JB 82-04

Paper Entitled

"Accurate Evaluation of Conditional Densities in Nonlinear Filtering"

From the Proceedings of

Filtering and Stochastic Control
Springer-Verlag Lecture Notes
in Control and Information Sciences
Vol. 42, pp. 54-68

New York
1982

JOHN BARAS

ACCURATE EVALUATION OF CONDITIONAL DENSITIES IN NONLINEAR FILTERING

W.E. Hopkins, Jr., G.L. Blankenship and J.S. Baras
Electrical Engineering Department
University of Maryland
College Park, Maryland 20742

ABSTRACT

Using some approximation formulas for stochastic Wiener function space integrals, it is possible to approximate the conditional densities which arise in the nonlinear filtering of diffusion processes to within $O(n^{-2})$, with $n \geq 1$ arbitrary, by n-fold ordinary integrals. The latter have the simple form of a "rectangular rule", but their accuracy is an order of magnitude better. The n-fold integral can be further decomposed into a recursion involving n one dimensional integrals. The sequence is recursive in the increments of the observation process in the filtering problem. It is not, however, recursive in time. The one dimensional integrals are naturally treated by an m-step Gaussian quadrature which has an error proportional to $nm!/[2^m(2m)!]$. (The proportionality constant can be estimated and optimized.) The computation of these individual integrals can be reduced further by exploiting certain inherent symmetries of the problem, and by doing some preliminary, "off-line" computing. The end result is a highly accurate, computationally efficient numerical algorithm for evaluating conditional densities for a substantial class of nonlinear filtering problems. By accepting slight reductions in accuracy, one can obtain an algorithm (apparently) fast enough, when efficiently coded, for "on-line," recursive filtering in real time.

1. THE PROBLEM

Consider the stochastic dynamical system

$$dx(t) = f(x(t))dt + g(x(t))dv(t) \tag{1.1a}$$

$$dy(t) = h(x(t))dt + dw(t) \tag{1.1b}$$

$$x(0) = x_0 \sim p_0(x), \quad 0 \le t \le T$$

written in the Itô calculus. Here f,g,h are smooth functions mapping R into R; v(t), w(t) are independent, R-valued standard Wiener processes mutually independent of the initial condition x_0 which has a smooth, bounded density $p_0(x)$. We call x the signal and y the observation process. The filtering problem is to determine an estimate of x(t) given observations y(s), $0 \le s \le t$, or, more precisely, the σ-algebra Y_t generated by $\{y(s), 0 \le s \le t\}$. For the estimate one usually takes the conditional mean $\hat{x}(t) = E[x(t)|Y_t]$ since this produces the minimum mean square error.

To compute $\hat{x}(t)$, it suffices to know $p(t,x|Y_t)$, the conditional density of x(t) given Y_t, if it exists. This satisfies a complex, nonlinear stochastic partial differential equation which is difficult to analyze or to treat numerically [1] [2]. Alternately, one can write

$$p(t,x|Y_t) = u(t,x)/[\int_{-\infty}^{\infty} u(t,z)dz] \tag{1.2}$$

where the unnormalized conditional density u satisfies

$$u(t,x) = [\tfrac{1}{2}\, \partial_{xx}(g^2 u) - \partial_x(fu) - \tfrac{1}{2}h^2\, u]dt + h(x)u\, dy(t) \tag{1.3}$$

$$u(0,x) = p_0(x), \quad 0 \le t \le T$$

(written in the Stratonovich calculus*) a linear stochastic PDE discovered by M. Zakai [3] (and independently by R. Mortensen and T. Duncan - see the remarks in [3]). The solution to this equation may be written in terms of a function space integral. Specifically,

$$u(t,x) = E_x\{\exp[\int_0^t h(x(s))dy(s) - \tfrac{1}{2}\int_0^t h^2(x(s))ds]\, p_0(x(t))\} \tag{1.4}$$

where $E_x\{\cdot\}$ is expectation over the paths of (1.1) starting at x(0)=x. That is, (1.4) is formally the solution of (1.3) as a short calculation using the Stratonovich calculus shows. (Existence and uniqueness of solutions to (1.3) are discussed in [4], among other papers.)
Our objective here is the evaluation of the function space integral (1.4) by quadrature type approximations.

The first step of the approximation is carried out by regarding (a transformed version of) (1.4) as a "stochastic Wiener integral" and applying the formulas in [5]. This leads to an approximation of the function space integral by an n-fold ordinary integral with an error $O(n^{-2})$. The second step of the approximation is reduction of the n-fold integral to a recursive sequence of one dimensional integrals which may be evaluated by Gaussian quadrature.

To describe the formulas in [5], we will use the following setup: Let $C_0([0,t])$ be the space of R-valued, continuous functions x(s) on [0,t] with x(0)=0 and let W be Wiener measure on C_0. If $F:C_0 \to R$ is a smooth functional, the Wiener integral

$$I = \int_{C_0} F(x)dW(x) \tag{1.5}$$

is defined as the sequential limit [6]

$$I = \lim_{\substack{\max|t_j - t_{j-1}| \to 0 \\ 1 \le j \le n \\ n \to \infty}} \int_R \overset{n}{\cdots} \int_R da_1 \dots da_n F(z_{sx}) \cdot \prod_{j=1}^{n} \frac{\exp[-(a_j - a_{j-1})^2/2(t_j - t_{j-1})]}{[2\pi(t_j - t_{j-1})]^{n-2}} \tag{1.6}$$

where $0<t_1<t_2<\dots<t_n = t$ and z_{sx} is a polynomial function on [0,t] passing through x at s=0 and a_j at t_j, j=1,2,...,n.

Except for a few simple cases, it is impossible to evaluate Wiener integrals explicitly. Approximation formulas suitable for numerical computations are, therefore, of considerable interest in applications. In [7] A.J Chorin presented some formulas of this type. His results were based on the use of parabolas to interpolate the Wiener paths and on expansions of the nonlinear functional F in a Taylor series with the quadrature formula adjusted to optimize the approximation of the first two terms. Chorin's formulas were of the form

$$\int_C F(x)dW(x) = \pi^{-n/2} \int_{R^n} F_n(u_1,\dots,u_n)e^{-u_1^2 - \dots - u_n^2} du_1 \dots du_n + O(n^{-2}) \tag{1.7}$$

where F_n had the simple form of a rectangle rule in specific cases.

The function space integral in (1.4) involves a random (Itô) process $\{y(s), 0 \le s \le t\}$, and the formulas of Chorin do not apply to it. In [5] these formulas were adapted and extended to cover this case. The result

Lemma. Let $\{w(s),\ 0\le s\le t\}$ be an R-valued standard Wiener process on $(\Omega,\mathcal{F},P)$ and let $\{f(s),\ g(s),\ 0\le s\le t\}$ be R-valued random processes non-anticipative with respect to w which have continuous paths almost surely and second moments uniformly bounded in $s\in[0,t]$. Let

$$dy(s) = f(s)ds + g(s)dw(s) \qquad (1.8)$$
$$y(0) = 0,\ 0\le s\le t.$$

Suppose $V:R\to R$ has derivatives up to order 4 with

$$\int_t^{t+1/n}|V''''(z(s))|^2 ds = O(n^{-1}) \qquad (1.9)$$

for all $t\in[0,T]$ and any continuous $z:[0,T]\to R$. Then for any $t\in[0,T]$

$$I = \int_C \exp[\int_0^t V(x(s))dy(s)]dW(x)$$

$$= (2\pi)^{-n/2}\int_{R^n}\{\exp[\sum_{i=1}^n V(x_{i-1}+\frac{u_n t}{\sqrt{2n}})\Delta y_{i-1}]\} \qquad (1.10)$$

$$\cdot \exp[-\tfrac{1}{2}(u_1^2+\ldots+u_n^2)]du_1\ldots du_n + e_n$$

where

$$x_i = t(u_1+\ldots+u_i)/n,\ i = 1,2,\ldots,n \qquad (1.11)$$

$$t_i = it/n,\ \Delta y_{i-1} = y(t_i)-y(t_{i-1})$$

The approximation error is

$$(Ee_n^2)^{1/2} = O(n^{-2}) \qquad (1.12)$$

where E is expectation with respect to the distribution of $\{y(s),0\le s\}$.

Remarks 1. The "ordinary" integral which appears in (1.4) admits a similar approximation with $\Delta t = n^{-1}$ replacing Δy_{i-1} in (1.10). The error, which is deterministic, is also $O(n^{-2})$.

2. The remarkable feature of formula (1.10), as noted in [7], is that it is no more complicated in structure nor does it require more computing effort than the standard "rectangle rule" [8] which has accuracy $O(n^{-1})$.

3. The evaluation of the n-fold integral in (1.10) may be reduced to a sequence of one dimensional integrals which are recursive in the increments Δy_{i-1}. This has some important implications in the filtering problem.

4. The simple form of (1.10), the error estimate, and the recursive evaluation depend on the fact that the underlying measure

process, it is necessary to make a change of coordinates or a change of measure (i.e., a Girsanov transformation) in (1.1) to take advantage of this structure.

2. COORDINATE TRANSFORMATIONS

If the coefficients in (1.1a) are sufficiently smooth, we can change coordinates (pathwise) so that the resulting diffusion is a Wiener process. When the coefficients are not smooth or when $x(t)$ is a multi-variable process, this procedure may not work, and a change of measure, i.e., a Girsanov transformation, may be required to implement our computational algorithms.

Suppose

(A1) $g(x) \geq g_0 > 0$ for some g_0 and all $x \in R$

$$\int_0^{\infty} dx/g(x) = \int_{-\infty}^{0} dx/g(x) = +\infty$$

(A2) $p_0(x) = \exp[\mu_0(x)]$

(A3) $f \in C^1(R)$, $g \in C^2(R)$

Consider the change of coordinates in (1.1)

$$z(t) = \phi[x(t)] = \int_0^{x(t)} dx/g(x) \tag{2.1}$$

Using Itô's formula

$$\begin{aligned} dz(t) &= (f/g - \tfrac{1}{2}g')(\phi^{-1}[z(t)])dt + dv(t) \\ dy(t) &= h(\phi^{-1}[z(t)])dt + dw(t) \end{aligned} \tag{2.2}$$

The associated Zakai equation is

$$\begin{aligned} \text{đ}u(t,z) = \{&\tfrac{1}{2}\partial_{zz}u + [\tfrac{1}{2}g' - f/g](\phi^{-1}(z))\partial_z u \\ &+[g(\tfrac{1}{2}g'' - (f/g)') - \tfrac{1}{2}h^2](\phi^{-1}(z))u\}dt \\ &+ h(\phi^{-1}(z))u\,\text{đ}y(t) \end{aligned} \tag{2.3}$$

(written in the Stratonovich calculus). We can eliminate the first order term in (2.3) by using the exponential transformation

$$\begin{aligned} v(t,z) &= u(t,z)\exp[-\psi(z)] \\ \psi(z) &= \int_0^z (f/g - \tfrac{1}{2}g')(\phi^{-1}(x))dx \end{aligned} \tag{2.4}$$

The equation for $v(t,z)$ is

$$\text{đ}v(t,z) = [\tfrac{1}{2}\partial_{zz}v(t,z) - V(z)v(t,z)]dt + H(z)v\,\text{đ}y(t) \tag{2.5}$$

where

$$V(z) = \tfrac{1}{2}[h^2+(f/g-\tfrac{1}{2}g')^2+g(f/g-\tfrac{1}{2}g')'](\phi^{-1}(z)) \tag{2.6}$$

$$H(z) = h(\phi^{-1}(z))$$

Since the Laplacean in (2.5) is "isolated," the fundamental solution of (2.5) involves Wiener measure, and we can apply the formulas of [5] to evaluate it. This is done in the next section.

Instead of changing coordinates in (1.1) we could have changed the coordinates in the Zakai equation (1.3) and then used an exponential transformation to eliminate the drift term. This would also lead to an equation like (2.5), but involving different functions $\psi(z)$, $V(z)$. While this procedure is perfectly valid in the context of numerical studies, the resulting equation cannot, in general, be associated with a nonlinear filtering problem like (1.1), and we shall not pursue it further.

In situations where f and g are not smooth, i.e., (A3) does not hold, then we must consider some more general type of weak transformation to accomplish the reduction of the Zakai equation. Also, when x(t) is an R^d-valued process with $d \geq 2$, it will be necessary to have the integrand in (2.4) be a gradient. That is, with g(x) dxd and non-singular $g^{-1}(x)f(x) - \frac{1}{2}g'(x)$ will have to be a gradient for the exponential transformation (2.4) to have a meaning.

3. APPROXIMATE EVALUATION OF THE CONDITIONAL DENSITY

Using the Feynman-Kac formula or the Kallianpur-Striebel formula as it is called in filtering theory, we can write the solution to (2.5) as

$$v(t,z) = E_z\{\exp[\int_0^t H(z(s))dy(s) - \frac{1}{2}\int_0^t H^2(z(s))ds] \cdot \exp[-\int_0^t V(z(s))ds+\mu_0(z(t))]\} \tag{3.1}$$

where E_z is expectation over Brownian paths starting at $z(0) = z$. Applying the formulas of Chorin and the Lemma, we can write

$$v(t,z) = I_n(t,z) + O(n^{-2}) \tag{3.2}$$

where

$$I_n(t,z) = (2\pi)^{-n/2}\int_{R^n} K_n(t,z,\underline{r},\underline{y})d\underline{r} \tag{3.3a}$$

$$\underline{y} = (y_0, y_1, \ldots, y_{n-1}), \quad y_i = y(t_{i+1}) - y(t_i)$$

$$\underline{r} = (r_0, r_1, \ldots, r_{n-1}), \quad t_i = it/n, \; i = 0,1,\ldots,n-1 \tag{3.3b}$$

$$\begin{aligned} K_n(t,z,\underline{r},\underline{y}) = \exp\{ & \sum_{i=0}^{n-1} [H(z+\langle\alpha_i^n(t),\underline{r}\rangle)y_i \\ & - \tfrac{1}{2}H^2(z+\langle\alpha_i^n(t),\underline{r}\rangle)t/n \\ & - V(z+\langle\alpha_i^n(t),\underline{r}\rangle)t/n] \\ & +(\mu_0-\psi)(\phi^{-1}(z+\langle\alpha_{n-1}^n(t),\underline{r}\rangle)) \\ & - \tfrac{1}{2}\langle\underline{r},\underline{r}\rangle\} \end{aligned} \tag{3.3c}$$

where for $i = 0,1,\ldots,n-1$

$$\alpha_i^n(t) = (\underset{0^{th}\text{ entry}}{t/\sqrt{2n}}, t/\sqrt{n},\ldots,\underset{(i-1)^{st}\text{ entry}}{t/\sqrt{n}}, 0,\ldots,\underset{(n-1)^{st}\text{ entry}}{0})^T \varepsilon R^n \tag{3.4}$$

and $\langle\alpha,\underline{r}\rangle$ is the Euclidean inner product. The error in (3.2) is interpreted as in the Lemma (even though it has a deterministic component which is $O(n^{-2})$). Efficient evaluation of the n-fold integral (3.3a) is our main objective here.

3.1 A Recursive Evaluation of $I_n(t,z)$

Let

$$w_i = z + \langle\alpha_i^n(t), \underline{r}\rangle, \quad i = 0,1,\ldots,n-1 \tag{3.5}$$

$$\underline{w} = [w_0, w_1, \ldots, w_{n-1}]^T$$

and note $dw_0 = (t/\sqrt{2n})dr_0$, $dw_i = (t/\sqrt{n})dw_i$, $i=1,\ldots,n-1$. So

$$I_n(t,z) = \left(\frac{n}{2\pi}\right)^{n/2} \frac{\sqrt{2}}{t^n} \int_{R^n} \hat{K}_n(t,z,\underline{w},\underline{y})d\underline{w} \tag{3.6}$$

where

$$\begin{aligned} \hat{K}_n(t,z,\underline{w},\underline{y}) = \exp\{ & \sum_{i=0}^{n-1} [H(w_i)y_i - \tfrac{1}{2}H^2(w_i)t/n \\ & - V(w_i)t/n] + (\mu_0-\psi)(\phi^{-1}(w_{n-1})) \\ & - \tfrac{1}{2}\frac{n}{t^2}[2(w_0-z)^2 + (w_1-w_0)^2 + \ldots + (w_{n-1}-w_{n-2})^2]\} \end{aligned} \tag{3.7}$$

If we now define

$$g(t,w,y) = H(w)y - \tfrac{1}{2}H^2(w)t/n - V(w)t/n \tag{3.8}$$

and the sequence of integrals

$$I_1(t,z,w_1) = c_n(t)\int_{-\infty}^{\infty}\exp\{-\frac{1}{2}\frac{n}{t^2}[2(w_0-z)^2+(w_1-w_0)^2]+g(t,w_0,y_0)\}dw_0 \qquad (3.9)_1$$

$$\vdots$$

$$I_k(t,z,w_k) = c_n(t)\int_{-\infty}^{\infty}\exp\{-\frac{1}{2}\frac{n}{t^2}(w_k-w_{k-1})^2+g(t,w_{k-1},y_{k-1})\} \qquad (3.9)_k$$

$$\cdot\ I_{k-1}(t,z,w_{k-1})dw_{k-1},\quad k = 2,3,\ldots,n-1,$$

then

$$I_n(t,z) = c_n(t)\int_{-\infty}^{\infty}\exp\ (\mu_0-\psi)(\phi^{-1}(w_{n-1})) + g(t,w_{n-1},y_{n-1})\}$$

$$\cdot I_{n-1}(t,z,w_{n-1})dw_{n-1} \qquad (3.9)_n$$

where $c_n(t) = (2^{1/2n}\sqrt{n})/(t\sqrt{2\pi})$.

Remarks 1. The integrals $I_1, I_2,\ldots,I_{n-1}$ are independent of the initial data; and so, $(3.9)_n$ has the form of a Green's representation

$$I_n(t,z) = \int_{-\infty}^{\infty}q_0(w)G(t,z;0,w)dw \qquad (3.10)$$

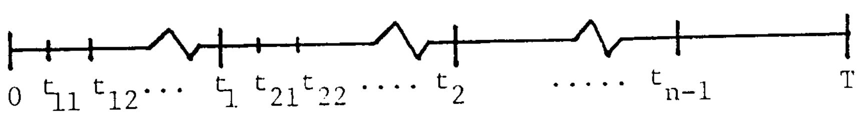
That is, $c_n(t)\exp[g(t,w_{n-1},y_{n-1})]I_{n-1}(t,z,w_{n-1})$ approximates the Green's function of the partial differential equation (2.5) for the (transformed) unnormalized conditional density.

2. The system (3.9) is recursive in the observations $y_0,y_1,\ldots,y_{n-1}$; that is, I_k depends on $y_0,y_1,\ldots,y_{k-1}$, and it is computed from I_{k-1} and y_{k-1}. Unfortunately, it is not recursive in n or in t. This last means that (3.9) does not constitute a recursive filter in the usual sense of the term.

At this point there is considerable flexibility in designing a numerical implementation of (3.9). In discussing possible designs we will show that the n-step recursive evaluation (3.9) has a significant computational advantage over (most) direct evaluations of the n-fold integral in (3.6).

3.2 Implementation

Suppose we require the evaluation of $v(t,z) = I_n(t,z) + 0(n^{-2})$ at points $z = a_1,a_2,\ldots,a_m$ for some $m\geq 1$, and $t=t_1, t_2,\ldots,t_N = T$. The t_j are prespecified observation times. Let $n_1,n_2,\ldots,n_N$ be integers with $n_j\geq 1$, $j = 1,2,\ldots,N$. Consider the discretization of the time interval shown below



Here $t_{11} = t_1/n_1$, $t_{12} = 2t_1/n_1,\ldots,t_{21} = (t_2-t_1)/n_2$, $t_{22}=2(t_2-t_1)/n_2,\ldots$. If we only want $v(t,a_i)$ at $t_1,t_2,\ldots,t_N$, then the mesh points t_{ij} can be chosen arbitrarily; that is, the integers $n_1,n_2,\ldots,n_N$ can be different. The recursion (3.9) is then used to evaluate $I_n(t,z)$ at $t=t_1,t_2,...,t_N$. If $v(t,z)$ is required at the intermediate times t_{ij}, then it is advantageous to take $n_1,n_2=...=n_N=n$, since this permits the use of parallel processors (n of them) to compute the intermediate values.

The operational count for this procedure is as follows: Each integral $I_k(t,z,w)$ is evaluated for each t at $(z,w) = (a_i,a_j), i,j=1,2,...,m$, a total of m^2 evaluations of I_k. For this purpose we can use <u>any</u> quadrature rule for one-dimensional integrals of the form $I(z,w) = \int_{-\infty}^{\infty} e^{-\phi(v)} F(z,w,v)dv$ with ϕ a given weight function; that is,

$$I(z,w) = \sum_{i=1}^{m} F(z,w,a_i)A_i \qquad (3.11)$$

Here A_i are weights associated with the rule, and the a_i are chosen in accordance with the prescription of the quadrature rule. By using the same rule for all the I_k no interpolation is needed to evaluate I_{k+1} in terms of I_k. At the final stage we require only m values of $v(t_1,z)$. Thus, on the initial interval $(0,t_1)$ we make $(n-1)m^2 + m$ evaluations of the integral (3.11) and $(n-1)m^3 + m^2$ function evaluations and the same number of multiplications and additions in (3.11) to compute (3.9). For example, if $m = 20$, $n = 20$, we must compute about 7500 integrals, implying about 150,000 additions and multiplications. While this is not insubstantial, it compares very favorably with product formulas for (3.6) that do not use the special structure in (3.9). By introducing a particular quadrature representation, we can obtain a more precise count of the elementary operations required in the evaluation.

4. FURTHER COMPUTATIONAL ASPECTS OF THE RECURSIVE FORMULAS

By modifying the notation we can streamline the representation of the recursion (3.9) and, in this way, make the underlying structure of the computation more evident. Referring to (3.9) a natural choice for the weight functions ϕ in (3.11) are the diagonal terms of the quadratic form in w_i in (3.7). Unfortunately, it will be seen in section 5 that the error associated with the resulting Hermite-Gauss quadrature rule is an increasing function of bounds on derivatives of the integrand in (3.11); these may not be finite because of the remaining cross terms in the quadratic form. This suggests selecting as weight functions a portion of each diagonal element, and choosing the proportionality constants to preserve the negative definiteness of the quadratic form

$$J^1 = (t/\sqrt{n})\,\mathrm{diag}[\sqrt{2/3},\ 1,\dots,1,\sqrt{2}]$$
$$J^2 = \mathrm{diag}[(1-\alpha_i)^{-\frac{1}{2}}] \tag{4.1}$$

(diag = diagonal matrix), and

$$\underline{w} = J\underline{\xi},\ J=J^1J^2 \triangleq \mathrm{diag}[J_i] \tag{4.2}$$

Then

$$I_n(t,z) = c_n(t,z)\int_{R^n} [\prod_{i=0}^{n-1} e^{-\xi_i^2}/\sqrt{\pi}]\tilde{K}_n(t,z,\underline{\xi},\underline{y})d \tag{4.3}$$

where

$$c_n(t,z) = 2\sqrt{2/3}\prod_{i=0}^{n-1}(1-\alpha_i)^{\frac{1}{2}}\, e^{-nz^2/t^2}/2^{n/2}$$

$$\tilde{K}(t,z,\underline{\xi},\underline{y}) = \exp\{[\mu_0-\psi)(\phi^{-1}(w_{n-1})) + \sum_{i=0}^{n-1} g(t,w_i,y_i)]_{w=J\underline{\xi}}$$
$$- \sum_{i=0}^{n-1} a_{ii}\xi_i^2 - 2\sum_{i=0}^{n-2} a_{i,i+1}\xi_i\xi_{i+1} - b_0\xi_0\}$$

$$b_0 = -\frac{2\sqrt{2/3}\sqrt{n}}{t\sqrt{1-\alpha_0}}\,z \tag{4.4}$$

$$a_{ii} = \alpha_i/(1-\alpha_i)$$

$$a_{i,i+1} = \begin{cases} -(1/\sqrt{6})/\sqrt{(1-\alpha_0)(1-\alpha_1)} & ,\ i = 0 \\ -(1/2)/\sqrt{(1-\alpha_i)(1-\alpha_{i+1})} & ,\ 1\le i\le n-3 \\ -(1/\sqrt{2})/\sqrt{(1-\alpha_{n-2})(1-\alpha_{n-1})} & ,\ i=n-2 \end{cases}$$

The transformation J^1 isolates a factor $\exp-\Sigma(1-\alpha_i)\xi_i^2$, which is then normalized by J^2. To express (4.3) recursively, define

$$r_i^1(\xi_i,t) = \exp[H(J_i\xi_i)]$$
$$r_i^2(\xi_i,t) = \exp[-\tfrac{1}{2}H^2(J_i\xi_i)t/n - V(J_i\xi_i)t/n]$$
$$q_0(\xi_0,z,t) = \exp[(2\sqrt{2n/3}/t\sqrt{1-\alpha_0})z\xi_0] \tag{4.5}$$
$$s_k(\xi_k,\xi_{k+1}) = \exp[a_{kk+1}\xi_k\xi_{k+1} - a_{kk}\xi_k^2],\ 0\le k\le n-2$$
$$s_{n-1}(\xi_{n-1}) = \exp[(\mu_0-\psi)(\phi^{-1}(J_{n-1}\xi_{n-1}))]$$

Here $w_i = J_i\xi_i$ and the dependence on α_i has been suppressed. Then (4.3) becomes

$$I_n(t,z) = c_n(t,z)\int [\prod_{i=0}^{n-1} e^{-\xi_i^2}/\sqrt{\pi}]\ [\prod_{i=0}^{n-1}(r_i^1(\xi_i,t))^{y_i}]$$
$$\cdot[\prod_{i=0}^{n-1} r_i^2(\xi_1,t)s_i(\xi_i,\xi_{i+1})]q(\xi_0,z,t)d\underline{\xi} \tag{4.6}$$

To express this recursively in terms of one dimensional integrals, let $\rho(\xi) = e^{-\xi^2}/\sqrt{\pi}$, and consider the sequence

$$I_1(\xi_1, z) = \tilde{c}_n(t,z)\int_{-\infty}^{\infty}\rho(\xi_0)(r_0^1(\xi_0,t))^{y_0} r_0^2(\xi_0,t) \cdot s_0(\xi_0,\xi_1)q(\xi_0,z,t)d\xi_0$$

$$\vdots$$

$$I_{k+1}(\xi_{k+1}, z) = \tilde{c}_n(t,z)\int_{-\infty}^{\infty}\rho(\xi_k)(r_k^1(\xi_k,t))^{y_k} r_k^2(\xi_k,t) \cdot s_k(\xi_k,\xi_{k+1})I_k(\xi_k,z)d\xi_k, \quad k=1,2,\ldots,n-2 \tag{4.7}$$

$$\vdots$$

$$I_n(t,z) = \tilde{c}_n(t,z)\int_{-\infty}^{\infty}\rho(\xi_{n-1})(r_{n-1}^1(\xi_{n-1},t))^{y_{n-1}} r_{n-1}^2(\xi_{n-1},t) \cdot s_{n-1}(\xi_{n-1})I_{n-1}(\xi_{n-1},z)d\xi_{n-1}$$

where

$$\tilde{c}_n(t,z) = [c_n(t,z)]^{1/n} \tag{4.8}$$

Note that the functions r_k^1, r_k^2, s_k do not depend on k, for k=1,...,n-2.

For fixed α_i, we can integrate $I = \int_{-\infty}^{\infty}\rho(\xi)f(\xi)d\xi$ with an M-point Gaussian formula (see, e.g., [9]):

$$I = S + e_M$$
$$S = \sum_{i=1}^{M} A_i f(a_i) \tag{4.9}$$

where $A_i > 0$ are weights and a_i are the (real) zeros of the M^{th} Hermite polynominal

$$H_M(x) = 2^M x^M + \ldots \tag{4.10}$$

and are symmetric about x = 0. Using this, the k^{th} term in (4.7), k = 1,2,...,n-2, becomes

$$I_{k+1}(a_j, a_i) = \tilde{c}_n(t,a_i)[S_k(a_j,a_i) + e_k] \tag{4.11}$$

where

$$S_k(a_j,a_i) = \sum_{\ell=1}^{M} A_\ell[r^1(a_\ell,t)]^{y_k} r^2(a_\ell,t) \cdot s(a_\ell,a_j)I_k(a_\ell,a_i)$$
$$\approx \sum_{\ell=1}^{M} A_\ell[r^1(a_\ell,t)]^{y_k} r^2(a_\ell,t) \cdot s(a_\ell,a_j)\tilde{c}_n(t,a_i)S_{k-1}(a_\ell,a_j) \tag{4.12}$$

(Note: we have dropped the k-dependence on r^1, r^2, s.)

The error in the second expression should be clear from (4.11). Now let $\underline{S}_k$ be the MxM matrix with elements $S_k(a_j,a_i), i,j=1,2,\ldots,M$.

$$\underline{S}_k = \{S_k(a_j,a_i)\}_{i,j=1}^{M} \tag{4.13}$$

and let

$$R(t,i,j,\ell) = r^2(a_\ell,t)s(a_\ell,a_j)\tilde{c}_n(t,a_i) \tag{4.14}$$

Then

$$S_k(a_j,a_i) = \sum_{\ell=1}^{M} A_\ell[r^1(a_\ell,t)]^{y_k}R(t,i,j,\ell) \cdot S_{k-1}(a_\ell,a_i) \tag{4.15}$$

The terms A_ℓ, $r^1(a_\ell,t)$, $R(t,i,j,\ell)$ can be precomputed "off-line". In this case, calculation of the matrix $\underline{S}_k$ consists of

(i) raising the elements of the vector $r^1(a_\ell,t)$ $\ell = 1,2,\ldots,M$ to the y_k power;

(ii) performing M^2 vector products of $\underline{x}_{ij} = \{A_\ell[r^1(a_\ell,t)]^{y_k} R(t,i,j,\ell),\ \ell = 1,\ldots,M\}$ and $\underline{v}_{k-1}^{i} = \{S_{k-1}(a_\ell,a_i),\ \ell = 1,\ldots,M\}$ for $i,j = 1,2,\ldots,M$.

To compute $I_{k+1}(a_j,a_i)$, it is necessary to multiply $S_k(a_j,a_i)$ by $\tilde{c}_n(t,a_i)$,

Thus, a total of $2M^3+M^2+M$ elementary operations is required to compute the M^2 entries $I_{k+1}(a_\ell,a_i)$, $i,j=1,\ldots,M$. This is for $k=1,2,\ldots,n-2$, a total of $(n-2)(2M^3+M^2+M)$ operations. The operation count for $I_1(a_\ell,a_i)$ is the same, adding $(2M^3+M^2+M)$ operations. Only $2M^2 + 2M$ operations are required to compute the M-vector $I_n(t,a_\ell)$, $\ell=1,\ldots,M$. Thus, the total number of elementary operations N_{oper} required to compute the approximation $I_n(t,z)$ of $v(t,z)$ at the points $z = a_\ell$, $\ell=1,\ldots,M$ is

$$N_{oper} = 2(n-1)M^3 + (n+1)(M^2+M) \tag{4.16}$$

For n=20, M=20 we have $N_{oper} = 312820$.

Note that it is possible to do the multiplications of $\underline{x}_{ij}$ and $\underline{v}_{k-1}^{i}$ and that of $S_k(a_j,a_i)$ and $\tilde{c}_n(t,a_i)$ using "parallel processing" which leads to considerable time savings.

Finally, referring back to (2.4) the evaluation of the unnormalized density $u(t,a_i)$ from $v(t,a_i)$, $i = 1,2,\ldots,M$, requires an additional M function evaluations and M multiplications.

Our final formula (4.11) for the approximation of $v(t,z)$ depends on the weight parameters α_i and involves two basic errors: (i) the error $O(n^{-2})$ in the approximation (1.10) of the function space integral by an n-fold "ordinary" integral; and (ii) the error e_M in the M-point Gaussian quadrature formula (4.9). (The error in the substitution (4.12) is clearly a linear multiple of e_M.) For obvious reasons we would like to have sharp estimates of these quantities, the last of which should be minimized with respect to $\underline{\alpha}$.

The analysis in [5] which led to the estimate $O(n^{-2})$ in (1.10) is so complex that a precise evaluation of the order expression appears to be out of the question. In specific cases one can probably emulate the analysis of Cameron [8], sec. 6, which shows that in some cases the $O(n^{-2})$ estimate is actually very conservative. (E.g., "Simpson's rule," included in our formulas, evaluates the Wiener integral of the functional $F[x] = [\int_0^1 x^2(s)dx]^2$ to within an error of $1/8\pi^4 n^{-3}$.)

We can say somewhat more about the error $e_M = e_M(t,z,\underline{\alpha})$. Natural bounds are of the form

$$|e_M(t,z,\underline{\alpha})| \le C(n,M)N(t,z,\underline{Y},n,\underline{\alpha}) \tag{5.1}$$

where C is the error coefficient and N is some bound on a derivative of the integrand in (4.6). There are several possible procedures, see, e.g., [10][11]. We shall follow the approach used by Lether in [11]. Since the recursive evaluation (4.7) does not reduce or increase the error relative to direct evaluation of the n-fold integral (4.3), we may as well apply Lether's procedure to the latter.

Because the weight function in (4.3) is normalized, and the error coefficient for each coordinate is (independent of the coordinate)

$$\varepsilon(M) = M!/2^M(2M)! \tag{5.2}$$

the error for fixed $(t,z,\underline{\alpha})$ is bounded by (from [11], equ.(7))

$$E_M(t,z,\underline{\alpha}) = \varepsilon(M) \sum_{i=0}^{n-1} \frac{\partial^{2M}}{\partial\xi_i^{2M}} K_n(\tilde{\xi}_0,\ldots,\tilde{\xi}_{n-1}) \tag{5.3}$$

for some point $(\tilde{\xi}_0,\ldots,\tilde{\xi}_{n-1})$ in R^n. Here K_n is the integrand in (4.3). (We have supressed the $t,z,\underline{Y},\underline{\alpha}$ variables in the argument of $\tilde{K}$.) Of course, we assume that $\tilde{K}$ and, therefore ρ_0, f,g,h, are sufficiently smooth for (5.3) to make sense.

The dependence of E_M on α may be found by computing upper bounds for the derivatives of $\tilde{K}_n$. Writing

$$\tilde{K}_n(\underline{\xi},\underline{\alpha}) = F(\underline{J\xi})G(\underline{\xi},\underline{\alpha}),\ F(\underline{w}) = \exp[(\mu_0-\psi)(\phi^{-1}(w_{n-1})) + \sum_{i=0}^{n-1} g(w_i)] \tag{5.4}$$

$$G(\underline{\xi},\underline{\alpha})=\exp[-(\sum_{i=0}^{n-1} a_{ii}(\alpha_i)\xi_i^2 + \sum_{i=0}^{n-2} a_{ii+1}(\alpha_i,\alpha_{i+1})\xi_i\xi_{i+1} + b_0(\alpha_0)\xi_0)]$$

assume F has continuous derivatives up to order $2M+1$, bounded by a constant F_0 possibly depending on the observation path $y(t)$. By completing the square with respect to ξ_i in $G(\underline{\xi},\underline{\alpha})$ and employing a bound for Hermite polynomials [12],

$$H_j(x)e^{-x^2} \le C_0 2^{j/2}\sqrt{j!}\,e^{-x^2/2} \quad , \quad C_0 \cong 1.086$$

and $d^i e^{-x^2}/dx^i = e^{-x^2} H_i(x)\cdot(-1)^i$, it may be shown that

$$E_M(t,z,\underline{\alpha}) \le C_1 \varepsilon(M) \sum_{i=0}^{n-1} \underline{P}_i(\underline{\alpha}) \sup_{\xi} e^{-Q_i(\underline{\xi},\underline{\alpha})} \tag{5.6}$$

$$C_1(n,t,z) = C_0 F_0^n \sqrt{2}\ (\tfrac{n}{2})^{-n/2}\ t^{-n} \exp(-nz^2/t^2)$$

$$P_i(\underline{\alpha}) = (\prod_{j=0}^{n-1}(1-\alpha_j)^{-1/2})(J_i^1)^{-m}(1-\alpha_i)^{-m} \sum_{j=0}^{2m} \binom{2m}{j}\sqrt{j!}\,(2J_i^1\ \alpha_i)^{j/2}$$

$$\cdot\ \exp[\tfrac{1}{8}\ b_i^2/a_{ii} + \tfrac{1}{4}\ B_i^T A_i^{-1} B_i](\underline{\alpha})$$

$$Q_i(\underline{\xi},\underline{\alpha}) = [(\xi + \tfrac{1}{2}\ A_i^{-1}B_i)^T A_i(\xi + \tfrac{1}{2}\ A_i^{-1}B_i)](\underline{\alpha})$$

where

$$A_i(\underline{\alpha}) = M_i(\alpha)A(\alpha),$$

$$A = (a^{ij}),\quad M_i = I_{nxn} - \frac{1}{2}\begin{bmatrix} & 0 & \\ & \vdots & \\ & 0 & \\ & a_{i-1,i}/a_{ii} & \\ 0 & 1 & 0 \\ & a_{ii+1}/a_{ii} & \\ & 0 & \\ & \vdots & \\ & 0 & \end{bmatrix}_{nxn}$$

$$B_0 = \begin{bmatrix} \frac{1}{2}b_0 \\ 0 \\ \\ 0 \end{bmatrix} \quad , \quad B_i = \begin{bmatrix} b_0 \\ 0 \\ \\ 0 \end{bmatrix}$$

Here a_{ij}, b_i are defined in (4.4) and A_i is quint-diagonal. By continuity arguments it may be shown that the matrices A_i are all positive definite in some nonempty subset S of the cube $\mathcal{D} = \prod_{i=0}^{n-1}(0,1)$. In this region the factors $\exp{-Q_i}$ in (5.6) may be bounded by unity. Also, the function $\Sigma\,\underline{P}_i(\underline{\alpha})$ is continuous in $\mathcal{D}$ and diverges to $+\infty$ as $\underline{\alpha} \to \partial\mathcal{D}$ Therefore, $\Sigma\,\underline{P}_i(\underline{\alpha})$ attains a minimum in S at some point $\underline{\alpha}^*(m)$.

This value α^* may be computed numerically and used to define the approximations in section 4, giving rise to the error bound

$$|e_M(t,z,\alpha^*)| \le E_M(t,z,\alpha^*) \le C_1\varepsilon(M) \sum_{i=0}^{n-1} \underline{P}_i(\alpha^*) \qquad (5.7)$$

It is robust in the sense that α^* does not depend on the data f,g,h, or the observation process y(t). Of course, the dependence of (5.7) on (m,n) is crucial. At present, the asymptotic behavior of α^* as $m \to \infty$ has not been determined, but analysis of similar integrals with A_i diagonal suggests that the performance should be good.

Remark : The assumption that $F(\underline{w})$ have bounded derivatives is not very restrictive. From (2.6) (3.8), (5.4) it is clear that we include the cases

(i) f,g,h and their derivatives bounded

(ii) g constant , f,h together with their derivatives of "polynomial" growth, and $\lim_{|x| \to \infty} \operatorname{sgn}(x) f(x) < \infty$.

More general classes of f,g,h may be identified from (2.6),(3.8),(5.4).

REFERENCES

1. H. Kushner, "Dynamical equations for optimal nonlinear filtering", J. Differential Equations, 2 (1967), pp. 179-190.
2. F. Levieux, "Conception d'algorithms parallelisables et convergents de filtrage récursif non-linéare", Appl. Math. Opt., 4 (1977), pp. 61-95.
3. M. Zakai, "On the optimal filtering of diffusion processes", Z. Wahrsch. Verw. Geb., 11 (1969), pp. 230-243.
4. J. S. Baras, G. L. Blankenship, W. E. Hopkins, Jr., "Existence, uniqueness, and asymptotic behavior of solutions to a class of Zakai equations with unbounded coefficients", to appear.
5. G. L. Blankenship, J. S. Baras, "Accurate evaluation of stochastic Wiener integrals with applications to scattering in random media and to nonlinear filtering", SIAM J. Appl. Math., 41 (1981), pp. 518-552.
6. R. H. Cameron, "A family of integrals serving to connect the Wiener and Feynman integrals", J. Math. Physics, 39 (1960), pp. 126-140.
7. A. J. Chorin, "Accurate evaluation of Wiener integrals", Math. Comp., 27 (1973), pp. 1-15.
8. R. H. Cameron, "A Simpson's rule for the numerical evaluation of Wiener's integrals in function space", Duke Math. J., 18 (1951) pp. 111-130.
9. A. H. Stroud, D. Secrest, Gaussian Quadrature Formulas, Prentice-Hall, Englewood Cliffs, N.J., 1966.
10. F. G. Lether, "Cross-product cubature error bounds", Math. Comp., 24 (1970), pp. 583-592.
11. F. G. Lether, "An error representation for product cubature rules" SIAM J. Numer. Anal., 7 (1970), pp. 363-365.
12. A. Erdelyi, ed., Bateman Manuscript Project, Higher Transcendental Functions, vol. II, Calif. Inst. Tech., McGraw-Hill, New York,