu}_t)$;

(ii) $\hat{g}^{\{k\}\nu}_t = G_k(\hat{g}_t, \cdots, \widehat{g^{k+1}_t})$.

*Proof.* We remark that, since $\lambda = 1$, it follows by (1.4) that the martingale derivative of the $k$th conditional moment is $\widehat{g^k_t}^\nu = \widehat{g^{k+1}_t} - \widehat{g^k_t}\hat{g}_t$. We use this formula in the proof of (i) and (ii).

(i) Evidently, $\widehat{g^0_t} = P_0 = 1$, and we assume, by induction, that $\widehat{g^n_t} = P_n(\hat{g}_t, \cdots, \hat{g}^{\{n-1\}\nu}_t)$ for an integer $n \geqq 0$. Then using (2.8), we obtain that

$$\begin{aligned} \widehat{g^{n+1}_t} = \hat{g}_t \widehat{g^n_t} + \widehat{g^n}^\nu_t &= \hat{g}_t P_n(\hat{g}_t, \cdots, \hat{g}^{\{n-1\}\nu}_t) + \sum_{i=0}^{n-1} \left[\frac{\partial P_n}{\partial \hat{g}^{\{i\}\nu}_t}\right] (\hat{g}^{\{i\}\nu}_t)^\nu \\ &= \hat{g}_t P_n(\hat{g}_t, \cdots, \hat{g}^{\{n-1\}\nu}_t) + \sum_{i=0}^{n-1} \left[\frac{\partial P_n}{\partial \hat{g}^{\{i\}\nu}_t}\right] \hat{g}^{\{i+1\}\nu}_t. \end{aligned}$$

Hence, $\widehat{g^{n+1}_t} = P_{n+1}(\hat{g}_t, \cdots, \hat{g}^{\{n\}\nu}_t)$, and (i) follows.

(ii) Again, identically, $\hat{g}_t^{\{0\}\nu} = \hat{g}_t = G_0(\hat{g}_t)$ and, by induction, we assume that (ii) holds for an integer $n \geqq 0$. Then, still using (2.8), we obtain

$$\hat{g}_t^{\{n+1\}\nu} = [G_n(\hat{g}_t, \cdots, \widehat{g_t^{n+1}})]^\nu = \sum_{i=1}^{n+1}\left[\frac{\partial G_n}{\partial \widehat{g_t^i}}\right](\widehat{g_t^i})^\nu$$
$$= \sum_{i=1}^{n+1}\left[\frac{\partial G_n}{\partial \widehat{g_t^i}}\right](\widehat{g_t^{i+1}} - \widehat{g_t^i}\hat{g}_t) = G_{n+1}(\hat{g}_t, \cdots, \widehat{g_t^{n+2}}).$$

Hence, (ii) holds for $n+1$, and our proof is concluded. □

Let $f_t$ be a semimartingale of the form (2.1). In general, the calculation of the martingale derivative $f_t^W$ involves the use of both a sample of $f_t$ and the corresponding sample of $W_t$. However, in case of the nonlinear filtering problem, the situation turns out to be different. We next show that, for every $t \in [0, T]$, the martingale derivative $\hat{g}_t^\nu$ *of* $\hat{g}_t$ is completely determined by the sample $\hat{g}_0^t$. No explicit information on the sample $\nu_0^t$ is required. In fact, more is true.

THEOREM 3.4. *Let $g(x)$ be the measurement function, and $\nu$ the innovation process for the filtering problem described by* (1.1) *and* (1.3). *Then, for every integer $k \geqq 0$ and for every $t \in [0, T]$, the* kth *martingale derivative $\hat{g}_t^{\{k\}\nu}$ is determined by the sample $\hat{g}_{t-\alpha}^t$, where $0 < \alpha < t$.*

*Proof.* First we note that Theorem 3.3(ii) implies that all iterated martingale derivatives $\hat{g}_t^{\{k\}\nu}$, $k = 0, 1, \cdots$, are almost surely continuous for all $t \in [0, T]$. For the sake of simplicity, we shall consider below only *continuous samples*. The null sets on which continuity does not hold can be dealt with by a standard "countable diagonal set" method.

By the definition of the martingale derivative, we have that $[\hat{g}^{\{n\}\nu}, \hat{g}]_t - [\hat{g}^{\{n\}\nu}, \hat{g}]_{t-\alpha} = \int_{t-\alpha}^t \hat{g}_s^{\{n+1\}\nu}\hat{g}_s^\nu\,ds$. Hence it follows by continuity that, for every $s \in [t-\alpha, t]$, the quantity $h_s^n := \hat{g}_s^{\{n+1\}\nu}\hat{g}_s^\nu$ is determined by the paths $(\hat{g}^{\{n\}\nu})_{t-\alpha}^t$ and $\hat{g}_{t-\alpha}^t$. We show next, by recursion, that this fact implies that all martingale derivatives can be calculated as required.

First, in case $n = 0$, we obtain $h_s^0 = (\hat{g}_s^\nu)^2$. Now, by (1.4), $\hat{g}_t^\nu = \lambda(\widehat{g_t^2} - \hat{g}_t^2)$ so that, since $\lambda > 0$, it follows by the Jensen inequality that $\hat{g}_t^\nu \geqq 0$ for all $t \in [0, T]$. Consequently, $\hat{g}_s^\nu$ is determined by $h_s^0$, and thus the path $(\hat{g}^\nu)_{t-\alpha}^t$ is determined by the path $\hat{g}_{t-\alpha}^t$.

Further, by recursion, we assume that, for some integer $n \geqq 0$, the path $(\hat{g}^{\{n\}\nu})_{t-\alpha}^t$ is determined by $\hat{g}_{t-\alpha}^t$. Then, for every $s \in [t-\alpha, t]$, the quantity $h_s^n$ is clearly determined by $\hat{g}_{t-\alpha}^t$. Hence, if $\hat{g}_s^\nu \neq 0$, then the martingale derivative $\hat{g}_s^{\{n+1\}\nu}$ is determined by $\hat{g}_{t-\alpha}^t$. Thus, it remains to consider the case $\hat{g}_s^\nu = 0$, which we next do. By continuity, it follows that there exists an element $\delta > 0$ such that either (i) $\hat{g}_u^\nu = 0$ for all $u \in [s-\delta, s]$; or (ii) $\hat{g}_u^\nu \neq 0$ for all $u \in [s-\delta, s)$. But then, in subcase (i) we have $\hat{g}_u^{\{k\}\nu} = 0$ for all $u \in [s-\delta, s]$ and $k = 2, 3, \cdots$. In subcase (ii), we have already shown that $\hat{g}_u^{\{n+1\}\nu}$ is determined by $\hat{g}_{t-\alpha}^t$ for all $u \in [s-\delta, s)$. Hence it follows, by continuity, that $\hat{g}_s^{\{n+1\}\nu}$ is determined as well. □

Combining Theorems 3.3(i) and 3.4, we directly obtain the following:

COROLLARY 3.5. *Let $g(x)$ be the measurement function for the filtering problem of* (1.1) *with measurement* (1.3). *Then, for every integer $k \geqq 0$ and for every $t \in [0, T]$, the conditional moment $\widehat{g_t^k}$ is determined by the path $\hat{g}_{t-\alpha}^t$, where $0 < \alpha < t$.*

It is interesting to note that the formulas involved in the calculation of the conditional moment $\widehat{g_t^k}$ from the first conditional moment path $\hat{g}_{t-\alpha}^t$, as described in Theorems 3.3 and 3.4, are independent of the functions $m(x, t)$ and $\sigma(x, t)$ which determine the information process (1.1). In fact, it can be shown that the same calculation scheme is valid for information processes more general than (1.1) as well.

Before concluding this section, we note that Corollary 3.5 can be generalized in the following sense. Let $f: R \to R$ be a continuous function, measurable with respect to the $\sigma$-field induced by $g$ on $R$. Then, for every $t \in [0, T]$, $\hat{f}(x_t)$ is determined by the path $\hat{g}^t_{t-\alpha}$, where $0 < \alpha < t$.

**4. Some asymptotic properties.** We start with a closer examination of the polynomials $G_n(x_1, \cdots, x_{n+1})$ of (3.2). Let $z$ be a random variable on the probability space $(\Omega, \mathcal{F}, P)$. We denote by $\overline{z^i}$ the $i$th moment of $z$. We next show that the polynomials $G_n$ are closely related to the Gaussian probability law, as follows.

LEMMA 4.1. *A random variable $z$ has a Gaussian distribution function if and only if, for every integer $n \geqq 2$, $G_n(\bar{z}, \cdots, \overline{z^n}, \overline{z^{n+1}}) = 0$.*

*Proof.* Assume first that $z$ is a Gaussian random variable. Then, all moments of $z$ are determined by the first two moments $\bar{z}$ and $\overline{z^2}$, so that, for every integer $i \geqq 0$, $\overline{z^i} = \overline{z^i}(\bar{z}, \overline{z^2})$. We let $\sigma^2 := \overline{(z-\bar{z})^2} \neq 0$, and consider the Gaussian random variable $z_\varepsilon$ defined as follows: $\bar{z}_\varepsilon = \bar{z} + \varepsilon$ and $\overline{(z_\varepsilon - \bar{z}_\varepsilon)^2} = \sigma^2$. Then, we have

$$\left.\frac{d\overline{z^i_\varepsilon}}{d\varepsilon}\right|_{\varepsilon=0} = (2\pi\sigma^2)^{-1/2} \frac{d}{d\varepsilon}\left.\left\{\int_{-\infty}^{\infty} u^i \exp\left[\frac{(u-\bar{z}-\varepsilon)^2}{2\sigma^2}\right] du\right\}\right|_{\varepsilon=0} = \frac{(\overline{z^{i+1}} - \overline{z^i}\bar{z})}{\sigma^2}.$$

Now, since the Gaussian distribution is symmetric, and since $G_2(\bar{z}, \overline{z^2}, \overline{z^3}) = \overline{(z-\bar{z})^3}$, we clearly have that $G_2(\bar{z}, \overline{z^2}, \overline{z^3}) = 0$. By induction, we assume now that there is an integer $n \geqq 2$ such that, for every Gaussian random variable $z_1$, $G_n(\bar{z}_1, \cdots, \overline{z_1^{n+1}}) = 0$. In particular, it follows then that, in the case $z_1 = z_\varepsilon$, $G_n(\bar{z}_\varepsilon, \cdots, \overline{z_\varepsilon^{n+1}}) = 0$ for every $\varepsilon$. Hence, $dG_n(\bar{z}_\varepsilon, \cdots, \overline{z_\varepsilon^{n+1}})/d\varepsilon = 0$ for every $\varepsilon$ as well. Now, by a direct calculation, we have $0 = dG_n(\bar{z}_\varepsilon, \cdots, \overline{z_\varepsilon^{n+1}})/d\varepsilon|_{\varepsilon=0} = (1/\sigma^2)[G_{n+1}(\bar{z}, \cdots, \overline{z^{n+2}})]$, which implies the necessity of our assertion.

Conversely, if $G_n(\bar{z}, \cdots, \overline{z^{n+1}}) = 0$ for every $n = 2, 3, \cdots$, then, since $G_n$ is monic in $\overline{z^{n+1}}$, it follows that all moments $\overline{z^i}$ of $z$ are determined by the first two moments $\bar{z}$ and $\overline{z^2}$. Moreover, by our previous discussion it is clear that the functions $\overline{z^i} = \overline{z^i}(\bar{z}, \overline{z^2})$ thus obtained are identical to those for the Gaussian case. Hence, $z$ has a Gaussian characteristic function, and our proof concludes. □

Motivated by Lemma 4.1, we shall call the polynomials $G_n$ of (3.2) the *Gaussian polynomials.*

*Example.* The first Gaussian polynomials are as follows:

$$G_0 = \bar{z}, \quad G_1 = \overline{(z-\bar{z})^2}, \quad G_2 = \overline{(z-\bar{z})^3}, \quad G_3 = \overline{(z-\bar{z})^4} - 3\left[\overline{(z-\bar{z})^2}\right]^2.$$

We return now to the nonlinear filtering problem of the information process (1.1) with the measurement process (1.3). Let $\hat{P}_t : R \to [0, 1]$ denote the conditional probability distribution of $g(x_t)$, conditioned on the $\sigma$-field generated by the measurement process $y_0^t$. Then, we have $\widehat{g_t^n} = \int u^n \, d\hat{P}_t(u)$. As usual, we shall say that $\hat{P}_t$ is *symmetric* if it satisfies the following. For every function $f: R \to R$ that satisfies $f(x - \hat{g}_t) = -f(\hat{g}_t - x)$ for all $x \in R$, one has that $\int f(u) \, d\hat{P}_t(u) = 0$. As a direct consequence of Theorem 3.3 and Lemma 4.1, we can now show that if $\hat{P}_t$ is symmetric, then it is necessarily Gaussian. This is proved in the following:

PROPOSITION 4.2. *The conditional probability measure $\hat{P}_t$ is almost surely symmetric for all $t \in [0, T]$ if and only if it is almost surely Gaussian for all $t \in [0, T]$.*

*Proof.* Assume first that $\hat{P}_t$ is symmetric. Then evidently, $\widehat{(g_t - \hat{g}_t)^3} = 0$ almost surely, so that $G_2(\hat{g}_t, \widehat{g_t^2}, \widehat{g_t^3}) = 0$ almost surely for all $t \in [0, T]$. But then, since by

Theorem 3.3, $\hat{g}_t^{\{2\}\nu} = G_2(\hat{g}_t, \widehat{g_t^2}, \widehat{g_t^3})$, we have that $\hat{g}_t^{\{2\}\nu} = 0$ almost surely for all $t \in [0, T]$. Hence, also $\hat{g}_t^{\{n\}\nu} = 0$ almost surely for all integers $n \geqq 2$ and all $t \in [0, T]$. Again, by Theorem 3.3, this implies that $G_n(\hat{g}_t, \cdots, \widehat{g_t^{n+1}}) = 0$ for all $n \geqq 2$ and all $t \in [0, T]$, so that, by Lemma 4.1, $\hat{P}_t$ is Gaussian. The converse direction is immediate. □

We consider next the asymptotic behavior of the conditional probability measure $\hat{P}_t$ as the constant $\lambda$ in (1.3) approaches infinity. Explicitly, we shall show that, as $\lambda \to \infty$, the probability law determined by $\hat{P}_t$ approaches the Gaussian probability law at the rate of $1/\lambda$ at least. To this end, we need the following notation. Let $f:[0, T] \to R$ be a function, which implicitly depends on $\lambda$. We shall say that $f \sim 1/\lambda$ if, for almost every $t \in [0, T]$, the following holds: For every $\alpha > 0$, $\lim_{\lambda \to \infty} \lambda^{1-\alpha} f(t) = 0$.

As before, we let $g(x)$ be the measurement function, and denote $\widehat{g_t^i} = E\{g^i(x_t) \mid y_0^t\}$. Also, $G_n$, $n = 0, 1, 2, \cdots$, are the Gaussian polynomials defined in (3.2). Clearly, as $\lambda \to \infty$, the conditional probability measure $\hat{P}_t$ degenerates into a deterministic measure. Thus, we expect by Lemma 4.1 that, for $n \geqq 1$, one should have $\lim_{\lambda \to \infty} E \mid G_n(\hat{g}_t, \cdots, \widehat{g_t^{n+1}}) \mid = 0$. In fact, the following stronger result is valid.

THEOREM 4.3. *Given the nonlinear filtering problem of the information process* (1.1) *with measurement process* (1.3), *the following holds true*: *For every* $0 \leqq \delta < 1$,

$$E^{1/(1+\delta)} |G_n(\hat{g}_t, \cdots, \widehat{g_t^{n+1}})|^{(1+\delta)} \sim \frac{1}{\lambda},$$

*where* $n = 1, 2, 3, \cdots$, *and* $t \in [0, T]$.

*Proof.* We first note that, if the condition $\lambda = 1$ in Theorem 3.3 is relaxed, then, for all $k = 0, 1, \cdots$ and $t \in [0, T]$, we have

$$\hat{g}_t^{\{k+1\}\nu} = \lambda^{k+1} G_{k+1}(\hat{g}_t, \cdots, \widehat{g_t^{k+2}}) = \lambda^k \sum_{i=1}^{k+1} \left(\frac{\partial G_k}{\partial \widehat{g_t^i}}\right) \widehat{g_t^{i\nu}},$$

where the last equality follows by (3.2) and (1.4). Then, applying the Minkowski inequality (e.g., Loève [1963, Ch. 3]), we obtain that, for every $0 \leqq \delta < 1$,

$$\begin{aligned} f_t := E^{1/(1+\delta)} |\lambda G_{k+1}(\hat{g}_t, \cdots, \widehat{g_t^{k+2}})|^{(1+\delta)} &= E^{1/(1+\delta)} |\lambda^{-k} \hat{g}_t^{\{k+1\}\nu}|^{(1+\delta)} \\ &\leqq \sum_{i=1}^{k+1} E^{1/(1+\delta)} \left[ \left| \frac{\partial G_k}{\partial \widehat{g_t^i}} \right|^{(1+\delta)} |\widehat{g_t^{i\nu}}|^{(1+\delta)} \right]. \end{aligned}$$

Applying now the Hölder inequality (e.g., Loève [1963, Ch. 3]), with exponents $2/(1-\delta)$ and $2/(1+\delta)$, to each summand in the above sum, we obtain

$$(\alpha) \qquad f_t \leqq \sum_{i=1}^{k+1} E^{(1-\delta)/2(1+\delta)} \left| \frac{\partial G_k}{\partial \widehat{g_t^i}} \right|^{2(1+\delta)/(1-\delta)} E^{1/2} \{\widehat{g_t^{i\nu}}\}^2.$$

Now, $\partial G_k / \partial \widehat{g_t^i}$ is a polynomial in $\hat{g}_t, \cdots, \widehat{g_t^{k+1}}$ and, by our assumptions, $E\{g(x_t)\}^{2n} < \infty$ for all integers $n \geqq 0$ and for every $t \in [0, T]$. Also, all relevant quantities are almost surely continuous functions of $t$ on the compact interval $[0, T]$ (and we also have $1 - \delta > 0$). It follows then that there exists a constant $M' \geqq 0$ such that, for every $i = 1, \cdots, k+1$ and for all $t \in [0, T]$,

$$(\beta) \qquad E^{(1-\delta)/2(1+\delta)} \left| \frac{\partial G_k}{\partial \widehat{g_t^i}} \right|^{2(1+\delta)/(1-\delta)} \leqq M'.$$

Further, by (1.4), we have

$$\int_0^t \widehat{g_s^i}\, d\nu_s = \widehat{g_t^i} - \int_0^t \Big[ \overline{im(x_s, s) g^{i-1}(x_s) g'(x_s)} + \left(\frac{i(i-1)}{2}\right) \overline{\sigma^2(x_s, s) g^{i-2}(x_s) g'(x_s)} + \left(\frac{i}{2}\right) \overline{\sigma^2(x_s, s) g^{i-1}(x_s) g''(x_s)} \Big]\, ds$$

for all integers $i \geqq 1$. We now square both sides of the last equation, and consider the expectation of the resulting quantities. On the left-hand side we obtain $\int_0^t E\{\widehat{g_s^{i\nu}}\}^2\, ds$. Also, by an application of the Hölder inequality and in view of our assumptions on (1.1) and (1.3), it follows that the expectation of the squared right-hand side is bounded by a constant $M'' \geqq 0$. Thus, we have, for all $i = 1, \cdots, k+1$, that $\int_0^T E\{\widehat{g_s^{i\nu}}\}^2\, ds \leqq M''$.

But then, for every $\alpha > 0$, $\lim_{\lambda\to\infty} \int_0^T \lambda^{-\alpha} E\{\widehat{g_s^{i\nu}}\}^2\, ds = 0$, so that

$$(\gamma) \qquad \lim_{\lambda\to\infty} \lambda^{-\alpha} E\{\widehat{g_t^{i\nu}}\}^2 = 0$$

for almost all $t \in [0, T]$. Finally, substituting $(\beta)$ and $(\gamma)$ into $(\alpha)$, it follows that, for every $\alpha > 0$,

$$\lim_{\lambda\to\infty} E^{1/(1+\delta)} |\lambda^{1-\alpha} G_{k+1}(\hat{g}_t, \cdots, \widehat{g_t^{k+2}})^{1+\delta}| \leqq \lim_{\lambda\to\infty} \left[ \sum_{i=1}^{k+1} M'(\lambda^{-\alpha} E^{1/2}\{\widehat{g_t^{i\nu}}\}^2) \right] = 0$$

for almost every $t \in [0, T]$, proving our assertion. □

Consider now Theorem 4.3 in the case of linear measurement, that is, when $g(x) = x$. In this case, substituting $n = 1$ and $\delta = 0$, and noting that $G_1 = \widehat{x_t^2} - \hat{x}_t^2$, we obtain that $E\{(x_t - \hat{x}_t)^2\} \sim 1/\lambda$, which is in accordance with the upper bound of Zakai and Ziv [1972]. Thus, Theorem 4.3 is a generalization of that result.

We conclude this section by showing that the conditional probability measure $\hat{P}_t$ can be replaced by a Gaussian measure, up to an error of the "order" of $1/\lambda$. To this end, we let $\Pi_t$, for every $t \in [0, T]$, be the Gaussian measure determined by its first two moments as follows: $\int x\, d\Pi_t(x) = \hat{g}_t$ and $\int x^2\, d\Pi_t(x) = \widehat{g_t^2}$. Given a function $f: R \to R$, we shall denote by $\overline{f_t} := \int f(x)\, d\Pi_t(x)$ its expectation with respect to $\Pi_t$. We now have the following:

THEOREM 4.4. *Given the nonlinear filtering problem of the information process* (1.1) *with measurement process* (1.3), *the following holds for every* $0 \leqq \delta < 1$:

$$E^{1/(1+\delta)} |\widehat{g_t^i} - \overline{g_t^i}|^{(1+\delta)} \sim \frac{1}{\lambda},$$

*where* $i = 0, 1, 2, \cdots$, *and* $t \in [0, T]$.

*Proof.* The cases $i = 0$, 1, and 2 are clearly implied by the construction of the probability measure $\Pi_t$. The proof proceeds by induction. Assume that the theorem holds for all integers $i = 1, \cdots, n$. Now by Lemma 4.1, $G_n(\overline{g_t}, \cdots, \overline{g_t^{n+1}}) = 0$ for all integers $n \geqq 2$, so that, since $G_n$ is monic in $\overline{g_t^{n+1}}$, we have that $\widehat{g_t^{n+1}} - \overline{g_t^{n+1}} = G_n(\overline{g_t}, \overline{g_t^2}, \cdots, \overline{g_t^n}, \widehat{g_t^{n+1}})$ for all integers $n \geqq 2$. The following calculation is intended to replace the arguments $\overline{g_t^n}, \cdots, \overline{g_t}$ in the last expression by $\widehat{g_t^n}, \cdots, \hat{g}_t$, and to compute the error caused by this manipulation. To this end, we represent $G_n(\overline{g_t}, \cdots, \overline{g_t^n}, \widehat{g_t^{n+1}}) = A_t + \overline{g_t^n} B_t$, where $A_t$ and $B_t$ are suitable polynomials in $\widehat{g_t^{n+1}}, \overline{g_t^n}, \cdots, \overline{g_t}$. By our assumptions on (1.1) and (1.3), it follows that $A_t$ and $B_t$ have all their moments bounded.

Now, let $\delta'$ be such that $\delta<\delta'<1$. Then, using the Minkowski and Hölder inequalities, we obtain:

$$E^{1/(1+\delta)}|\widehat{g_t^{n+1}}-\widehat{g_t^{n+1}}|^{(1+\delta)}=E^{1/(1+\delta)}|A_t+\widehat{g_t^n}B_t+(\widehat{g_t^n}-\widehat{g_t^n})B_t|^{(1+\delta)}$$
$$\leqq E^{1/(1+\delta)}|A_t+\widehat{g_t^n}B_t|^{(1+\delta)}+E^{1/(1+\delta')}|\widehat{g_t^n}-\widehat{g_t^n}|^{(1+\delta')}E^{1/\gamma}|B_t|^{\gamma},$$

where $\gamma:=(1+\delta)(1+\delta')/(\delta'-\delta)$. Applying now the induction assumption, and the fact that all moments of $B_t$ are bounded, it follows that

$$E^{1/(1+\delta)}|\widehat{g_t^{n+1}}-\widehat{g_t^{n+1}}|^{(1+\delta)}\leqq E^{1/(1+\delta)}|A_t+\widehat{g_t^n}B_t|^{(1+\delta)}+f_t,$$

where $f_t\sim 1/\lambda$.

By a similar procedure, we replace, for all $i=1,\cdots,n$, all appearances of $\widehat{g_t^i}$ by $\widehat{g_t^i}$, retaining the corresponding errors $f_t$. Thus, after a finite number of steps, we obtain

$$E^{1/(1+\delta)}|\widehat{g_t^{n+1}}-\widehat{g_t^{n+1}}|^{(1+\delta)}\leqq E^{1/(1+\delta)}|G_n(\widehat{g_t^{n+1}},\widehat{g_t^n},\cdots,\hat{g}_t)|^{(1+\delta)}+h_t,$$

where $h_t\sim 1/\lambda$. But then, it follows by Theorem 4.3, that $E^{1/(1+\delta)}|\widehat{g_t^{n+1}}-\widehat{g_t^{n+1}}|^{(1+\delta)}\sim 1/\lambda$, concluding our proof. □

Finally, we note that Theorem 4.4 can be directly extended to the case of polynomials in $g(x)$ and, also, to functions which are limits, in a suitable sense, of such polynomials.

**Acknowledgment.** The use of the quadratic variation in the proofs of the present paper was suggested by Moshe Zakai, Department of Electrical Engineering, Technion-Israel Institute of Technology, Haifa, Israel. The author is most grateful to him for this suggestion, and for many valuable and stimulating discussions on mentioned and related topics.

## REFERENCES

M. Fujisaki, G. Kallianpur and H. Kunita [1972], *Stochastic differential equations for the nonlinear filtering problem*, Osaka J. Math., 9, pp. 19–40.

I. I. Gikhman and A. V. Skorokhod [1972], *Stochastic Differential Equations*, Springer-Verlag, Berlin.

A. H. Jazwinski [1970], *Stochastic Processes and Filtering Theory*, Academic Press, New York.

T. Kailath [1969], *A general likelihood formula for random signals in Gaussian noise*, Trans. IEEE, IT-15, pp. 350–361.

H. J. Kushner [1967], *Dynamical equations for optimal nonlinear filtering*, J. Differential Equations, 2, pp. 179–190.

M. Loève [1963], *Probability Theory*, Van Nostrand, Princeton NJ.

P. A. Meyer [1966], *Probabilité et potential*, Hermann, Paris.

———, [1975], *Un course sur les intégrales stochastiques*, Lecture notes in Mathematics, Springer-Verlag, Berlin.

R. I. Stratonovich [1960], *Conditional Markov processes*, Theory Prob. Appl., 5, pp. 156–178.

M. Zakai [1969], *On the optimal filtering of diffusion processes*, Z. Wahrsch Verw. Geb., 11, pp. 230–243.

———, [1975], *Lecture notes on nonlinear filtering*, Technion-Israel Institute of Technology, Haifa, Israel. (In Hebrew).

M. Zakai and J. Ziv [1972], *Lower and upper bounds on the optimal filtering error of certain diffusion processes*, Trans. IEEE, IT-18, pp. 325–331.