THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 50, No. 4, April, 1971
Printed in U.S.A.

Some Extensions of the Innovations Theorem*

By THOMAS KAILATH†

(Manuscript received December 4, 1970)

Consider an observed process that is the sum of a Wiener noise *process and the integral of a not necessarily gaussian* signal *process. The innovations process is defined as the difference between the observed process and the integral of the causal minimum-mean-square-error estimate of the signal process. Then if the integral of the expected value of the absolute magnitude of the signal process is finite, we show that the innovations process is also a Wiener process. The present conditions are a substantial weakening of those previously used in which the integral of the signal variance had to be finite. The new result is obtained by using some recent results in martingale theory. These results also enable us to obtain similar results when the Wiener process is replaced by a square-integrable martingale.*

I. INTRODUCTION AND DISCUSSION OF RESULTS

We shall be concerned with stochastic processes of the form

$$dy = z\,dt + dw, \quad \text{i.e.,} \quad y(t) = \int_0^t z(s)\,ds + w(t), \qquad 0 \leqq t \leqq T \tag{1}$$

where $w(\cdot)$ is a Wiener process with

$$E[w(t)] = 0, \;\; E[w(t)w(s)] = \min(t, s) \qquad (2)‡$$

and $z(\cdot)$ is a not necessarily gaussian process such that, for every $0 \leqq s < t \leqq T$

$$w(t) - w(s) \perp\!\!\!\perp \mathfrak{F}_s = \sigma\{y(\tau),\; 0 \leqq \tau \leqq s\} \tag{3}$$

* This work was supported by the Applied Mathematics Division of the Air Force Office of Scientific Research under contract AF 44-620-69-C-0101 and in part by Bell Telephone Laboratories (Summer 1969) and a Guggenheim Fellowship.

† Stanford University, Stanford, California 94305.

‡ Later we shall replace $w(\cdot)$ by a square-integrable martingale (cf., Theorem 3).

where $\sigma\{\cdot\}$ denotes the smallest sigma-field generated by $\{\cdot\}$ and, following K. Itô, the symbol $\perp\!\!\!\perp$ will be used to denote statistical independence. In our earlier work,[1] we also assumed that

$$E \int_0^T |z(t)|^2 \, dt < \infty. \tag{4}$$

This assumption not only enabled us to define the conditional expectation,

$$E[z(t) \mid \mathfrak{F}_t] = \hat{z}(t), \quad \text{say} \tag{5}$$

but also to conclude that $|\hat{z}(\cdot)|$ and $|\hat{z}(\cdot)|^2$ (as well as $|z(\cdot)|$ and $|z(\cdot)|^2$) have finite integrals a.s. Then in Appendix I of Ref. 1 and independently in Refs. 2 through 4, the following result was proved:

Theorem 1: *Under the assumptions* (1) *through* (4), *the 'innovations' process* $\{\nu(t), \mathfrak{F}_t\}$ *where*

$$\nu(t) = y(t) - \int_0^t \hat{z}(s) \, ds, \qquad 0 \leqq t \leqq T \tag{6}$$

is a Wiener process, relative to the sigma-fields $\mathfrak{F}_t$ *generated by* $y(\cdot)$. *Moreover,* $\nu(\cdot)$ *has the same statistics as the original Wiener process* $w(\cdot)$.

The reasons for the name "innovations" and some applications of the concept are discussed in Refs. 1 through 5, and the references therein, so that we shall not pursue them here. Our interest will be in relaxing the conditions under which the theorem is true. In particular we shall show that the condition (4) can be replaced by the weaker condition

$$E \int_0^T |z(t)| \, dt < \infty. \tag{7}$$

This is a substantial relaxation of (4), since now we can even consider cases in which $z(\cdot)$ need not have finite variance and need not be square-integrable in t a.s., (examples of this type are easy to find for gaussian $z(\cdot)$, cf., Ref. 6).

Nevertheless, even this weakening is not the best possible. Such a claim seems surprising, since to begin with, the standard definitions (see e.g., J. L. Doob[7]) of conditional expectation require that

$$E|z(t)| < \infty.$$

In our problem, however, L. Shepp has pointed out by example that even this condition on $z(\cdot)$ is not always necessary.[8] In Shepp's example,

$z(t) = \eta,$ a Cauchy distributed random variable, independent of $w(\cdot)$.

Then it is easy to verify that $z(\cdot)$ as defined below by equation (8) can be used to define the innovations process $\nu(\cdot)$, even though $E \mid z(t) \mid$ does not exist. Nevertheless, Bayes' theorem shows that

$$\hat{z}(t) = \frac{\int_0^t \eta \exp\left[\eta w(t) - \frac{\eta^2 t}{2}\right] p(\eta)\, d\eta}{\int_0^t \exp\left[\eta w(t) - \frac{\eta^2 t}{2}\right] p(\eta)\, d\eta}. \tag{8}$$

It is true that there is a generalized definition of conditional expectation (see, e.g., p. 342 of Ref. 9) that can be used when $E \mid z \mid$ does not exist, but the exact sense in which equation (8) fits this generalized definition and the more general problem of trying to define $\hat{z}(\cdot)$ under just the assumptions (1) through (3) requires further investigation. So much for the assumption (4). To see that equation (3) is also not always necessary, we can let $z(t) = w(T)$, $0 \leqq t \leqq T$. By using the results of Ref. 6, it can be shown that even in this case we can define a variable $\hat{z}(\cdot)$ such that the process $\nu(\cdot)$ of equation (6) is still a Wiener process. Having discussed such examples however, let us quote precisely the results that we can prove. The first is:

Theorem 2: *Under the assumptions* (1)-(3) *and* (7), *the innovations process* $\{\nu(t), \mathfrak{F}_t\}$ *defined by equation* (6), *is a Wiener process with the same statistics as* $w(\cdot)$.

Besides extending the range of applicability of the innovations process, Theorem 2 is of interest, at least to us, because of the relatively new techniques used to prove it. Thus Theorem 1 was proved in Ref. 1 by using a martingale theorem of Lévy and Doob (cf., p. 384 of Ref. 7) to the effect that a continuous-path, finite-variance martingale with conditional variance equal to t must be a Wiener process. The assumption (4) that was made in Theorem 1 enabled us, with some computation,[1] to establish the finite-variance and conditional variance properties. In Theorem 2, the process $z(\cdot)$ may itself have infinite variance and we cannot easily apply the theorem of Lévy and Dobb. Fortunately, however, some recent developments in martingale theory, due to H. Kunita and S. Watanabe[10] and to P. Meyer,[11,12] have provided a generalized form of the Lévy-Doob theorem that we can use to prove Theorem 2, and in fact with less computation than was needed in Ref. 1 for Theorem 1. Moreover, the results in Refs. 10 through 12 have suggested our second result, which in fact includes the first one.

Theorem 3: Let $dx = z\,dt + dM$, where

$$E \int_0^T |\, z(t) \,|\, dt < \infty, M(t) - M(s) \perp\!\!\!\perp \mathfrak{F}_s \tag{9}$$

and $M(\cdot)$ is a square-integrable martingale. Then $\{\mu(t, \omega), \mathfrak{F}_t, P\}$, where

$$\mu(t) = x(t) - \int_0^t \hat{z}(s)\, ds, \qquad \hat{z}(t) = E[z(t) \mid \mathfrak{F}_t],$$

is a locally square-integrable martingale with quadratic variation the same as that of M.

P. Frost has obtained the result of Theorem 3 under the stronger hypotheses that $E|z(\cdot)\,|^2$ is integrable and $M(\cdot)$ is continuous.[13]

A special form of Theorem 3 was obtained earlier[5] by use of a result of Doob (See p. 449 of Ref. 7) that a continuous square-integrable martingale $\{M(t), \mathcal{B}_t\}$ satisfying

$$E\{[M(t) - M(s)]^2 \mid \mathcal{B}_s\} = E\left\{\int_0^t |\, G(\tau, \omega)\,|^2\, d\tau \mid \mathcal{B}_s\right\}$$

can be written as a stochastic integral with respect to a Wiener process. In other words, this leads us to extend equation (1) by allowing a stochastic coefficient for dw; we may note that Kunita and Watanabe have shown that a (locally) square-integrable martingale of a Wiener process can also be written as a stochastic integral.

Finally we may remark that the proof of Theorem 3 also yields the following.

Corollary 4: Theorem 3 remains valid if hypothesis (7) is replaced by

$$\hat{z}(\cdot) \text{ exists and } \int_0^T \hat{z}(t)\, dt < \infty \text{ a.s.} \tag{10}$$

Note that condition (7) guarantees (10), not just for conditional expectations with respect to the sigma-fields $\{\mathfrak{F}_t\}$ generated by $dx = z\,dt + dw$, but with respect to any family of sigma-fields $\{A_t\}$; clearly the condition (7) is too strong and does not, as we also noted earlier, exploit the special nature of our problem. We have as yet found no condition on $z(\cdot)$ simpler than (7) that will imply (10). It seems that what will be needed is a condition on $x(\cdot)$ rather than on $z(\cdot)$ and $M(\cdot)$ separately. In fact, we may observe that in all our examples [in the lines below (7)], the process $x(\cdot)$ is absolutely continuous with respect to a Wiener process; however, we have not yet worked out all the implications of absolute continuity.

II. PROOFS OF THE THEOREMS

We shall begin with some definitions; general references for the following material are the paper of Kunita and Watanabe,[10] and the book[11] and lecture notes[12] of Meyer.

All our random variables will be defined on a probability space $(\Omega, \mathcal{B}, P)$. We shall also assume that we have an increasing and right-continuous family, $\mathcal{B}_t$, $0 \leqq t \leqq T$, of subsigma-fields of $\mathcal{B}$, each $\mathcal{B}_t$ containing all the null sets; we shall write $\mathcal{B}_T = \mathcal{B}$.

A process $\{M(t, \omega)\}$ will be called an L_2-*martingale* or a *square-integrable martingale* if $M(\cdot)$ has a.s. right-continuous sample paths* and if

$$EM^2(t, \omega) < \infty \quad \text{and} \quad E[M(t, \omega) \mid \mathcal{B}_s] = M(s, \omega) \quad \text{a.s.}$$

Every L_2-martingale $\{M(t, \omega)\}$ has associated with it an increasing function, called its *quadratic variation* $[M(t), M(t)]$, that can be calculated as (cf., p. 92 of Ref. 12)

$$[M(t), M(t)] = \lim_{\max |t_{i+1} - t_i| \to 0} \sum_{0}^{n-1} [M(t_{i+1}) - M(t_i)]^2 \tag{11}^\dagger$$

where $(t_0, t_1, \cdots, t_n)$ is any partition of $(0, t)$ and the convergence of the sum is in L^1, i.e., in the norm $||A||^2 = E|A|$. When $M(\cdot)$ has continuous sample-paths, the quadratic variation is often written $\langle M, M \rangle$, and can be identified as the unique increasing process such that process $M^2(t) - \langle M(t), M(t) \rangle$ is a martingale.

A positive random variable $\tau(\omega)$ is called a *stopping time* if for every $t > 0$, the event $\{\tau(\omega) \leqq t\}$ belongs to $\mathcal{B}_t$, i.e., it depends only upon the history up to t.

A example of a stopping time is the first time $\langle M(t) \rangle$ exceeds a given level. An important property of a stopping time is that if $M(t)$ is a martingale process, then the "stopped" process $M(t \wedge \tau)$ is also a martingale.‡

A process $M(t, \omega)$, $0 \leqq t \leqq T$, is said to be a (continuous) local L_2-martingale, or a locally square-integrable martingale, if there exists a sequence of stopping times $\tau_1 \leqq \tau_2 \leqq \cdots$, increasing to T such that the stopped processes are all (continuous) L_2-martingales. The unique

* Meyer (p. 45 of Ref. 11) has shown that if $E[M(t, \omega)]$ is right-continuous (and if, as assumed above, $\{\mathcal{B}_t\}$ is right-continuous) then there always exists a right-continuous modification of $M(\cdot)$. If the paths of $M(\cdot)$ are a.s. continuous, we shall call the martingale continuous.

† This notation is a shorthand (cf. Meyer[12]) for the usual procedure involving a sequence of partitions.

‡ We use the notation $a \wedge b = \min(a, b)$.

(continuous) increasing function $\lim_{n\to\infty} [M(t\Lambda\tau_n)]$ will be called the function associated with the local L_2-martingale M, or often, the quadratic variation of M.

2.1 *Example*

The Itô integral

$$I(t) = \int_0^t f(s, \omega)\, dw\,(s, \omega) \tag{12a}$$

can be defined as a continuous function of t, $0 \leqq t \leqq T$, for every $f(\cdot\,,\,\cdot)$ such that $f(\cdot\,,\,\cdot)$ is independent of future increments of $w(\cdot\,,\,\cdot)$ and

$$\int_0^T f^2(s, \omega)\, ds < \infty \quad \text{a.s.} \tag{12b}$$

If *in addition* $f(\cdot)$ obeys

$$E\left[\int_0^T |\, f(s, \omega)\, |^2\, ds\right] < \infty \tag{13}$$

then the Itô integral $I(t)$ is an L_2-martingale and its quadratic variation is

$$[I(t), I(t)] = \langle I(t), I(t)\rangle = \int_0^t f^2(s, \omega)\, ds.$$

By defining

$$\tau_n = \begin{cases} \min\left\{t\colon \int_0^t f^2(s, \omega)\, ds \geqq n\right\}, \\ T, \text{ if } \{\cdot\} \text{ is empty,} \end{cases}$$

we see that the Itô integral under equation (12) is always a continuous local L_2-martingale, with quadratic variation

$$[M(t), M(t)] = \lim_{n\to\infty} \int_0^{t\Lambda\tau_n} f^2(s, \omega)\, ds = \int_0^t f^2(s, \omega)\, ds.$$

2.2 *Proof of Theorem 3*

We note first that, with $\tilde{z}(t) = z(t) - \hat{z}(t)$,

$$\mu(t) = y(t) - \int_0^t \hat{z}(s)\, ds = \int_0^t \tilde{z}(s)\, ds + M(t)$$

is a martingale relative to the fields $\{\mathfrak{F}_t\}$ because for $\tau < t$,

$$E[\nu(t) \mid \mathfrak{F}_\tau] = \nu(\tau) + E\left[\int_0^t [\tilde{z}(s)\, ds + dM(s)] \mid \mathfrak{F}_\tau\right],$$

$$= \nu(\tau) + \int_\tau^t E[\tilde{z}(s) \mid \mathfrak{F}_s \mid \mathfrak{F}_\tau]\, ds + E[M(t) - M(\tau) \mid \mathfrak{F}_\tau],$$

$$= \nu(\tau) + \int 0\, ds + 0 = \nu(\tau). \tag{14}$$

The interchange of $E[\cdot]$ and the integral is justified by Fubini's theorem and the assumption (7).

Now define a sequence of $\mathfrak{F}_t$-measurable stopping times by

$$T_n = \{\inf t \, : \, t \geqq 0, \;\; \mu(t) \geqq n\} \Lambda T. \tag{15}$$

Since $\mu(t) < \infty$ a.s., the $\{T_n\}$ increase to T. Also it is clear that, for each n, $\mu(t \Lambda T_n)$ is a square-integrable martingale. Therefore, $\mu(\cdot)$ is a locally square-integrable martingale. To find its quadratic variation, we first show that

$$[\mu(t\Lambda T_n), \;\; \mu(t\Lambda T_n)] \;\; = \;\; [M(t\Lambda T_n), \;\; M(t\Lambda T_n)]. \tag{16}$$

To obtain equation (16), we use the definition (12), the fact that $[\cdot,\cdot]$ is zero for a continuous process of bounded variation, and the inequality

$$\sum_i [\mu(t_{i+1}) - \mu(t_i)]^2 \leq \sum \left[\int_{t_i}^{t_{i+1}} z(s)\right]^2 + \sum [M(t_{i+1}) - M(t_i)]^2.$$

Finally by letting n tend to infinity, we see that

$$[\mu(t), \;\; \mu(t)] \;\; = \;\; [M(t), \;\; M(t)], \;\; 0 \leqq t \leqq T.$$

This completes the proof of Theorem 3.

2.3 *Proof of Theorem 2*

When $M(t)$ is a Wiener process, its quadratic variation is t. To obtain Theorem 2 from Theorem 3, we now use a theorem of Lévy and Doob (See p. 384 of Ref. 7), as extended by Kunita and Watanabe (See p. 217 of Ref. 10), that a continuous locally square-integrable martingale with quadratic variation t must be a Wiener process.

It is interesting to point out that the Kunita-Watanabe proof of the extended result is also considerably simpler than the original proof because the new proof utilizes the powerful new tools of the stochastic integral for L_2-martingales and the Itô differential rule for such processes.

2.4 *Proof of Corollary 4*

We note that hypothesis (10) suffices to ensure that $\mu(\cdot)$ is a.s. finite, so that the $\{T_n\}$ of equation (15) are well defined and tend to T, and also to ensure that $\mu(t \wedge T_n)$ is a square-integrable martingale. These are the essential ingredients of the proof of Theorem 3.

The present wider conditions under which we have been able to establish the innovations result will, of course, extend the range of problems in which the innovations can be used. In particular, we have applied Theorem 2 to show[14] that a general likelihood-ratio formula, derived in Ref. 5, for processes obeying equations (1) through (4) remains valid if equation (4) is replaced by the weaker conditions that $\int z^2(t)\, dt < \infty$ a.s. and $\int E\, |\, z(t)\, |\, dt < \infty$.

REFERENCES

1. Kailath, T., "A General Likelihood-Ratio Formula for Random Signals in Gaussian Noise," *IEEE Trans. on Infm. Theory, IT-15*, No. 3, (May 1969), pp. 350–361.
2. Frost, P. A., "Estimation in Continuous-Time Nonlinear Systems," Ph.D. Dissertation, Dept. of Electrical Engineering, Stanford University, Stanford, Calif., June 1968.
3. Shiryaev, A. N., "Stochastic Equations of Nonlinear Filtration for Purely Discontinuous Markov Processes," *Prob. Pered. Informatsii, 2*, (1966), pp. 3–22.
4. Kallianpur, G., personal communication, August 1969.
5. Kailath, T., "A Further Note on a General Likelihood-Ratio Formula," *IEEE Trans. on Infm. Theory, IT-16*, No. 4, (July 1970), pp. 393–396.
6. Kailath, T., "Likelihood Ratios for Gaussian Processes," *IEEE Trans. on Infm. Theory, IT-16*, No. 3, (May 1970), pp. 276–288.
7. Doob, J. L., *Stochastic Processes*, New York: J. Wiley & Sons, Inc., 1953.
8. Shepp, L., personal communication.
9. Loève, M., *Probability Theory*, New Jersey: Van Nostrand Book Co., Third Edition, 1963.
10. Kunita, H. and S. Watanabe, "On Square Integrable Martingales," *Nagoya Math. J., 30*, August 1967, pp. 209–245.
11. Meyer, P., *Probability and Potentials*, Waltham, Mass.: Blaisdell Book Co., 1966.
12. Meyer, P., *Intégrales Stochastiques*, I–IV, Lecture Notes in Mathematics *39*, Springer, Berlin, 1967, pp. 72–162.
13. Frost, P. A., "The Innovations Process and Its Application to Non-Linear Estimation and Detection of Signals in Additive White Noise," *Proc. UMR-Mervin J. Kelly Communications Conference*, Rolla, Missouri, October 1970, pp. 7-3-1 to 7-3-6.
14. Kailath, T., "The Structure of Radon-Nikodym Derivatives with Respect to Wiener and Related Measures," *Annals Math. Stat.*, to appear, 1971.