NONLINEAR FILTERING: THE EXACT DYNAMICAL EQUATIONS
SATISFIED BY THE CONDITIONAL MODE

Harold J. Kushner*

*This research was supported in part by National Aeronautics and Space Administration under Grant No. NGR-40-002-015, in part by the Air Force Office of Scientific Research, Office of Aerospace Research, United States Air Force, under AFOSR Grant No. AF-AFOSR-693-66 and in part by National Science Foundation under Grant No. GK-967.
The signal is a stochastic process satisfying the stochastic differential equation $dx = f(x)dt + dz$ and observations $\dot{y} = g(x) + \xi$ are taken, where $\xi$ is white noise. The exact dynamical equation for the mode of the conditional density of $x_t$ is derived and discussed.
I. INTRODUCTION: The System and Observation Model.

A filtering problem will be studied for the following system. The signal, $x_t$, satisfies the vector Itô stochastic equation

$$dx = f(x)dt + dz.$$  

(1)

The vector valued observations $\dot{y}_s = h(x_s) + \dot{w}_s$, $s \leq t$, are available at time $t$. For convenience $t = 0$ is the initial time.

$$y_t = \int_0^t h(x_s)ds + w_t$$

$z_t$ and $w_t$ are independent Wiener processes; $\dot{w}_t$ and $\dot{z}_t$ may be considered to be independent white noises. The symbol $d$ denotes a differential; e.g., $dw = \dot{w}_t dt - w_t dt$. The subscript $t$ indicates a functional dependence on time, and will be omitted occasionally. $E(dw)(dw)' = \Sigma_t dt$, where $\Sigma_t$ is independent of $x$, and $E(dz)(dz)' = V(x,t)dt$.

Equation (1) and all subsequent stochastic equations are to be interpreted in the sense of Itô [1] (see also Doob [2]*).

*An alternative interpretation of the stochastic differential equation has been given by Stratonovich [3], and also studied by Wong and Zakai [4] and Clark [5]. Results derived with one interpretation can be transformed by known formulas to results obtained with the other interpretation; see Clark [5] for a discussion specific to the filtering problem.
Notation and the Problem. The conditional density of $x_t$, given $y_s$, $s \leq t$, is written simply as $P(a,t)$. For simplicity the dependence of $P(a,t)$ on $y_s$, $s \leq t$, is suppressed in the notation. The mode of $P(a,t)$, written $a_t$, is a stochastic process (supposing that there is a mode), and the object of this paper is the derivation of a dynamical equation satisfied by $a_t$. As in the well known linear-gaussian case, the equation contains the observation as a driving term.

The result of the paper is new and is, in a sense, an extension of earlier work (Kushner [6],[7]) where dynamical equations for the conditional density $P(a,t)$ and for conditional expectations of functions $E^t h(a) \equiv \int h(a)P(a,t)da$ are derived. The derivation here depends upon the result in [6]. The result in [6], derived formally, pertains to $P(a,t)$; paper [7] provides a rigorous justification of the equations for the conditional expectations $E^t h(a)$. (See also Stratonovich [8].) The simplest of all of the dynamical equations for filtering is, of course, the well known result of Kalman and Bucy [9] for the linear-gaussian case (for which [7] provides a rigorous justification).

The derivation in this paper is formal, although the rules of the calculus of Itô are followed for manipulating differentials of functionals of Wiener processes.

Whether it is more useful to know the conditional mode, rather than the conditional mean, for some filtering problem, is a
controversial matter depending on the specific problem of concern, and is of no interest here. Despite the controversy, it is useful to compare different methods of processing observed data, or at least to have available alternative means for such processing. Furthermore, the estimate of the conditional mode has the intuitive appeal attached to a 'maximum likelihood' estimate. The derivation is for a continuous time problem. Whether or not it has relevance for the discrete problem remains to be seen. (The result does, of course, describe the movement of the mode for a continuous system with discrete observations -- 'between' the observations.) In any case, other than a straightforward application of Bayes' rule, at present there is no really satisfactory dynamical system representation for a filter for a really non-linear and discrete problem.

As will appear, for the non-linear problem, a system corresponding to the exact equation cannot be built with finite components (which is also true for the equation for the conditional mean). Nevertheless, it is useful to have an explicit exhibition of the exact equation, so that approximations derived by any method can be compared, or for its usefulness in suggesting approximations, or to know what must be approximated. This is especially true in view of the great importance and difficulty of the problem of processing 'non-linear' data. The issue of adequate approximations is still wide open.

An only slightly different problem, in discrete and continuous time, was considered in several interesting papers (Bryson
Let the system be \( x_{n+1} = f_n(x_n) + \xi_n \), with observation \( \theta_n = g(x_n) + \psi_n \).

Then, the discrete time equivalent to our problem is to find the sequence \( x_n \) maximizing the sequence \( P(x_n|\theta_1, \ldots, \theta_n) \). The problem in [10], [11], and [12] is to find the sequence of last values \( x_n \) maximizing the sequence \( P(x_1, \ldots, x_n|\theta_1, \ldots, \theta_n) \), and was 'approximately' resolved via approximations to a two-point boundary problem.

An advantage to be claimed by approximating the equations of the sequel, rather than using approximations derived indirectly, as in [10], [11], [12], is that the exact meaning of all dropped or substituted terms is clear; there are no longer multipliers or other quantities which are not defined in terms of known quantities. Of course, the numerical results of the cited works stand by themselves, although their relation to the mode is not clear.

It is implicitly supposed that \( P(a,t) \) is sufficiently differentiable with respect to the components of 'a'. Then, at \( a_t \), the gradient satisfies \( P_{a}(a_t,t) \equiv 0 \). Subscripts \( a_i \) denote a derivative, subscript 'a' a gradient, the subscript \( aa \) denotes the matrix of second partials, and \( P_{a_ia_\alpha} \) denotes \( [P_{a_\alpha}]_{a_i} \). It is supposed that \( P_{aa}(a_t,t) \) is not singular, for if it is, either the mode is not unique, or a finite jump in \( a_t \) can occur in an infinitesimal of time, or other analytical problems arise.

Actually, it is not necessary to suppose that \( P(a_t) \)
is unimodal. Supposing that $a_0$ is the location of a local maximum of $P(a,0)$, $(P_a(a_0,0) = 0)$, the result, equation (2), gives the movement of the location of the local maximum $a_t$ corresponding to initial location $a_0$, as long as $a_t$ moves continuously and $P_{aa}(a_t,t)$ is non-singular.
II. THE MAIN RESULT.

The derivation, a computation using Itô's calculus and the result of [6], appears in the appendix. The general result is that the vector process (the conditional mode) $a_t$ satisfies the stochastic differential equation* (2). The functions $f$, $V$, $g$, $P$ and their derivatives are evaluated at $(a_t, t)$. The prime ' is transpose, $g'_a$ is the matrix whose $i$th column is the gradient (with respect to $a$) of $g_i$, the $i$th component of the vector $g$. $dy_idy_j$ is to be interpreted as $E dy_i dy_j = \sigma_{ij} dt$ (see appendix), the $i,j$th element of $\Sigma$.

$$\text{da} = -P \cdot P_{aa}^{-1} g'_a \Sigma^{-1} (dy_idt - P_{aa}^{-1} f_a) \cdot \text{dt}$$

$$+ P_{aa}^{-1} (dy'_a \Sigma^{-1} g'_a) P_{aa}^{-1} \Sigma^{-1} dy - \frac{1}{2} P_{aa}^{-1} P_{aa}$$

where $u$ is a vector with components

$$u_i = (dy'_a \Sigma^{-1} g'_a P_{aa}^{-1} a_i a_i (P_{aa}^{-1} \Sigma^{-1} dy)$$

and

$$L*P = - \sum_i (f_i P) a_i + \frac{1}{2} \sum_{i,j} (v_{ij} P) a_i a_j$$

Recall that $\partial Q/\partial t = L*Q$ is Kolmogorov's equation for the density.

*If the reader prefers 'white noise' to differentials of Wiener processes, he may (at the risk of falling into one of the many traps laid by the 'white noise' concept) divide (2) and subsequent equations through by $dt$. 
of the conditional process $x_t$.

Next, write the equations for scalar valued observations. In this case, $dy$ is a scalar differential and $(dy)^2$ is to be interpreted as $\sigma^2 dt = E dt$ (see appendix) and

\[
\begin{align*}
d a &= -P\cdot \frac{p^{-1}}{\sigma^2} g'_a(dy-gdt) - P^{-1}(L*P)_a dt \\
(4) &+ P^2\sigma^2 \frac{P^{-1}}{\sigma^2} g'_{aa} P^{-1} g'_a dt - P^2\frac{P^{-1}}{2\sigma^2} dt,
\end{align*}
\]

where (note that $g'_a$ is a column vector)

\[
\begin{align*}
r'_i &= (g_{aa}P^{-1})P_{ai} (P^{-1}g'_a),
\end{align*}
\]

and $g_{aiaj}$ is the $i,j$th element of the matrix $g_{aa}$.

If, in addition, $x_t$ is a scalar valued process, then (4) simplifies to the equation, where the abbreviation $q = -P/P_{aa}$ is used, $P_a = 0$ and $V$ is scalar,

\[
\begin{align*}
da &= \frac{qg_a}{\sigma^2} (dy-gdt) + [f_q + \frac{qV_{aa}}{2} - \frac{3V_a}{2} - \frac{VP_{aa}}{2P_{aa}}] \\
&+ \frac{q^2g_{aa}g_a}{\sigma^2} dt - \frac{q^2P_{aa}g_a^2}{2P_{aa}\sigma^2} dt.
\end{align*}
\]

In view of the complexity of (2), the remainder of the discussion of qualitative properties will concern only the scalar
case (5). If the realizations of the processes $P_{aa}$ and $P_{aaa}/P_{aa}$ were available, then the system (5) could be built*, and we would have a dynamical system whose input is the observation, and whose output is the conditional mode. If the system were completely linear and gaussian, then $P_{aaa} = 0$ and $P_{aa}$ is the negative of the conditional covariance of the conditional mean (which then coincides with $a_t$); i.e., $q = -P_{aa}$ satisfies the ricatti equation.

There are two cases in which (3), (4) and (5) can be immediately checked. One is the completely linear-gaussian case, the other is the case where $f = 0$ and $V = 0$ (no dynamics). In the latter case, $x_t \equiv x$ and the conditional mode satisfies

\begin{equation}
F_a(y_t, t, a_t) = 0
\end{equation}

where ($s$ indicates time dependence)

\begin{equation}
F(y_t, t, x) = P(x, 0) \cdot \exp \left( \int_0^t g_s \sum_s (d y_s - \frac{1}{2} g_s ds) \right).
\end{equation}

(6) follows by noting that $(y_i y = y_i y - y_i a + a t)$

*It is, in fact, possible to do this in an approximate sense when the noises $\tilde{w}_t$ and $\tilde{z}_t$ are 'wide band'. See Clark [5] and Stratonovich [3]. We cannot pursue this matter here, but the general idea is to transform (5) by adding some terms, then divide by $dt$, suppose that $\tilde{w}_t$ is 'wideband' and use ordinary analog components. The transformations are discussed in [3] and [5]; the wider the 'bandwidth', the better the approximation. See also Wong and Zakai [4]. Also, if the noises $\tilde{w}_n - \tilde{w}_n$ are truly independent for small $\Delta$, then (5) may be simulated on a digital computer directly.
\[
P(s,t) = \lim_{\Delta \to 0} \frac{\exp - \frac{1}{2\Delta} \sum_{i=1}^{n} (\delta_i y - g(x)\Delta) \cdot \Sigma_{i\Delta}^{-1} (\delta_i y - g(x)\Delta) \cdot P(x,0)}{\int \exp - \frac{1}{2\Delta} \sum_{i=1}^{n} (\delta_i y - g(x)\Delta) \cdot \Sigma_{i\Delta}^{-1} (\delta_i y - g(x)\Delta) \cdot P(x,0) dx}
\]

where the denominator does not depend on \( x \), and the second order terms in \( \delta_i y \) in numerator and denominator cancel each other. Also \( P/P_{aa} = F/F_{aa} \), and \( P_{aaa}/P_{aa} = F_{aaa}/F_{aa} \), since \( F = (\text{constant}) P \). Now applying Itô's Lemma and the implicit function theorem to (6) yields the differential of \( a_t \) in terms of \( dt \) and \( dy \). Since a more general case is treated in the appendix, the details here are left to the reader.

Returning to (5), denote \(-P/P_{aa}\) by \( q \). It is possible to derive differential equations satisfied by \( q \) and by \( P_{aaa}/P_{aa} \). These, in turn, will involve new terms (e.g., \( P_{aaa}/P_{aa} \)), and so on.

In view of the identification, in the gaussian linear case, of \( q \) with the conditional variance, an equation for \( q \) (and its possible resemblance to the Ricatti equation) would be of interest. First note that all functions below of the form \( P/P_{aa} \), \( P_{aa} \), \( F \), \( f \), \( V \), etc. are evaluated at \((a_t, t)\) at time \( t \). Also for a time function of the form \( L(a_t, t) = F(t) \), \( dF \) is a total stochastic differential:

\[
dL = dF = L(a_{t+dt}, t+dt) - L(a_t, t).
\]

Then via an application of Itô's Lemma (see appendix)

\[
-dq_t = d(F(P_{aa}^{-1}) = (dF)/P_{aa} - (dP)/P_{aa}^2/F_{aa}^2
\]

(7)

\[
+ P(dP_{aa}^2/P_{aa}^2 - (dP)(dP_{aa})/P_{aa}^2,
\]
where, by virtue of Itô's Lemma, the products of the differentials are to be replaced by the expectations of the products up to $O(dt)$ terms. An evaluation of (7) (see appendix) gives

$$dq = \frac{-1}{P_{aa}} [L*P + q(L*P)_{aa}]dt$$

$$- \frac{g_{aa}}{2\sigma^2} [q^2 + q^3 \frac{P_{aaa}}{P_{aa}}]dt$$

$$(8)$$

$$+ q^3 \frac{g_a g_{aaa}}{\sigma^2} dt - q \frac{P_{aaa}}{P_{aa}} (L*P)_{a} dt - \frac{q^2 g_{aa}}{\sigma^2} [g_{aa} - (E^t g - q g_{aa})]dt$$

$$- 3q^3 \frac{g_a P_{aaa}}{P_{aa} \sigma^2} [g_{aa} - \frac{g_a P_{aaa}}{2P_{aa}}]dt + \frac{(dy^t g dt)q^2}{\sigma^2} [g_{aa} - \frac{P_{aaa} g_{aa}}{P_{aa}}]$$

where $E^t g \equiv E[g(x_t)|y_s, s \leq t]$.

The first term of (8) equals

$$\frac{q}{2} V_{aa} - \frac{V_2}{2} + q[-f_{aaa} q + 2f_a + f \frac{P_{aaa}}{P_{aa}} + \frac{V_{aaa}}{2} q -$$

$$(9)$$

$$-3V_{aa} - 2V_a \frac{P_{aaa}}{P_{aa}} - \frac{VP_{aaa}}{2P_{aa}}]$$

Note that (8) reduces to the riccati equation if the system is completely linear and gaussian, for then $V_a = V_{aa} = V_{aaa} = g_{aa} = g'_{aa} = 0$, and at $a = a_t$, $P_{aaa} = 0$ and $P_{aaa}/P_{aa} = -3/q$. For comparison, (10) and (11) are the exact equations for the conditional mean ($m_t$) and variance ($m_{2t}$) for the general scalar
non-linear problem (see [6], [7] for the method of derivation).

\[(10) \quad dm = E^t f dt + \frac{1}{\sigma^2} (dy - E^t g dt)(E^t(x-m)g)\]

\[dm_2 = [v^2 + 2E^t(x-m)f - (E^t(x-m)g)^2/\sigma^2]dt\]

\[(11) \quad + \frac{(dy - E^t g dt)}{\sigma^2} E^t [(x-m)^2 - m_2]g.\]

The question of useful approximations to (5) and (8) is still quite open. It seems customary in the derivation of 'approximate' or 'linearized' filters to drop all terms with too many derivatives. The effect of this or any other procedure is not known, although some numerical investigations and comparisons are currently being conducted on the system (10), (11). A rather interesting and detailed theoretical treatment of linearization for a class of discrete time problems appears in [13].

Observations. The movement of the conditional mean and covariance involve functionals of the functions \(f, V\) and \(g\), (i.e., \(E^t f\), \(E^t gh\), etc.). The movement of the conditional mode depends directly on the properties of \(f, V\) and \(g\) in the neighborhood of \(a_t\), and indirectly, via \(q\), on the entire function.

Consider the special case where the instantaneous values satisfy (at \(a_t\)) \(P_{aa} = V = g_a = g_{aa} = 0\); i.e., symmetry of \(P\), no
observations and no system driving noise. Then (5) reduces to

\begin{equation}
(12) \quad da = (f - qf_a)dt
\end{equation}

\begin{equation}
(13) \quad dq = q[-f_{aa}q + 2f_a]dt.
\end{equation}

The behavior exhibited by (12) is slightly surprising at first. Suppose \( f(x) \) is given by the quadratic of Figure 1, where \( f_{aa} > 0 \), all \( a \). Then, for \( f(a_t) = 0 \) at time \( t \), the instantaneous change in \( a_t \) is in the negative direction, as contrasted to the instantaneous change in \( m_t \), which must be positive (recall \( P(x,t) \) is supposed symmetric). In fact, for this example, the mode tends to \( x' \) while the mean tends to \( +\infty \). This behavior is explained by the fact that, owing to the dynamics \( f_{aa} > 0 \), the 'density of probability' is thinning out faster for \( x > x' \), than for \( x < x' \). The approximate solutions of [10], [11], [12] do not exhibit this behavior (although their problem is slightly different, it appears that this qualitative feature should be preserved). Even if there are no observations, the 'odd' terms contribute to \( dq \), and there does not appear to be any a priori reason why such terms should be neglected in comparison to other terms (the usual linearizations do neglect the odd terms). The complexity of (8) is rather disturbing, but helps to emphasize that great care may be necessary in choosing an approximation.

In any case, there are only 2 unknown terms in (8),
\( q_3 = P_{aa}/P_{aa} \) and \( q_4 = P_{aaaa}/P_{aa} \). Equations for these may also be written, and will involve higher derivatives. As the order of the derivative increases, the less relevant to the mode it becomes. Nevertheless, some approximation, preserving the important qualitative effects of the higher derivatives on the motion of the estimate of the mode, must ultimately be made, and is currently under investigation for a system of the type of (10) and (11).

To understand an approximation, it must be studied in the context of some particular problem. Hopefully, we will be more enlightened in the future. The question of approximations would take up too much space here.
III. APPENDIX. Background: Itô's Lemma and related results.

The discussion is purely heuristic and no conditions for validity are given. Let \( w_t = \{w_{it}\} \) be a set of independent normed Wiener processes. By the notation

\[
U_{it} = \int_0^t F_i(U_{is}, s)\,ds + \int_0^t \sum_j G_{ij}(U_{is}, s)\,dw_j, \quad i = 1, \ldots, n.
\]

is meant that there exists a process \( \{U_{it}\} = U_t \) which satisfies (where the stochastic integral is understood in the sense of Itô [1],[2])

\[
U_{it} = U_{io} + \int_0^t F_i(U_{is}, s)\,ds + \int_0^t \sum_j G_{ij}(U_{is}, s)\,dw_j, \quad i = 1, \ldots, n.
\]

Let \( Q(U, s) \) be a doubly differentiable function of \( U_1, \ldots, U_n, s \); then, roughly speaking, Itô's Lemma states that

\[
Q(U_t, t) = Q(U_0, 0) + \int_0^t dQ(U_s, s),
\]

where

\[
dQ = \frac{\partial Q}{\partial t} + \sum_j \frac{\partial Q}{\partial U_j} dU_j + \frac{1}{2} \sum_{j, k} \frac{\partial^2 Q}{\partial U_j \partial U_k} (dU_j)(dU_k).
\]
where by the second order terms \((dU_j)(dU_k)\) is meant the expectation (given the \(U_i\)) up to first order terms or

\[
E[\sum_k G_{ik} dw_k \sum_k G_{jk} dw_k] = \sum_k G_{ik} G_{jk} dt .
\]

In view of the definition of the product \((dU_j)(dU_k)\), the term is equal to any other with the same expectation, conditioned on \(\{U_{it}\}\). Higher order products are said to be zero, in that they need not be taken into account in computing the integral \((A3)\). This result will be basic: It defines a differential so that, with a proper interpretation of the integral, \((A2)\) holds, and \(dU\) has the intuitive interpretation \(dU = U(t+dt) - U(t)\). In the sequel, \(t\) denotes time dependence, \(i\) a subscript, \(\alpha_i\) or \(\alpha_i\), etc., derivatives; other notation, e.g., the subscript \(aa\), has been already given in the text.

Suppose that the \(F_i\) and \(G_{ij}\) depend on a constant vector parameter \(\alpha = (\alpha_i)\) and are sufficiently differentiable with respect to it. Then

\[
(A4) \quad dU_{i\alpha_k} = F_{i\alpha_k} (U_t, \alpha_s) dt + \sum_j G_{ij\alpha_k} (U_t, \alpha_s) dw_j .
\]

Now, suppose that \(\alpha_t\) is a process satisfying

\[
(A5) \quad d\alpha_j = A_j (U_t, \alpha_t, t) dt + \sum_k B_{jk} (U_t, \alpha_t, t) dw .
\]
Then (A4) is now only a 'partial' differential. Define \( d'L(\alpha_t, t) \) = \( d'L \) as the stochastic differential of \( L(c, t) \), where \( c \) is a constant, but evaluated at \( c = \alpha_t \). Intuitively \( d'L(\alpha_t, t) = L(\alpha_t, t + dt) - L(\alpha_t, t) \).

With the foregoing conventions regarding products of two or more differentials, the following formal result is seen to hold.

\[
dU_1(\alpha_t, t) = U_1(\alpha_{t+dt}, t+dt) - U_1(\alpha_t, t) =
\]

\[
= [U_1(\alpha_t, t+dt) - U_1(\alpha_t, t)] + [U_1(\alpha_{t+dt}, t+dt) - U_1(\alpha_t, t+dt)]
\]

\[
= d'U_1(\alpha_t, t) + \sum_j U_{i\alpha_j}(\alpha_t, t+dt) d\alpha_j +
\]

\[
+ \frac{1}{2} \sum_{j,k} U_{ij\alpha_j}(\alpha_t, t+dt)(d\alpha_j)(d\alpha_k).
\]

In the last term on the right, the \( dt \) argument contributes a 3′rd order term; also \( U_{ij\alpha_j}(\alpha_t, t+dt) = U_{ij\alpha_j}(\alpha_t, t) + d'U_{ij\alpha_j}(\alpha_t, t) \).

Thus, dropping the \( (a_t, t) \) arguments,

\[
dU_1 = d'U_1 + \sum_j U_{i\alpha_j} d\alpha_j + \sum_j (d'U_{i\alpha_j})(d\alpha_j)
\]

(A6)

\[
+ \frac{1}{2} \sum_{j,k} U_{ij\alpha_j\alpha_k}(d\alpha_j)(d\alpha_k),
\]
or
\[ U_i(\alpha_t, t) = U_i(\alpha_0, 0) + \int_0^t d U_i. \]

If the \( dU_i, d'U_i, d''U_i, \alpha_j, \) and \( \alpha_j, \alpha_k \) were given, then the \( d\alpha_j \) could be computed. In fact, \( d\alpha_j \) can be no more than a second order polynomial in the other differentials.

**Derivation of (2).**

In [6], it is formally shown that the conditional density \( P(a, t) \) satisfies

\[ (A7) \quad dP(a, t) = P(a, t)(dy - E^t g dt)' \Sigma_0(g(a) - E^t g) + (L^*P(a, t)) dt \]

(see text for definitions of \( E^t g \) and \( L^*P \).) The conditional moments satisfy equations derivable from \((A7)\), and these equations are justified rigorously in [7]. It is supposed that \( P(a, t) \) is sufficiently differentiable with respect to \( a \). **It is also supposed** that \( a_t \) satisfies a stochastic differential equation (i.e., it has a stochastic differential). If this is not assumed, then the same result can be obtained by a longer related argument, requiring the taking of formal limits, and which does not seem to be more advantageous than the present argument.

The following scalar, vector and matrix equations may be verified from \((A7)\). The arguments \((a, t)\) are mostly omitted.
(A8) \[ d'P'_a(a,t) = P_g \sum \left( (dycb - gdt) + (L^*_P)_a dt \right) \]

\[ + \text{ terms linear in } P_a \]

(A9) \[ d'P_{aa}(a,t) = (dycb - gdt) \left( \sum (g-E^t g) P_{aa} \right) \]

\[ + \text{ terms linear in } P_a \]

(A10) \[ d'P_{ai aj a_k}(a,t) = (dycb - gdt) \left( \sum \left[ g_{aj} P_{ai a_k} + g_{ak} P_{ai aj} \right] \right) \]

\[ + \text{ terms linear in } P_a \]

By definition of \( a_t \), \( P_a(a_t, t) \equiv 0 \) and \( P_a(a_t + \delta_t a_t, t + \Delta t) \equiv 0 \), where \( \delta_t a = a_{t+\Delta t} - a_t \), or, equivalently, \( dP_a(a_t, t) \equiv 0 \).

Since \( dU_i = dP_{ai} \equiv 0 \),

\[ 0 = d'P_{ai}(a_t, t) + \sum_j P_{ai aj}(a_t, t) da_j \]

(A11)

\[ + \sum_j (d'P_{ai aj}(a_t, t))(da_j) + \frac{1}{2} \sum_{j, k} P_{ai aj a_k}(a_t, t)(da_j)(da_k). \]

In vector notation, and dropping the arguments,

(A12) \[ 0 = d'P_a + (P_{aa} + d'P_{aa}) (da) + \frac{1}{2} \cdot s, \]
where \( s \) is a vector with \( i \)th component

\[
s_i = (da)'P_{a_i a a}(da).
\]

It is clear that, if the differentials actually satisfy (A12), then \( da \) may be obtained in the following way: From (A12),

\[
da = -(I + P_{a a}^{-1}dP_a)^{-1}P_{aa}^{-1}(d'P_a)
\]

(A13)

\[
- \frac{1}{2} (I + P_{a a}^{-1}d'P_a)^{-1}P_{a a}^{-1} s.
\]

The first term of (A13) equals \( -P_{a a}^{-1}[d'P_a - d'P_{a a}P_{a a}^{-1}d'P_a] \).

Substitute this into (A13), and then substitute the entire right side of (A13) into the \( da \) terms of \( s \). Obviously, in this latter substitution only the first order terms \( (-P_{a a}^{-1}d'P_a) \) will matter, for the others will put products of more than two differentials into the term \( s \).

Thus, finally, retaining only differentials up to the second order,

(A14) \[
da = -P_{a a}^{-1}d'P_a + P_{a a}^{-1}(d'P_a)P_{a a}^{-1}d'P_a - \frac{1}{2} P_{a a}^{-1}s,
\]

where

(A15) \[
\tilde{s}_i = (d'P_a)'P_{a a}^{-1}d'P_{a a}^{-1}P_{a a}^{-1}(d'P_a).
\]
The substitution of the appropriate quantities from (A8) to (A10) (without the zero $P_a$ terms) into (A15) and keeping only terms whose expectations are no greater than $dt$, yields the equivalent of (2).

The evaluations of (7) requires merely $dP_{aa}$ and $dP$ (with, again, the substitution of second order terms by their average values up to $O(dt)$). As was done in the previous procedure (here $a, g$ are scalar valued),

$$dP = P(a_{t+dt}, t+dt) - P(a_t, t)$$

(A16) $$= [P(a_{t+dt}, t+dt) - P(a_t, t)] + [P(a_{t+dt}, t+dt) - P(a_t, t+dt)]$$

$$= d'P + a_{da} + P_{aa}(da)^2/2 + (d'P_a)(da).$$

The second term of (A10) is zero. $dP_{aa}$ is derived similarly. In the interest of saving space the substitutions will not be carried out, except for noting that

$$dP = \frac{(dy-E^tgdt) \ P(g-E^tg) + \{L^*P + P_{aa} \frac{\sigma^2 g_a^2}{2} + \frac{Pq_a}{2}\}}{\sigma^2} dt$$

$$dP_{aa} = \frac{(dy-E^tgdt)}{\sigma^2} [(g-E^tg)_{Paa} + Pg_{aa} + P_{aaa} qga]$$

$$+ \{(L^*P)_{aa} + P_{aaa} \frac{-L^*P}{P_{aa}} + \frac{q^2 g_a g_a}{\sigma^2} - \frac{q^2 P_{aaa} g_a^2}{2p_{aa} \sigma^2} + \frac{P_{aaa} q^2 g_a^2}{2\sigma^2}\}$$

$$+ \frac{(q^2 g_a)}{\sigma^2} (3g_a P_{aa} + g_{aaa} P + (g-E^t g)P_{aaa})] dt.$$
CONCLUSIONS

The exact dynamical equation satisfied by the conditional mode has been derived. The equation depends on terms $P/P_{aa'}$, $P_{aa'a}/P_{aa'}$, etc., which in turn, are given by other stochastic equations. The system, for a non-linear problem is not finite dimensional, but is at least exact, and is the most useful starting point for the study of approximations.
REFERENCES


