

Research article

Open Access

Universal nonlinear filtering using Feynman path integrals II: the continuous-continuous model with additive noise

Bhashyam Balaji

Address: Radar Systems Section, Defence Research and Development Canada, Ottawa, 3701 Carling Avenue, Ottawa, ON K1A 0Z4, Canada

Email: Bhashyam Balaji - Bhashyam.Balaji@gmail.com

Published: 10 February 2009

Received: 24 April 2008

PMC Physics A 2009, 3:2 doi:10.1186/1754-0410-3-2

Accepted: 10 February 2009

This article is available from: <http://www.physmathcentral.com/1754-0410/3/2>

© 2009 Balaji

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

In this paper, the Feynman path integral formulation of the continuous-continuous filtering problem, a fundamental problem of applied science, is investigated for the case when the noise in the signal and measurement model is Gaussian and additive. It is shown that it leads to an independent and self-contained analysis and solution of the problem. A consequence of this analysis is the configuration space Feynman path integral formula for the conditional probability density that manifests the underlying physics of the problem. A corollary of the path integral formula is the Yau algorithm that has been shown to have excellent numerical properties. The Feynman path integral formulation is shown to lead to practical and implementable algorithms. In particular, the solution of the Yau partial differential equation is reduced to one of function computation and integration.

PACS Codes: 02.50.Ey, 02.50.Fz, 05.10.Gg, 89.90.+n, 93E10, 93E11

I Introduction

1.1 Motivation

The fundamental dynamical laws of physics, both classical and quantum mechanical, are described in terms of variables continuous in time. The continuous nature of the dynamical variables has been verified at all length scales probed so far, even though the relevant dynamical variables, and the fundamental laws of physics, are very different in the microscopic and macroscopic realms. In practical situations, one often deals with macroscopic objects whose state is phenomenologically well-described by classical deterministic laws modified by external disturbances that can be modelled as random noise, or Langevin equations. Even when there is no underlying fundamental dynamical law, the Langevin equation provides an effective description of the state variables in many applications. It is therefore natural to consider the problem of the

evolution of a state of a signal of interest described by a Langevin equation called the state process.

When the state model noise is Gaussian (or more generally multiplicatively Gaussian) the state process is a Markov process. Since the process is stochastic, the state process is completely characterized by a probability density function. The Fokker-Planck-Kolmogorov forward equation (FPKfe) describes the evolution of this probability density function (or equivalently, the transition probability density function) and is the complete solution of the state evolution problem.

However, in many applications the signal, or state variables, cannot be directly observed. Instead, what is measured is a nonlinearly related stochastic process called the measurement process. The measurement process can often be modelled by yet another continuous stochastic dynamical system called the measurement model. In other words, the observations, or measurements, are discrete-time samples drawn from a different Langevin equation called the measurement process.

The *conditional probability density function* of the state variables, given the observations, is the complete solution of the filtering problem. This is because it contains all the probabilistic information about the state process that is in the measurements and in the initial condition [1]. This is the Bayesian approach, i.e., the *a priori* initial data about the signal process contained in the initial probability distribution of the state is incorporated into the solution. Given the conditional probability density, optimality may be defined under various criterion. Usually, the conditional mean, which is the least mean-squares estimate, is studied due to its richness in results and mathematical elegance. The solution of the optimal nonlinear filtering problem is termed universal, if the initial distribution can be arbitrary.

1.2 Fundamental Stochastic Filtering Results

When the state and measurement processes are linear, the linear filtering problem was solved by Kalman and Bucy [2,3]. The celebrated Kalman filter has been successfully applied to a large number of problems in many different areas.

Nevertheless, the Kalman filter suffers from some major limitations. The Kalman filter is not optimal even for the linear filtering case if the initial distribution is not Gaussian. It may still be optimal for a linear system under certain criteria, such as minimum mean square error, but not a general criterion. In other words, the Kalman filter is not a universal optimal filter, even when the filtering problem is linear. Secondly, the Kalman filter cannot be an optimal solution for the general nonlinear filtering problem since it assumes that the signal and measurement models are linear. The extended Kalman filter (EKF), obtained by applying the Kalman filter to a linearized model, cannot be a reliable solution, in general. Thirdly, even when the EKF estimates the state well in some cases, it gives no reliable indication of the accuracy of the state estimate, i.e., the conditional variance is unreliable. Finally, the Kalman filter assumes that the conditional prob-

ability distribution is Gaussian, which is a very restrictive assumption; for instance, it rules out the possibility of a multi-modal conditional probability distribution.

The continuous-continuous nonlinear filtering problem (i.e., continuous-time state and measurement stochastic processes) was studied in [4-6] and [7]. This led to a stochastic differential equation, called the Kushner equation, for the conditional probability density in the continuous-continuous filtering problem. It was noted in [8,9], and [10] that the stochastic differential equation satisfied by the unnormalized conditional probability density, called the Duncan-Mortensen-Zakai (DMZ) equation, is linear and hence considerably simpler than the Kushner equation. The robust DMZ equation, a partial differential equation (PDE) that follows from the DMZ equation via a gauge transformation, was derived in [11] and [12].

A disadvantage of the robust DMZ equation is that the coefficients depend on the measurements. Thus, one does not know the PDE to solve prior to the measurements. As a result, real-time solution is impossible. A fundamental advance was made in tackling the general nonlinear filtering problem by S-T. Yau and Stephen Yau. In [13], it was proved that the robust DMZ equation is equivalent to a partial differential equation that is independent of the measurements, which is referred to as the Yau Equation (YYe) in this paper. Specifically, the measurements only enter as initial condition at each measurement step. Thus, no on-line solution of a PDE is needed; all PDE computations can be done off-line.

However, numerical solution of partial differential equations presents several challenges. A naïve discretization may not be convergent, i.e., the approximation error may not vanish as the grid size is reduced. Alternatively, when the discretization spacing is decreased, it may tend to a different equation, i.e., be inconsistent. Furthermore, the numerical method may be unstable. Finally, since the solution of the YYe is a probability density, it must be positive which may not be guaranteed by the discretization.

A different approach to solving the PDE was taken in [14] and [15]. An explicit expression for the fundamental solution of the YYe as an ordinary integral was derived. It was shown that the formal solution to the YYe may be written down as an ordinary, but somewhat complicated, multi-dimensional integral, with an infinite series as the integrand. In addition, an estimate of the time needed for the solution to converge to the true solution was presented.

1.3 Outline of the Paper

In this paper, the (Euclidean) Feynman path integral (FPI) formulation is employed to tackle the continuous-continuous nonlinear filtering problem. Specifically, phrasing the stochastic filtering problem in a language common in physics, the solution of the stochastic filtering problem is presented. In particular, no other result in filtering theory (such as the DMZ equation, the robust DMZ equation, etc.) is used. The path integral formulation leads to a path integral formula for the transition probability density for the general additive noise case. A corollary of the FPI for-

malism is the path integral formula for the fundamental solution of the YYe and the Yau algorithm – a fundamental result of nonlinear filtering theory. It is noted that this paper provides a detailed derivation of results that were used in [16].

The following point needs to be emphasized to readers familiar with the discussion of standard filtering theory – the FPI is different from the Feynman-Koc path integral. In filtering theory literature, it is the Feynman-Koc formalism that is often used. The Feynman-Koc formulation is a rigorous formulation and has led to several rigorous results in filtering theory. However, in spite of considerable effort it has not been proven to be directly useful in the development of reliable practical algorithms with desirable numerical properties. It also obscures the physics of the problem.

In contrast, it is shown that the FPI leads to formulas that are eminently suitable for numerical implementation. It also provides a simple and clear physical picture. Many path integral manipulations have no counterpart in the Feynman-Koc approach. Finally, the theoretical insights provided by the FPI are highly valuable, as evidenced by numerous examples in modern theoretical physics (see, for instance, [17]), and shall be employed in subsequent papers.

The outline of this paper is as follows. In the following section, the filtering problem is reformulated in a language common in physics. In Section 3, the path integral formula for the transition probability density is derived for the general additive noise case. The Yau algorithm is then derived in the following section. In Sections 5 and 6 some conceptual remarks and numerical examples are presented. The conclusions are presented in Section 7. In Appendix 1, aspects of continuous-continuous filtering are reviewed.

For more details on the path integral methods, see any modern text on quantum field theory, such as [17], and especially [18] which discusses application of FPI to the study of stochastic processes.

2 The continuous filtering problem: a physical reformulation

In this section, the filtering problem is stated in a language commonly used in theoretical physics.

2.1 Langevin Equation

Consider an ensemble of systems with state variables described by the Langevin equation:

$$\dot{\mathbf{x}}(t) = f(\mathbf{x}(t), t) + e(\mathbf{x}(t), t)\mathbf{v}(t), \quad \mathbf{x}(0) = \mathbf{x}_0, \quad Q(t) \in \mathbb{R}^{p \times p}. \tag{1}$$

Here, $\mathbf{x}(t) \in \mathbb{R}^n$, the drift $f(\mathbf{x}(t), t) \in \mathbb{R}^n$, the diffusion vielbein $e(\mathbf{x}(t), t) \in \mathbb{R}^{n \times p}$, and $\mathbf{v}(t)$ is δ -correlated with covariance matrix $Q(t) \in \mathbb{R}^{p \times p}$. When the diffusion vielbein is independent of the

state $\mathbf{x}(t)$, the noise is termed additive. It is the additive noise that is studied here, since it enables the use of functional methods common in quantum field theory.

Due to the random noise, each system leads to a different vector $x(t)$ that depends on time. Although only one realization of the stochastic process is ever observed, it is meaningful to speak about an ensemble average. For fixed times $t = t_i, i = 1, 2, \dots, r$, the probability density of finding the random vector $\mathbf{x}(t)$ in the (n -dimensional) interval $x_i \leq \mathbf{x}(t_i) \leq x_i + dx_i (1 \leq i \leq r)$ is given by

$$W_r(t_r, x_r; \dots; t_1, x_1) = \left\langle \prod_{i=1}^r \delta^n(\mathbf{x}(t_i) - x_i) \right\rangle, \tag{2}$$

where x_i is an n -dimensional column vector and $\langle \cdot \rangle$ denotes averaging with respect to the signal model noise $\nu(t)$. The complete information on the random vector $\mathbf{x}(t)$ is contained in the infinite hierarchy of such probability densities. The quantity of interest here is the conditional probability density

$$\begin{aligned} p(t_r, x_r | t_{r-1}, x_{r-1}, \dots; t_1, x_1) &= \langle \delta^n(\mathbf{x}(t_r)) \rangle |_{x(t_{r-1})=x_{r-1}, \dots, x(t_1)=x_1}, \quad x(t_r) \equiv x_r \\ &= \frac{W_r(t_r, x_r; \dots; t_1, x_1)}{\int W_n(t_r, x_r; \dots; t_1, x_1) \{d^n x_r\}}. \end{aligned} \tag{3}$$

Now the process described by the Langevin equation with δ -correlated Langevin force is a Markov process, i.e., the conditional probability density depends only on the value at the immediate previous time:

$$p(t_n, x_n | t_{n-1}, x_{n-1}; \dots; t_1, x_1) = p(t_n, x_n | t_{n-1}, x_{n-1}). \tag{4}$$

It can be shown that the transition probability satisfies the Fokker-Planck-Kolmogorov forward equation (FPKfe) (see, for instance, [19])

$$\begin{aligned} \frac{\partial p}{\partial t}(t, x) &= -\sum_{i=1}^n \frac{\partial}{\partial x_i} [f_i(x(t), t)p(t, x)] + \frac{1}{2} \sum_{i,j=1}^n \sum_{a=1}^p \frac{\partial^2}{\partial x_i \partial x_j} [(e(x(t), t)Q(t))_{ia} e_{aj}^T(x(t), t)p(t, x)], \\ &\equiv \mathcal{L}_p(t, x). \end{aligned} \tag{5}$$

Finally, the Gaussian noise process $\nu(t)$ can be represented by the following path integral measure

$$[d\rho(\nu(t))] = [D\nu(t)] \exp \left[-\frac{1}{2} \sum_{a,b=1}^p \int \nu_a(t) [Q^{-1}(t)]_{ab} \nu_b(t) dt \right], \quad \nu \in \mathbb{R}^p. \tag{6}$$

where $\nu(t)$ is a real vector for each t . This leads to a configuration space FPI formula for the fundamental solution of the FPKfe (see, for instance, [18]). The path integral formula for the fundamental solution for the FPKfe is applied to the continuous-discrete filtering problem with additive (state model) noise in [20,21].

2.2 The Continuous-Continuous Filtering Problem

Similarly, the continuous-continuous model can be written as follows:

$$\begin{cases} \dot{\mathbf{x}}(t) = f(\mathbf{x}(t), t) + e(\mathbf{x}(t), t)\mathbf{v}(t), & \mathbf{x}(0) = \mathbf{x}_0, & Q(t) \in \mathbb{R}^{p \times p} \\ \dot{\mathbf{y}}(t) = h(\mathbf{x}(t), t) + n(\mathbf{x}(t), t)\boldsymbol{\mu}(t), & \mathbf{y}(0) = \mathbf{y}_0, & R(t) \in \mathbb{R}^{q \times q} \end{cases} \quad (7)$$

Here, $\mathbf{y}(t) \in \mathbb{R}^m$, the measurement model drift $h(\mathbf{x}(t), t) \in \mathbb{R}^m$, the diffusion vielbein $n(\mathbf{x}(t), t) \in \mathbb{R}^{m \times q}$, and $\boldsymbol{\mu}(t)$ is δ -correlated with covariance matrix $W(t) \in \mathbb{R}^{q \times q}$.

Thus, in continuous-continuous filtering, the continuous-time measurement stochastic process needs to be incorporated as well. Consider another ensemble of systems with state variables whose time evolution is governed by the measurement process. The measurement noise means that each system in the ensemble leads to a different time-dependent vector $\gamma(t)$. Thus, even though only one realization of the measurement stochastic process is observed, it is still meaningful to talk about an ensemble average of the measurement process (in addition to one over the state process). Thus, the quantity of interest in continuous-continuous filtering is the *conditional probability density*

$$P(t_2, \mathbf{x}_2; \gamma_2 | t_1, \mathbf{x}_1; \gamma_1) = \langle\langle \delta^n(\mathbf{x}(t_2) - \mathbf{x}_2) \delta^m(\mathbf{y}(t_2) - \gamma_2) \rangle\rangle_{\boldsymbol{\mu}} |_{\mathbf{x}(t_1)=\mathbf{x}_1, \mathbf{y}(t_1)=\gamma_1}, \quad (8)$$

where $\langle \cdot \rangle_{\boldsymbol{\mu}}$ denotes averaging with respect to the measurement noise $\boldsymbol{\mu}(t)$. A crucial difference between the state and measurement stochastic process is that, unlike the state, the measurement samples are known.

Note that the conditional transition probability density is the complete solution to the continuous-continuous filtering problem, since if the initial distribution is $u(t_{i-1}, \mathbf{x}' | Y_{i-1})$, where Y_{i-1} is the set of all measurements prior to t_{i-1} , then the evolved conditional probability distribution is

$$u(t, \mathbf{x} | Y_i) = \int P(t, \mathbf{x}; \gamma_i | t_{i-1}, \mathbf{x}'; \gamma_{i-1}) u(t_{i-1}, \mathbf{x}' | Y_{i-1}) \{d^n \mathbf{x}'\}. \quad (9)$$

In the following sections, the path integral formulas for $P(t_2, \mathbf{x}_2; \gamma_2 | t_1, \mathbf{x}_1; \gamma_1)$ are derived. In Section 4 it shall be shown that it leads to the Yau algorithm. It shall be shown that the YYe plays the same role here that the FPKfe does in continuous-discrete filtering.

3 Path integral formula for the conditional transition probability density

The path integral formula for the conditional transition probability density shall now be derived using functional methods. Note that implicit in the use of these formal functional methods is the use of the Feynman convention, or symmetric discretization for the drift.

As noted in Section 2, the transition probability density is computed by averaging over the signal and measurement ensembles, i.e.,

$$\begin{aligned}
 P(t, x; \gamma | t_0, x_0; \gamma_0) &= \int [d\rho(v(t, t_0))] [d\rho(\mu(t, t_0))] \\
 &\times [\mathcal{D}\dot{x}(t)] J \delta^n[\dot{x}(t) - f(x(t), t) - e(t)v(t)] \\
 &\times [\mathcal{D}\dot{y}(t)] J_\gamma \delta^m[\dot{y}(t) - h(x(t), t) - n(t)\mu(t)] \\
 &\times \delta^n[x(t) - x] |_{x(t_0)=x_0} \delta^m[y(t) - \gamma] |_{y(t_0)=\gamma_0} .
 \end{aligned} \tag{10}$$

From the assumptions of the signal and measurement noise processes, it is evident that

$$[d\rho(v(t, t_0))] = [\mathcal{D}v(t, t_0)] \exp\left(-\frac{1}{2} \sum_{a,b=1}^p \int_{t_0}^t v_a(t) [Q^{-1}(t)]_{ab} v_b(t) dt\right), \tag{11}$$

$$[d\rho(\mu(t, t_0))] = [\mathcal{D}\mu(t, t_0)] \exp\left(-\frac{1}{2} \sum_{c,d=1}^p \int_{t_0}^t \mu_c(t) [W^{-1}(t)]_{cd} \mu_d(t) dt\right). \tag{12}$$

The Jacobian J follows from the functional derivative of the Langevin equation:

$$\begin{aligned}
 \frac{\delta}{\delta x_j(t')} \left[\dot{x}_i(t) - f_i(x(t), t) - \sum_{a=1}^p e_{ia}(t) v_a(t) \right] &= \left[\delta_{ij} \frac{d}{dt} - \frac{\partial f_i}{\partial x_j}(x(t), t) \right] \delta(t - t'), \\
 &= -\frac{d}{dt'} \left[\delta_{ij} \delta(t - t') - \theta(t - t') \frac{\partial f_i}{\partial x_j}(x(t), t) \right].
 \end{aligned} \tag{13}$$

Hence,

$$\begin{aligned}
 J &= \det\left(-\frac{d}{dt'}\right) \det\left(\delta_{ij} \delta(t - t') - \theta(t - t') \frac{\partial f_i}{\partial x_j}(x(t), t)\right), \\
 &= \mathcal{N} \det\left(\delta_{ij} \delta(t - t') - \theta(t - t') \frac{\partial f_i}{\partial x_j}(x(t), t)\right),
 \end{aligned} \tag{14}$$

where \mathcal{N} is an irrelevant constant, or,

$$\begin{aligned} \ln J &= \ln \det \left[\delta_{ij} \delta(t-t') - \theta(t-t') \frac{\partial f_i}{\partial x_j}(x(t), t) \right], \\ &= -\frac{1}{2} \sum_{i=1}^n \int \frac{\partial f_i}{\partial x_i}(x(t), t) dt. \end{aligned} \tag{15}$$

The Jacobian J_γ is trivial (as the measurement model drift is γ -independent) and can be absorbed into the measure.

It is noteworthy that J is not trivial. In quantum field theory, nontrivial Jacobians usually imply that there is an anomaly, as in the case of chiral anomalies in gauge theories. However, there is no reason for an anomaly here; after all, this is not even a quantum field theoretical system. The puzzle is resolved by noting that path integral anomalies in quantum field theory arise from the "multiplicative" part in the change of variables (i.e., $\psi(x) \rightarrow U(x) \psi(x)$). In contrast, the nontrivial Jacobian term here arises from the additive term; the multiplicative term does not contribute to the Jacobian, in accordance with expectations.

Thus, so far,

$$\begin{aligned} P(t, x; \gamma | t_0, x_0; \gamma_0) &= \int_{\gamma(t_0)=\gamma_0}^{\gamma(t)=\gamma} \int_{x(t_0)=x_0}^{x(t)=x} [\mathcal{D}x(t)] [\mathcal{D}\gamma(t)] [d\rho(v(t, t_0))] [d\rho(\mu(t, t_0))] \\ &\quad \times \delta^n \left(\dot{x}_i(t) - f_i(x(t), t) - \sum_{a=1}^p e_{ia}(t) v_a(t) \right) \exp \left(-\frac{1}{2} \sum_{i=1}^n \int_{t_0}^t \frac{\partial f_i}{\partial x_i}(x(t), t) dt \right) \\ &\quad \times \delta^m \left(\dot{y}_k(t) - h_k(x(t), t) - \sum_{c=1}^q n_{kc}(t) \mu_c(t) \right). \end{aligned} \tag{16}$$

Using the Fourier integral version of the delta function

$$\begin{aligned}
 P(t, x; \gamma | t_0, x_0; \gamma_0) &= \int_{\gamma(t_0)=\gamma_0}^{\gamma(t)=\gamma} \int_{x(t_0)=x_0}^{x(t)=x} [\mathcal{D}x(t)][\mathcal{D}\gamma(t)][\mathcal{D}v(t, t_0)][\mathcal{D}\mu(t, t_0)] \\
 &\quad [\mathcal{D}\lambda(t, t_0)] \exp \left(i \sum_{i=1}^n \int_{t_0}^t \lambda_i(t) \left[\dot{x}_i(t) - f_i(x(t), t) - \sum_{a=1}^p e_{ia}(t) v_a(t) \right] dt \right) \\
 &\quad [\mathcal{D}\kappa(t, t_0)] \exp \left(i \sum_{k=1}^n \int_{t_0}^t \kappa_k(t) \left[\dot{y}_k(t) - h_k(x(t), t) - \sum_{c=1}^q n_{kc}(t) \mu_c(t) \right] dt \right) \\
 &\quad \times \exp \left(-\frac{1}{2} \sum_{a,b=1}^p \int_{t_0}^t v_a(t) (Q^{-1}(t))_{ab} v_b(t) dt \right) \times \exp \left(-\frac{1}{2} \sum_{i=1}^n \int_{t_0}^t \frac{\partial f_i}{\partial x_i}(x(t), t) dt \right) \\
 &\quad \times \exp \left(-\frac{1}{2} \sum_{c,d=1}^q \int_{t_0}^t \mu_c(t) (W^{-1}(t))_{cd} \mu_d(t) dt \right).
 \end{aligned} \tag{17}$$

Integrating over $v(t, t_0)$, and $\mu(t, t_0)$ leads to

$$\begin{aligned}
 P(t, x; \gamma | t_0, x_0; \gamma_0) &= \int_{\gamma(t_0)=\gamma_0}^{\gamma(t)=\gamma} \int_{x(t_0)=x_0}^{x(t)=x} [\mathcal{D}x(t)][\mathcal{D}\gamma(t)][\mathcal{D}\lambda(t, t_0)][\mathcal{D}\kappa(t, t_0)] \exp \left(-\frac{1}{2} \sum_{i=1}^n \int \frac{\partial f_i}{\partial x_i}(x(t), t) dt \right) \\
 &\quad \times \exp \left(-\int_{t_0}^t \left[\sum_{i,j=1}^n \sum_{a,b=1}^p \lambda_i(t) e_{ia}(t) Q_{ab}(t) e_{bj}^T(t) \lambda_j(t) - i \sum_{i=1}^n \lambda_i(t) (\dot{x}_i(t) - f_i(x(t), t)) \right] dt \right) \\
 &\quad \times \exp \left(-\int_{t_0}^t \left[\sum_{k,l=1}^m \sum_{c,d=1}^q \kappa_k(t) n_{kc}(t) W_{cd}(t) n_{dl}^T(t) \kappa_l(t) - i \sum_{k=1}^m \kappa_k(t) (\dot{y}_k(t) - h_k(x(t), t)) \right] dt \right)
 \end{aligned} \tag{18}$$

Integrating over $\lambda(t, t_0)$ and $\kappa(t, t_0)$, it is clear that

$$P(t, x; \gamma | t_0, x_0; \gamma_0) = \int_{\gamma(t_0)=\gamma_0}^{\gamma(t)=\gamma} \int_{x(t_0)=x_0}^{x(t)=x} [\mathcal{D}x(t)][\mathcal{D}\gamma(t)] \exp[-S(t_0, t)], \tag{19}$$

where the action is given by

$$\begin{aligned}
 S(t_0, t) &= \frac{1}{2} \sum_{i,j=1}^n \sum_{a,b=1}^p \int_{t_0}^t [\dot{x}_i(t) - f_i(x(t), t)] (e_{ia}(t) Q_{ab}(t) e_{bj}^T(t))^{-1} [\dot{x}_j(t) - f_j(x(t), t)] dt \\
 &\quad + \frac{1}{2} \int_{t_0}^t \left[\sum_{k,l=1}^m \sum_{c,d=1}^q [\dot{y}_k(t) - h_k(x(t), t)] (n_{kc}(t) W_{cd}(t) n_{dl}^T(t))^{-1} [\dot{y}_l(t) - h_l(x(t), t)] + \sum_{i=1}^n \frac{\partial f_i}{\partial x_i}(x(t), t) \right] dt.
 \end{aligned} \tag{20}$$

4 Derivation of the Yau algorithm

Observe that the path integral formula derived in the previous section is over both the state and measurement variables. In this section, we shall show that in some cases it is possible to reduce

the result to a path integral over the state variables only. This has the advantage of being pre-computable since it is independent of measurements. It shall be shown to lead to the Yau algorithm.

4.1 Sampled Continuous Measurements

Suppose measurements are available at time t_{i-1} and t_i , and that the conditional transition probability density at time t_i is desired. Further, assume that there are no measurements available between t_{i-1} and t_i . The general path integral formula (Equation 19) cannot be simplified unless some additional assumptions are made.

First, consider the case where $e(t)Q(t)e^T(t)$ and $n(t)W(t)n^T(t)$ are $\hbar_\nu I_{n \times n}$ and $\hbar_\mu I_{m \times m}$ and $h(x(t))$ is not explicitly time dependent. Then, the contribution to the action due to the measurement process is

$$-\frac{1}{2\hbar\mu} \sum_{k=1}^m \int_{t_{i-1}}^{t_i} dt [\dot{y}_k - h_k(x(t))]^2 = -\frac{1}{2\hbar\mu} \sum_{k=1}^m \int_{t_{i-1}}^{t_i} dt [\dot{y}_k^2(t) + h_k^2(x(t)) - 2h_k(x(t))\dot{y}_k(t)]. \tag{21}$$

The quantity of interest is the state and we seek to integrate over the measurement variables.

Now the first term is independent of the state variables. The second term can be added to the action term that is independent of $y(t)$. It remains to investigate the contribution of the third term:

$$\frac{1}{\hbar\mu} \sum_{k=1}^m \int_{t_{i-1}}^{t_i} h_k(x(t))\dot{y}_k(t)dt. \tag{22}$$

There are two issues in this evaluation. Firstly, this can be evaluated via the usual integration by parts, but it is important to note that it is valid only for symmetric discretization. Secondly, since the measurements are sampled, there are two possible interpretations when $t_i - t_{i-1} = \epsilon$, where ϵ is an infinitesimal:

$$\frac{1}{\hbar\mu} \sum_{k=1}^m \int_{t_{i-1}}^{t_i} h_k(x(t))\dot{y}_k(t)dt = \begin{cases} \frac{1}{\hbar\mu} \sum_{k=1}^m \int_{t_{i-1}}^{t_i} h_k(x(t))\dot{y}_k(t)dt, \\ \frac{1}{\hbar\mu} \sum_{k=1}^m \int_{t_{i-1}-\epsilon}^{t_i-\epsilon} h_k(x(t))\dot{y}_k(t)dt. \end{cases} \tag{23}$$

This leads to two possibilities:

$$\frac{1}{\hbar \mu} \sum_{k=1}^m \int_{t_{i-1}}^{t_i} h_k(x(t)) \dot{\gamma}_k(t) dt = \begin{cases} \frac{1}{\hbar \mu} \sum_{k=1}^m h_k \left(\frac{1}{2} [x(t_i) + x(t_{i-1})] \right) [\gamma_k(t_i) - \gamma_k(t_{i-1})], \\ \frac{1}{\hbar \mu} \sum_{k=1}^m h_k \left(\frac{1}{2} [x(t_{i-1}) + x(t_{i-2})] \right) [\gamma_k(t_{i-1}) - \gamma_k(t_{i-2})]. \end{cases} \quad (24)$$

Finally, the residual Gaussian path integral over $\gamma(t)$ can be ignored as it is independent of the state. Therefore, the path integral formula simplifies to

$$P(t_i, x_i; \gamma_i | t_{i-1}, x_{i-1}; \gamma_{i-1}) = \tilde{P}(t_i, x_i | t_{i-1}, x_{i-1}) \times \begin{cases} \exp \left(\frac{1}{\hbar \mu} \sum_{k=1}^m h_k \left(\frac{1}{2} [x(t_i) + x(t_{i-1})] \right) [\gamma_k(t_i) - \gamma_k(t_{i-1})] \right), \\ \exp \left(\frac{1}{\hbar \mu} \sum_{k=1}^m h_k \left(\frac{1}{2} [x(t_{i-1}) + x(t_{i-2})] \right) [\gamma_k(t_{i-1}) - \gamma_k(t_{i-2})] \right), \end{cases} \quad (25)$$

where

$$\tilde{P}(t_i, x_i | t_{i-1}, x_{i-1}) = \int_{x(t_{i-1})=x_{i-1}}^{x(t_i)=x_i} [Dx(t)] \exp(-S(t_{i-1}, t_i)), \quad (26)$$

and

$$S(t_{i-1}, t_i) = \frac{1}{2} \int_{t_{i-1}}^{t_i} dt \left[\frac{1}{\hbar \nu} \sum_{i=1}^n [\dot{x}_i(t) - f_i(x(t))]^2 + \sum_{i=1}^n \frac{\partial f_i}{\partial x_i}(x(t)) + \frac{1}{\hbar \mu} \sum_{k=1}^m h_k^2(x(t)) \right]. \quad (27)$$

Secondly, if the conditions above are relaxed to allowing explicit time dependence of the drift term in the state model, i.e., $f(x(t), t)$, and $C = (n(t)W(t)n^T(t))$, then the path integral formula becomes

$$P(t_i, x_i; \gamma_i | t_{i-1}, x_{i-1}; \gamma_{i-1}) = \tilde{P}(t_i, x_i | t_{i-1}, x_{i-1}) \times \begin{cases} \exp \left(\frac{1}{\hbar \mu} \sum_{k,l=1}^m h_k \left(\frac{1}{2} [x(t_i) + x(t_{i-1})] \right) (C^{-1})_{kl} [\gamma_l(t_i) - \gamma_l(t_{i-1})] \right), \\ \exp \left(\frac{1}{\hbar \mu} \sum_{k,l=1}^m h_k \left(\frac{1}{2} [x(t_{i-1}) + x(t_{i-2})] \right) (C^{-1})_{kl} [\gamma_l(t_{i-1}) - \gamma_l(t_{i-2})] \right), \end{cases} \quad (28)$$

where

$$\tilde{P}(t_i, x_i | t_{i-1}, x_{i-1}) = \int_{x(t_{i-1})=x_{i-1}}^{x(t_i)=x_i} [\mathcal{D}x(t)] \exp(-S(t_{i-1}, t_i)), \tag{29}$$

and the action is given by

$$S(t_{i-1}, t_i) = \frac{1}{2} \int_{t_{i-1}}^{t_i} dt \left[\frac{1}{\hbar_{\mathbf{v}}} \sum_{i=1}^n [\dot{x}_i(t) - f_i(x(t), t)]^2 + \sum_{i=1}^n \frac{\partial f_i}{\partial x_i}(x(t), t) + \sum_{k,l=1}^m h_k(x(t))(C^{-1})_{kl} h_l(x(t)) \right]. \tag{30}$$

Finally, consider the case when $e(t)Q(t)e(t)$ is time-independent and invertible, but otherwise arbitrary, and $C = (n(t)W(t)n^T(t))$. Note that C is a constant, symmetric matrix and assumed to be invertible. The path integral formula is given by Equations 28 and 29 and with the action $S(t_{i-1}, t_i)$ is given by

$$S(t_{i-1}, t_i) = \frac{1}{2} \int_{t_{i-1}}^{t_i} dt \sum_{i,j=1}^n \sum_{a,b=1}^p [\dot{x}_i(t) - f_i(x(t), t)] [e_{ia}(t)Q_{ab}(t)e_{bj}^T(t)]^{-1} [\dot{x}_j(t) - f_j(x(t), t)] + \frac{1}{2} \int_{t_{i-1}}^{t_i} dt \left[\sum_{k,l=1}^m h_k(x(t))(C^{-1})_{kl} h_l(x(t)) + \sum_{i=1}^n \frac{\partial f_i}{\partial x_i}(x(t), t) \right]. \tag{31}$$

4.2 The Yau Algorithm

In Section 2 it was noted that if $v_{i-1}(t_{i-1}, x_{i-1})$ is the conditional probability density at time t_{i-1} , then the conditional probability density at time t_i is given by

$$v_i(t_i, x_i) = \int P(t_i, x_i; \gamma_i | t_{i-1}, x_{i-1}; \gamma_{i-1}) v_{i-1}(t_{i-1}, x_{i-1}) \{d^n x_{i-1}\}. \tag{32}$$

For simplicity, let us first consider the case $e(t)Q(t)e^T(t)$ and $n(t)W(t)n^T(t)$ are $\hbar_{\mathbf{v}} I_{n \times n}$ and $\hbar_{\mu} I_{m \times m}$, and $h(x(t))$ is not explicitly time dependent. Then

$$v_i(t_i, x_i) = \left\{ \int \tilde{P}(t_i, x_i | t_{i-1}, x_{i-1}) \exp\left(-\frac{1}{\hbar_{\mathbf{v}}} \sum_{k=1}^m h_k\left(\frac{1}{2}[x(t_i) + x(t_{i-1})]\right) [\gamma_k(t_i) - \gamma_k(t_{i-1})]\right) v_{i-1}(t_{i-1}, x_{i-1}) \{d^n x_{i-1}\}, \right. \\ \left. \int \tilde{P}(t_i, x_i | t_{i-1}, x_{i-1}) \exp\left(-\frac{1}{\hbar_{\mathbf{v}}} \sum_{k=1}^m h_k\left(\frac{1}{2}[x(t_{i-1}) + x(t_{i-2})]\right) [\gamma_k(t_{i-1}) - \gamma_k(t_{i-2})]\right) v_{i-1}(t_{i-1}, x_{i-1}) \{d^n x_{i-1}\}. \right. \tag{33}$$

When $t_i - t_{i-1}$ is small

$$h_k \left(\frac{1}{2} [x(t_i) + x(t_{i-1})] \right) \approx h_k(x(t_{i-1})) + O(\sqrt{\tau}), \tag{34}$$

and Equation 33 becomes

$$v_i = (t_i, x_i) = \begin{cases} \int \tilde{P}(t_i, x_i | t_{i-1}, x_{i-1}) \exp \left(-\frac{1}{\hbar_{\mathbf{v}}} \sum_{k=1}^m h_k(x(t_{i-1})) [\gamma_k(t_i) - \gamma_k(t_{i-1})] \right) v_{i-1}(t_{i-1}, x_{i-1}) \{d^n x_{i-1}\}, \\ \int \tilde{P}(t_i, x_i | t_{i-1}, x_{i-1}) \exp \left(-\frac{1}{\hbar_{\mathbf{v}}} \sum_{k=1}^m h_k(x(t_{i-1})) [\gamma_k(t_{i-1}) - \gamma_k(t_{i-2})] \right) v_{i-1}(t_{i-1}, x_{i-1}) \{d^n x_{i-1}\}. \end{cases} \tag{35}$$

Following the method originally used by Feynman, the partial differential equation satisfied by $\tilde{P}(t, x | t_0, x_0)$ may be derived (see, for instance, [20]). In particular, $\tilde{P}(t_i, x_i | t_{i-1}, x_{i-1})$ is the fundamental solution of the Yau Equation (YYe):

$$\begin{cases} \frac{\partial \tilde{P}}{\partial t}(t, x | t_0, x_0) = \frac{\hbar_{\mathbf{v}}}{2} \sum_{i=1}^n \frac{\partial^2 \tilde{P}}{\partial x_i^2}(t, x | t_0, x_0) - \sum_{i=1}^n \frac{\partial}{\partial x_i} [f_i(x) \tilde{P}(t, x | t_0, x_0)] - \frac{1}{2\hbar_{\mu}} \sum_{k=1}^m h_k^2(x) \tilde{P}(t, x | t_0, x_0), \\ \tilde{P}(t_0, x | t_0, x_0) = \delta^n(x - x_0). \end{cases} \tag{36}$$

This implies that $v_i(t_i, x)$ is the solution at t_i of

$$\begin{cases} \frac{\partial v_i}{\partial t}(t, x) = \frac{1}{2} \sum_{l=1}^n \frac{\partial^2 v_i}{\partial x_l^2}(t, x) - \sum_{l=1}^n f_l(x) \frac{\partial v_i}{\partial x_l}(t, x) - \left(\sum_{l=1}^n \frac{\partial f_l}{\partial x_l}(x) + \frac{1}{2} \sum_{l=1}^m h_l^2(x) \right) v_i(t, x), \\ v_i(t_{i-1}, x) = v_{i-1}(t_{i-1}, x) \begin{cases} \exp \left(\sum_{j=1}^m h_j(x) [\gamma_j(t_i) - \gamma_j(t_{i-1})] \right), \\ \exp \left(\sum_{j=1}^m h_j(x) [\gamma_j(t_{i-1}) - \gamma_j(t_{i-2})] \right). \end{cases} \end{cases} \tag{37}$$

This is precisely the Yau algorithm.

Likewise, it is straightforward to see that for the general case studied in Section 4.1, the Yau algorithm is extended to this case as follows: $v_i(t_i, x_i)$ is the solution at t_i of the PDE

$$\left\{ \begin{aligned} \frac{\partial v_i}{\partial t}(t, x) &= \frac{1}{2} \sum_{l,j=1}^n \sum_{a,b=1}^p \left(e_{la}(t) Q_{ab}(t) e_{bj}^T(t) \right) \frac{\partial^2 v_i}{\partial x_l \partial x_j}(t, x) \\ &\quad - \sum_{l=1}^n \frac{\partial}{\partial x_l} [f_l(x, t) v_i(t, x)] - \frac{1}{2} \sum_{k,l=1}^m h_k(x) (C^{-1})_{kl} h_l(x(t)) v_i(t, x), \\ v_i(t_{i-1}, x) &= v_{i-1}(t_{i-1}, x) \begin{cases} \exp \left(\sum_{j,k=1}^m h_j(x) (C^{-1})_{jk} [\gamma_k(t_i) - \gamma_k(t_{i-1})] \right), \\ \exp \left(\sum_{j,k=1}^m h_j(x) (C^{-1})_{jk} [\gamma_k(t_{i-1}) - \gamma_k(t_{i-2})] \right). \end{cases} \end{aligned} \right. \quad (38)$$

This is a straightforward generalization of the YYe to the state model with explicit time dependence.

5 Some remarks

Following are some remarks on the FPI solution of the filtering problem:

- Note that the FPI formulation has given a complete and self-contained solution of the continuous-continuous filtering problem. For instance, the DMZ equation or its variants were not used as an input. On the contrary, the FPI formula naturally led us to the YYe and the Yau algorithm.

- Note also that the DMZ equation (and variants) cannot be solved reliably in real-time as the various approximations assume that drift is bounded and require solution of a stochastic PDE that depend on measurements. In contrast, the general FPI formula presented in Section 3 can potentially be used for an efficient and reliable real-time solution. This is because the measurement time interval is usually small so that the simplest approximation of the path integral (termed the Dirac-Feynman approximation, see Section 6) is adequate. In contrast, in quantum mechanics and quantum field theory one is interested in the large time case.

- Unlike the Yau algorithm, the PI formula is valid even for the general time-dependent case for large measurement time interval. In other words, one can compute the conditional transition probability density using the conventional methods (see, for instance, [22]).

- The YYe can be viewed as a local expression of the path integral formula. That is, a path integral is a global object, while the PDE is a local one.

- In this paper, the signal noise and measurement noise are assumed to be additive. This is a stronger condition than the orthogonality of the diffusion vielbein assumed in [13]. The solution for the general case has been presented in [16].

- It is noted that other algorithms can also be solved using the FPI formulas with obvious changes. Usually, they require the solution of the FPKfe, which corresponds to the case $h(x) = 0$. Note that the FPKfe arises naturally in the solution of the continuous-discrete filtering problem [20]. Of course, as noted in Appendix 1, it would be unnecessary since the Yau algorithm has the best numerical properties. What is interesting to note is that the path integral formula naturally leads to the algorithm with the best numerical properties.

- The Yau algorithm also has the form of the "prediction" and "correction" part, as in continuous-discrete filtering [20]. Specifically, the prediction part is the solution of the YYe, whereas the correction part is the multiplicative factor in the initial condition. However, it is crucial to note that *the prediction part contains the measurement model drift*. In contrast, the measurement model plays no role in the prediction part in continuous-discrete filtering.

6 Examples

6.1 The Dirac-Feynman Approximation

From the discussion in the previous sections, it is evident that computing the path integral requires computing the Lagrangian, L , defined by $S = \int dt L(x, \dot{x}, t)$. The simplest (and crudest) approximation (for the case $e(t)Q(t)e^T(t) = \hbar_\nu I_{n \times n}$ and $n(t)W(t)n^T(t) = \hbar_\mu I_{m \times m}$) is to use the following approximation that is valid for infinitesimal time step:

$$\tilde{P}(t'', x'' | t', x') = \exp \left(-\frac{t''-t'}{2} \left\{ \frac{1}{\hbar_\nu} \sum_{i=1}^n \left[\frac{x''_i - x'_i}{t''-t'} - f_i \left(\frac{x''+x'}{2} \right) \right]^2 + \sum_{i=1}^n \frac{\partial f_i}{\partial x_i} \left(\frac{x''+x'}{2} \right) + \frac{1}{\hbar_\mu} \sum_{k=1}^m h_k^2 \left(\frac{x''+x'}{2} \right) \right\} \right). \tag{39}$$

This is the Dirac-Feynman (DF) approximation. The algorithm that follows from applying the DF approximation to the Yau algorithm is the Dirac-Feynman-Yau (DFY) algorithm.

6.2 Example 1

As an example, consider the following continuous-continuous filtering model that has been studied in [23] (and [24])

$$\begin{aligned} \dot{x}(t) &= a \cos(bx(t)) + \sigma_x \mathbf{v}(t), \\ \dot{y}(t) &= \tan^{-1} x(t) + \sigma_y \boldsymbol{\mu}(t). \end{aligned} \tag{40}$$

The Lagrangian for this model is given by

$$\frac{1}{2\sigma_x^2} (\dot{x}(t) - a \cos(bx(t)))^2 - \frac{ab}{2} \sin(bx(t)) + \frac{1}{2\sigma_y^2} (\tan^{-1} x(t))^2. \tag{41}$$

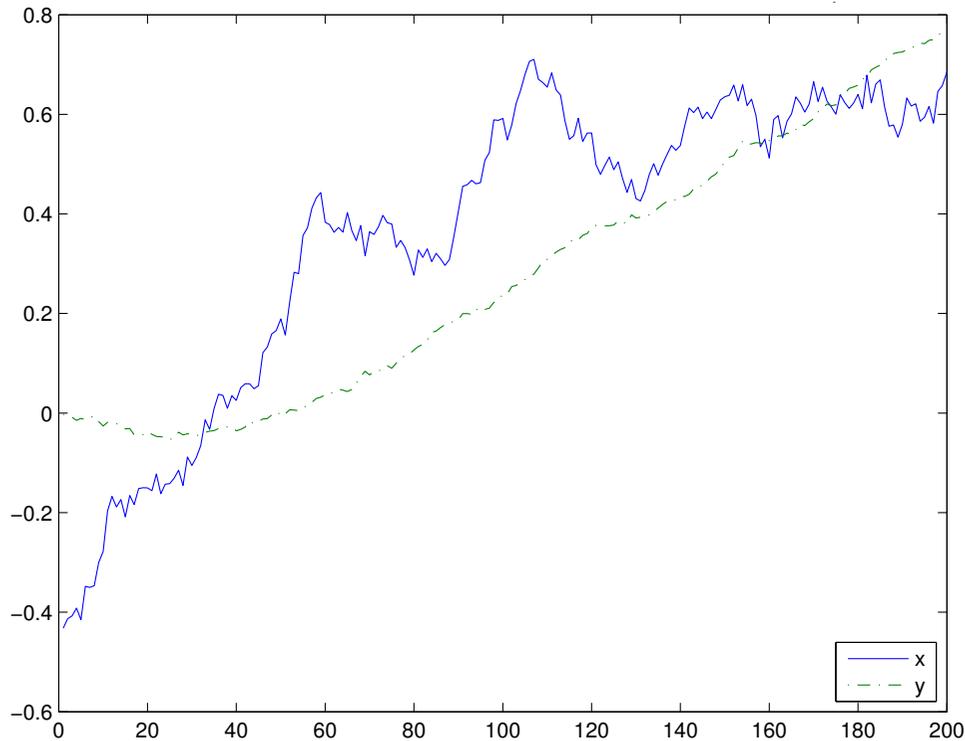


Figure 1
A sample of state and measurement processes given by Equation 40.

The parameters chosen were as in the reference. Specifically, $a = 1.2$, $b = 3$, $\sigma_x = 0.3$, spatial grid spacing $\Delta x = 0.01$ and extent $[-1.5, 1.5]$, temporal grid spacing $\Delta t = 0.01$ with 200 time steps.

In the first set, the measurement noise was set as $\sigma_y = 0.05$. Figure 1 shows a sample of state and measurement processes. The conditional mean, computed using the DFY algorithm, is plotted in Figure 2. Since there was negligible difference in performance between the pre-measurement and post-measurement forms, only the former was employed. Also plotted are 2σ limits. The conditional mean and variance were computed from the computed conditional probability density. The fact that the target was mostly within the 2σ limits of the conditional mean shows that the tracking performance of this algorithm is good.

In the next set, the measurement noise was set as $\sigma_y = 0.0125$. For this "small noise" case, most of the algorithms studied in [23] failed. In Figure 3 is plotted a sample of signal and measurement processes. The conditional mean and the 2σ limits computed using the DFY algorithm for this instance is plotted in Figure 4.

It is seen that good tracking performance is maintained for the small noise case even when the crudest path integral approximation is used in the Yau algorithm.

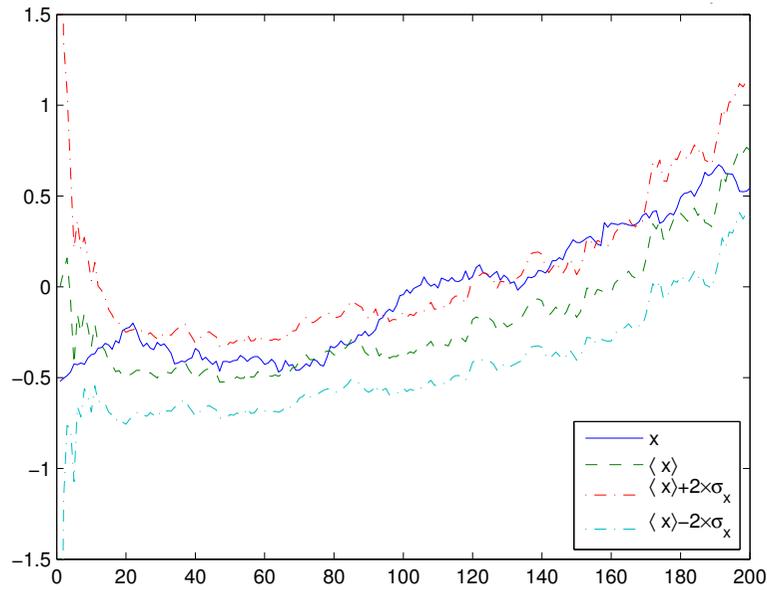


Figure 2
Conditional mean and 2σ limits computed using the DFY algorithm.

6.3 Example 2: Cubic Sensor Problem

The cubic sensor problem is defined by the following signal and measurement model:

$$\begin{aligned} \dot{x}(t) &= \sigma_x \mathbf{v}(t), \\ \dot{y}(t) &= x^3(t) + \sigma_y \boldsymbol{\mu}(t). \end{aligned} \tag{42}$$

It is a well-studied nonlinear filtering problem because it is one of the simplest examples of a filtering problem that is not finite dimensional (see, for instance, [25] and references therein).

For the simulation of the cubic sensor problem, the following model parameters were chosen (as in [25])

$$\begin{aligned} \Delta t &= 0.001, \\ \Delta x &= 0.01, \\ x_0 &= N(0, 0.01), \\ \sigma_x &= 1, \\ \sigma_y &= 1, \\ T &= [0, 20]. \end{aligned} \tag{43}$$

The EKF is a sub-optimal filter which approximates the conditional probability by a Gaussian. For the cubic sensor problem the EKF is given by

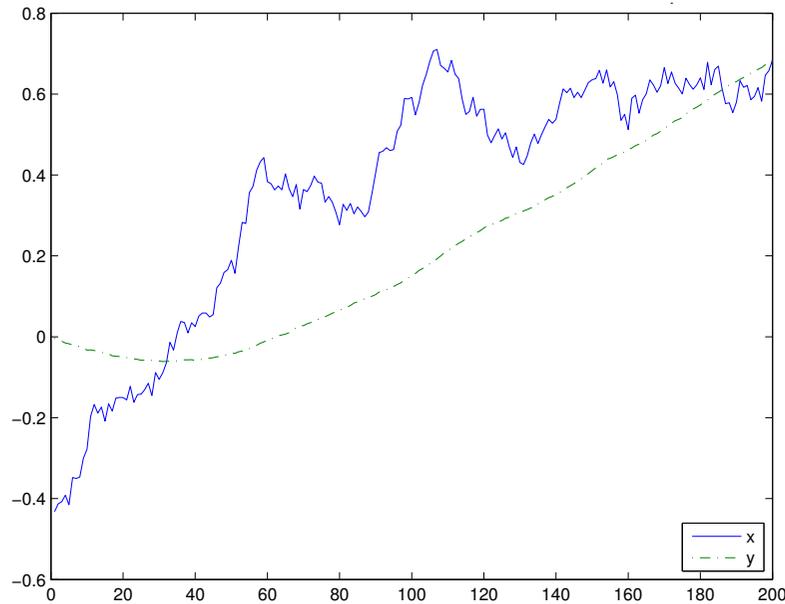


Figure 3
A sample of state and measurement processes given by Equation 40 – small measurement noise case.

$$\begin{aligned} d\hat{x} &= 3\hat{x}^2 P(dy - \hat{x}^3 dt), \\ dP &= (1 - 9\hat{x}^4 P^2)dt. \end{aligned} \tag{44}$$

The Yau Lagrangian for the cubic sensor problem is

$$L = \frac{1}{2\sigma_x^2} (\dot{x}^2) + \frac{1}{2\sigma_y^2} x^6. \tag{45}$$

The DF approximation the Yau Lagrangian is

$$\tilde{P}(t'', x'' | t', x') = \exp \left(-\frac{(t''-t')}{2\sigma_x^2} \left(\frac{x''-x'}{t''-t'} \right)^2 - \left(\frac{t''-t'}{2\sigma_y^2} \right) \left(\frac{x''+x'}{2} \right)^6 \right). \tag{46}$$

Figure 5 shows the performance of the DFY algorithm. Specifically, the conditional mean along with the two standard deviation bounds computed using the computed conditional probability density is plotted. Observe that the EKF fails completely in this case. As noted in [25], this is because the EKF considers only the first two moments (which vanish here); it is the fourth central moment that plays a crucial role in this example (for the chosen initial condition). Also, note that the state is within the 2σ region for most of the time. This shows that, unlike the EKF, the path integral filter has a reliable error analysis.

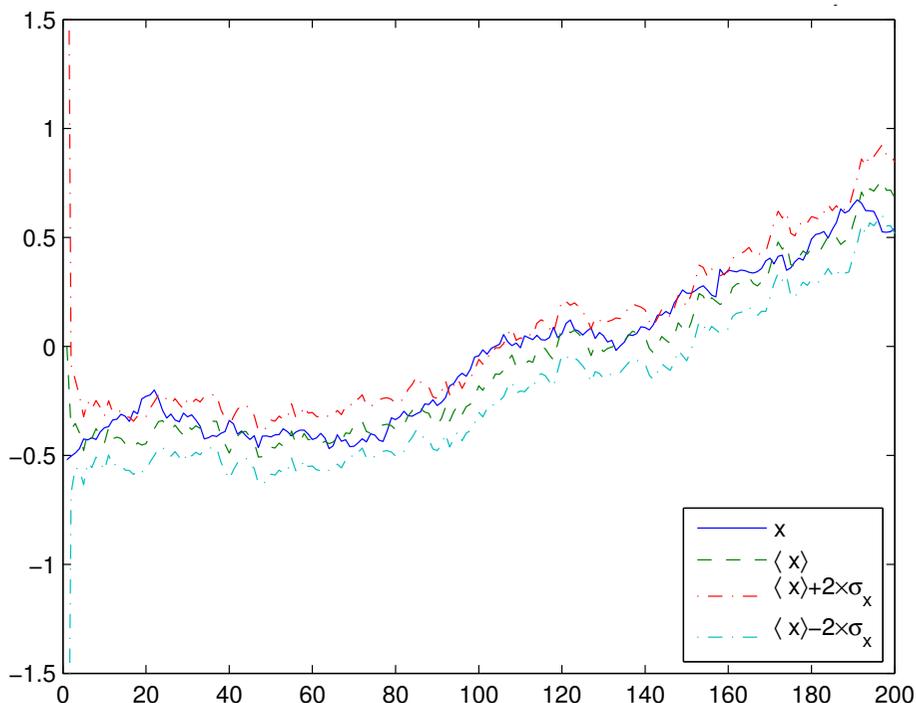


Figure 4
Conditional mean and 2σ limits computed using the DFY algorithm with small measurement noise.

After an initial period, the performance of the path integral approximation is seen to be excellent and comparable to that obtained using PDE methods in [25]. However, the crucial point is that the path integral solution is equally simple for the higher-dimensional case with more complicated models (e.g., colored noise), whereas a PDE solution would be significantly harder, if not impossible, to implement in real-time.

6.4 Comments

It is remarkable to note that very good performance is obtained using the crudest approximation. Of course, when the time step is large, it will fail (unlike the path integral formula itself). However, the practical situation is that the time steps are often small. Therefore, the DF approximation may be adequate in most cases.

The implementation of this method is trivial. The contrast with other methods, such as those studied in [23], is striking. For instance, many of those methods require off-line computation of complicated partial differential equations with uncertain numerical properties.

The results obtained in this paper used single time-step. More accuracy can be obtained quite simply using multiple time steps. Also, the computation of the transition probability density can be done off-line, but the on-line computation was not an onerous burden for the examples studied here.

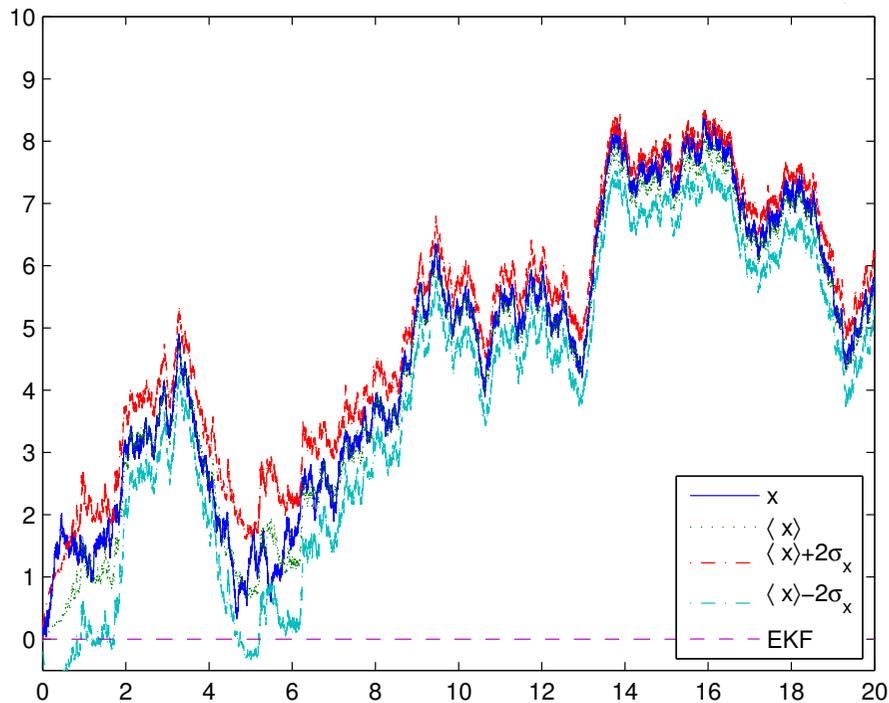


Figure 5
A simulation result for the cubic sensor problem.

It is important to note also that the transition probability density matrix (or tensor in the general case) is sparse (sparsity determined by h_u, h_v). This is of great importance in higher dimensional filtering problems because

- Sparse matrix storage requirements are small,
- The relevant transition probability density matrix elements can be computed based on the conditional density in the previous step, and
- Sparse matrix computations are very fast.

Note that unlike some other approximation techniques studied in [23], the conditional probability density is obviously always positive (provided, of course, that the initial distribution is positive).

Finally, a comment on measures of performance. Note that a good tracker is one that furnishes not only a good estimate of a state but also provides a reliable measure of the quality of the estimate. For the linear, Gaussian case, the conditional mean and the variance are Gaussian and the Kalman estimates are optimal and provide a complete description. However, for the general non-linear case, such a concise description is not possible. For instance, the conditional probability

density may be highly skewed. It may be multi-modal, in which case the conditional mean is not a very meaningful quantity. A more general measure is to indicate domains of "significant" probability mass; a good filtering solution is then one that guarantees that the state is in the region of significant probability mass with a very high confidence. For the purposes of the paper, the conditional variance was chosen.

7 Conclusion

In this paper, the formal path integral solution to the continuous-continuous nonlinear filtering problem has been presented. The solution is universal, i.e., the initial distribution may be arbitrary. Since the path integral measure is manifestly positive, positivity is maintained if the initial distribution is positive.

A path integral formulation has several advantages. It is well known that Feynman path integrals have led to theoretical insights in other areas including quantum mechanics, quantum field theory and even pure mathematics. It is demonstrated in this paper that it is possible to express the fundamental solution of the YYe in terms of Feynman path integrals. Finally, Feynman path integrals are very suitable for numerical implementation. Practical path integral filtering techniques, especially for solving large dimensional problems, will be presented in subsequent papers.

Appendix I continuous-continuous filtering and the Yau equation

In this section, the main results of (continuous-continuous) nonlinear filtering theory are summarized. For the general case (e.g., not the finite-dimensional filter case) and from a practical point of view the most important results are the YYe and the Yau algorithm.

A.1 The Continuous-Continuous Model

The signal and observation model in continuous-continuous filtering is the following:

$$\begin{cases} dx(t) = f(x(t), t)dt + e(x(t), t)dv(t), & x(0) = x_0, \\ dy(t) = h(x(t), t)dt + n(t)dw(t), & \gamma(0) = 0. \end{cases} \quad (47)$$

Here \mathbf{x} , \mathbf{v} , \mathbf{y} , and \mathbf{w} are \mathbb{R}^n , \mathbb{R}^p , \mathbb{R}^m and \mathbb{R}^q valued stochastic processes, respectively, and $e(\mathbf{x}(t), t) \in \mathbb{R}^{n \times p}$ and $n(t) \in \mathbb{R}^{m \times q}$. These are defined in the Itô sense. The Brownian processes \mathbf{v} and \mathbf{w} are assumed to be independent with $p \times p$ and $q \times q$ covariance matrices $Q(t)$ and $W(t)$, respectively. We denote $n(t)W(t)n^T(t)$ by $R(t)$, a $m \times m$ matrix. Also, f is referred to as the drift, e as the diffusion vielbein, and eQe^T as the diffusion matrix.

In this section, some of the relevant work on continuous-continuous filtering is summarized. Hence, it is assumed that $n = p$, $m = q$, f and h are vector-valued C^∞ smooth functions, $e(\mathbf{x}, t)$ is an

orthogonal matrix-valued C^∞ smooth function, $Q(t)$ is a $n \times n$ identity matrix, and $n(t)$ and $R(t)$ are $m \times m$ identity matrices. No explicit time dependence is assumed in the model.

A.2 The DMZ Stochastic Differential Equation

The unnormalized conditional probability density, $\sigma(t, x)$ of the state given the observations $\{Y(s): 0 \leq s \leq t\}$ satisfies the DMZ equation:

$$d\sigma(t, x) = \mathcal{L}_0 \sigma(t, x) dt + \sum_{i=1}^n \mathcal{L}_i \sigma(t, x) dy_i(t), \quad \sigma(0, x) = \sigma_0. \tag{48}$$

Here

$$\begin{aligned} \mathcal{L}_0 &= \frac{1}{2} \sum_{i=1}^n \frac{\partial^2}{\partial x_i^2} - \sum_{i=1}^n f_i(x) \frac{\partial}{\partial x_i} - \sum_{i=1}^n \frac{\partial f_i}{\partial x_i}(x) - \frac{1}{2} \sum_{i=1}^m h_i^2(x), \\ &= - \sum_{i=1}^n \frac{\partial}{\partial x_i} (f_i(x)) + \frac{1}{2} \sum_{i=1}^n \frac{\partial^2}{\partial x_i^2} - \frac{1}{2} \sum_{i=1}^m h_i^2(x), \\ &= \frac{1}{2} \left(\sum_{i=1}^n D_i^2 - \eta(x) \right), \end{aligned} \tag{49}$$

where \mathcal{L}_i is the zero-degree differential operator of multiplication by $h_i(x)$, $i = 1, \dots, m$, σ_0 is the probability density of the initial time t_0 , and

$$\begin{aligned} D_i &\equiv \frac{\partial}{\partial x_i} - f_i(x), \\ \eta(x) &\equiv \sum_{i=1}^n f_i^2(x) + \sum_{i=1}^n \frac{\partial f_i}{\partial x_i}(x) + \sum_{i=1}^m h_i^2(x). \end{aligned} \tag{50}$$

The DMZ equation is to be interpreted in the Stratanovich sense. Note that

$$D_i = \frac{\partial}{\partial x_i} - f_i(x), \tag{51}$$

$$= e^{F_i} \frac{\partial}{\partial x_i} e^{-F_i}, \quad \text{where } F_i = F_i(x) = \int_0^x f_i(t) dt. \tag{52}$$

Hence,

$$D_i^2 = e^{F_i} \frac{\partial^2}{\partial x_i^2} e^{-F_i}, \tag{53}$$

and

$$\mathcal{L}_0 = \frac{1}{2} \left(\sum_{i=1}^n e^{F_i} \frac{\partial^2}{\partial x_i^2} e^{-F_i} - \eta(x) \right). \tag{54}$$

A.3 The Robust DMZ Partial Differential Equation

Following [11] and [12] introduce a new unnormalized density

$$u(t, x) = \exp \left(- \sum_{i=1}^m h_i(x) \gamma_i(t) \right) \sigma(t, x). \tag{55}$$

Under this transformation, the DMZ SDE is transformed into the following time-varying PDE

$$\left\{ \begin{aligned} \frac{\partial u}{\partial t}(t, x) &= \frac{1}{2} \sum_{i=1}^n \frac{\partial^2 u}{\partial x_i^2}(t, u) + \sum_{i=1}^n \left(-f_i(x) + \sum_{j=1}^m \gamma_j(t) \frac{\partial h_j}{\partial x_i}(x) \right) \frac{\partial u}{\partial x_i}(t, x) \\ &\quad - \left(\sum_{i=1}^n \frac{\partial f_i}{\partial x_i}(x) + \frac{1}{2} \sum_{i=1}^m h_i^2(x) - \frac{1}{2} \sum_{i=1}^m \gamma_i(t) \Delta h_i(x) + \sum_{i=1}^m \sum_{j=1}^n \gamma_i(t) f_j(x) \frac{\partial h_i}{\partial x_j}(x) \right. \\ &\quad \left. - \frac{1}{2} \sum_{i,j=1}^m \sum_{k=1}^n \gamma_i(t) \gamma_j(t) \frac{\partial h_i}{\partial x_k}(x) \right) u(t, x), \\ u(0, x) &= \sigma_0(x). \end{aligned} \right. \tag{56}$$

This is called the robust DMZ equation. Here Δ is the Laplacian. The solution of a PDE when the initial condition is a delta function is called the fundamental solution.

A.4 The Yau Equation

Recently, it was proved that the real-time solution of the general nonlinear filtering problem can be obtained reliably [13,26]. Let $\mathcal{P} = \{ \tau_0 < \tau_1 < \dots < \tau_k = \tau \}$ be a partition of the time interval $[\tau_0, \tau]$, and let the norm of the partition \mathcal{P}_k be defined as $|\mathcal{P}_k| = \sup_{1 \leq i \leq k} \{ \tau_i - \tau_{i-1} \}$. If $u_i(t, x)$ satisfies the equation

$$\left\{ \begin{aligned} \frac{\partial u_l}{\partial t}(t, x) &= \frac{1}{2} \sum_{i=1}^n \frac{\partial^2 u_l}{\partial x_i^2}(t, x) + \sum_{i=1}^n \left(-f_i(x) + \sum_{j=1}^m \gamma_j(\tau_l) \frac{\partial h_j}{\partial x_i}(x) \right) \frac{\partial u_l}{\partial x_i}(t, x) \\ &\quad - \left(\sum_{i=1}^n \frac{\partial f_i}{\partial x_i}(x) + \frac{1}{2} \sum_{i=1}^m h_i^2(x) - \frac{1}{2} \sum_{i=1}^m \gamma_i(\tau_l) \Delta h_i(x) + \sum_{i=1}^m \sum_{j=1}^n \gamma_i(\tau_l) f_j(x) \frac{\partial h_i}{\partial x_j}(x) \right. \\ &\quad \left. - \frac{1}{2} \sum_{i,j=1}^m \sum_{k=1}^n \gamma_i(\tau_l) \gamma_j(\tau_l) \frac{\partial h_i}{\partial x_k}(x) \frac{\partial h_j}{\partial x_k}(x) \right) u_l(t, x), \\ u_l(\tau_{l-1}, x) &= u_{l-1}(\tau_{l-1}, x), \end{aligned} \right. \tag{57}$$

in the time interval $\tau_{l-1} \leq t \leq \tau_l$, then the function $\tilde{u}_l(t, x)$ defined as

$$\tilde{u}_l(t, x) = \exp \left(\sum_{i=1}^m \gamma_i(\tau_l) h_i(x) \right) u_l(t, x) \tag{58}$$

satisfies the parabolic partial differential equation

$$\begin{aligned} \frac{\partial \tilde{u}_l}{\partial t}(t, x) &= \frac{1}{2} \sum_{i=1}^n \frac{\partial^2 \tilde{u}_l}{\partial x_i^2}(t, x) - \sum_{i=1}^n f_i(x) \frac{\partial \tilde{u}_l}{\partial x_i}(t, x) - \left(\sum_{i=1}^n \frac{\partial f_i}{\partial x_i}(x) + \frac{1}{2} \sum_{i=1}^m h_i^2(x) \right) \tilde{u}_l(t, x), \\ &= - \sum_{i=1}^n \frac{\partial}{\partial x_i} [f_i(x) \tilde{u}_l(t, x)] + \frac{1}{2} \sum_{i=1}^n \frac{\partial^2 \tilde{u}_l}{\partial x_i^2}(t, x) - \frac{1}{2} \sum_{i=1}^m h_i^2(x) \tilde{u}_l(t, x), \end{aligned} \tag{59}$$

in the same time interval. The converse of the statement is also true. In [27], it was also noted that it is sufficient to use the previous observation, i.e., $u_l(t, x)$ satisfies Equation 57 if and only if $\tilde{u}_l(t, x)$ defined as

$$\tilde{u}_l(t, x) = \exp \left(\sum_{i=1}^m \gamma_i(\tau_{l-1}) h_i(x) \right) u_l(t, x) \tag{60}$$

satisfies Equation 59 in the time interval $\tau_{l-1} \leq t \leq \tau_l$. We refer to Equations 58 (60) and Equation 59 as the post-measurement (pre-measurement) forms of the YYe.

Observe that Equation 57 is obtained by setting $\gamma(t)$ to $\gamma(\tau_l)$ in Equation 56. It was proved that the solution of Equation 57 approximates very well the solution of the robust DMZ equation

(Equation 56), i.e., it converges to $u(t, x)$ in both pointwise sense and L^2 sense. Thus, solving Equation 56 is equivalent to solving Equation 59. Finally,

$$\sigma(\tau, x) = \lim_{|\mathcal{P}_k| \rightarrow 0} \tilde{u}_k(\tau_k, x). \tag{61}$$

Thus, the solution of the YYe (as $|\mathcal{P}_k| \rightarrow 0$) is the desired unnormalized conditional probability density.

Observe that when $h(x) = 0$, it is simply the FPKfe. However, unlike the FPKfe, the YYe does not satisfy the current conservation condition, i.e., the right-hand term is not a total divergence. This means that it does not conserve probability. This fundamental difference is traced to the fact that *the FPKfe evolves the normalized probability density (and preserves the normalization), while the YYe evolves the unnormalized conditional probability density*. Therefore, this distinction is made between the two equations in this paper.

A.5 The Yau Algorithm

We may summarize the real-time algorithm, based on both the pre- and post-measurement forms of the YYe, of Yau as follows. Suppose measurements are available at times

$$\dots < \tau_0 < \tau_1 < \tau_2 < \dots < \tau_k = \tau. \tag{62}$$

We seek the solution $u_i(t, x)$, which is the solution of the robust DMZ equation. Let the initial distribution be $u(\tau_0, x) = \sigma_0(x)$. Then, evolve the initial distribution to the first measurement instant, τ_1 , using the YYe:

$$\left\{ \begin{array}{l} \frac{\partial \tilde{u}_1}{\partial t}(t, x) = \frac{1}{2} \sum_{i=1}^n \frac{\partial^2 \tilde{u}_1}{\partial x_i^2}(t, x) - \sum_{i=1}^n f_i(x) \frac{\partial \tilde{u}_1}{\partial x_i}(t, x) - \left(\sum_{i=1}^n \frac{\partial f_i}{\partial x_i}(x) + \frac{1}{2} \sum_{i=1}^m h_i^2(x) \right) \tilde{u}_1(t, x), \\ \tilde{u}_1(\tau_0, x) = \sigma_0(x) \left\{ \begin{array}{l} \exp\left(\sum_{j=1}^m [\gamma_j(\tau_0) - \gamma_j(\tau_{-1})] h_j(x) \right) \quad \text{(Pre-Measurement)} \\ \exp\left(\sum_{j=1}^m [\gamma_j(\tau_1) - \gamma_j(\tau_0)] h_j(x) \right) \quad \text{(Post-Measurement)}. \end{array} \right. \end{array} \right. \tag{63}$$

The solution of equation 63 at time τ_1 is $\tilde{u}_1(\tau_1, x)$. Note that $u_1(\tau_1, x)$ is given by

$$u_1(\tau_1, x) = \tilde{u}_1(\tau_1, x) \left\{ \begin{array}{l} \exp\left(-\sum_{i=1}^m \gamma_i(\tau_0) h_i(x) \right) \quad \text{(Pre-Measurement)} \\ \exp\left(-\sum_{i=1}^m \gamma_i(\tau_1) h_i(x) \right) \quad \text{(Post-Measurement)}. \end{array} \right. \tag{64}$$

Next, solve the YYe to the next measurement instant τ_2 with initial condition $\tilde{u}_2(\tau_1, x)$, i.e.,

$$\left\{ \begin{aligned} \frac{\partial \tilde{u}_2}{\partial t}(t, x) &= \frac{1}{2} \sum_{i=1}^n \frac{\partial^2 \tilde{u}_2}{\partial x_i^2}(t, x) - \sum_{i=1}^n f_i(x) \frac{\partial \tilde{u}_2}{\partial x_i}(t, x) - \left(\sum_{i=1}^n \frac{\partial f_i}{\partial x_i}(x) + \frac{1}{2} \sum_{i=1}^m h_i^2(x) \right) \tilde{u}_2(t, x), \\ \tilde{u}_2(\tau_1, x) &= \tilde{u}_1(\tau_1, x) \begin{cases} \exp\left(\sum_{j=1}^m (\gamma_j(\tau_1) - \gamma_j(\tau_0)) h_j(x)\right) & \text{(Pre-Measurement)} \\ \exp\left(\sum_{j=1}^m (\gamma_j(\tau_2) - \gamma_j(\tau_1)) h_j(x)\right) & \text{(Post-Measurement)}, \end{cases} \end{aligned} \right. \tag{65}$$

to obtain $\tilde{u}_2(\tau_2, x)$. In fact, for $i \geq 2$, $u_i(\tau_i, x)$ can be computed from $\tilde{u}_i(\tau_i, x)$, where $\tilde{u}_i(t, x)$ satisfies the equation

$$\left\{ \begin{aligned} \frac{\partial \tilde{u}_i}{\partial t}(t, x) &= \frac{1}{2} \sum_{i=1}^n \frac{\partial^2 \tilde{u}_i}{\partial x_i^2}(t, x) - \sum_{j=1}^n f_j(x) \frac{\partial \tilde{u}_i}{\partial x_j}(t, x) - \left(\sum_{j=1}^n \frac{\partial f_j}{\partial x_j}(x) + \frac{1}{2} \sum_{j=1}^m h_j^2(x) \right) \tilde{u}_i(t, x), \\ \tilde{u}_i(\tau_{i-1}, x) &= \tilde{u}_{i-1}(\tau_{i-1}, x) \begin{cases} \exp\left(\sum_{j=1}^m (\gamma_j(\tau_{i-1}) - \gamma_j(\tau_{i-2})) h_j(x)\right) & \text{(Pre-Measurement)} \\ \exp\left(\sum_{j=1}^m (\gamma_j(\tau_i) - \gamma_j(\tau_{i-1})) h_j(x)\right) & \text{(Post-Measurement)}. \end{cases} \end{aligned} \right. \tag{66}$$

The initial condition in Equation 66 follows from noting that (since $u_i(\tau_{i-1}, x) = u_{i-1}(\tau_{i-1}, x)$)

$$\begin{aligned} \tilde{u}_i(\tau_{i-1}, x) &= u_i(\tau_{i-1}, x) \exp\left(\sum_{j=1}^m \gamma_j(\tau_i) h_j(x)\right), \\ &= \exp\left(-\sum_{j=1}^m \gamma_j(\tau_{i-1}) h_j(x)\right) \tilde{u}_{i-1}(\tau_{i-1}, x) \exp\left(\sum_{j=1}^m \gamma_j(\tau_i) h_j(x)\right), \end{aligned} \tag{67}$$

$$\begin{aligned} \tilde{u}_i(\tau_{i-1}, x) &= u_i(\tau_{i-1}, x) \exp\left(\sum_{j=1}^m \gamma_j(\tau_i) h_j(x)\right), \\ &= \exp\left(-\sum_{j=1}^m \gamma_j(\tau_{i-1}) h_j(x)\right) \tilde{u}_{i-1}(\tau_{i-1}, x) \exp\left(\sum_{j=1}^m \gamma_j(\tau_i) h_j(x)\right), \end{aligned} \tag{68}$$

Thus the solution of the robust DMZ equation at the i th step is given by

$$u_i(\tau_i, x) = \tilde{u}_i(\tau_i, x) \begin{cases} \exp\left(-\sum_{j=1}^m \gamma_j(\tau_{i-1}) h_j(x)\right) & \text{(Pre-Measurement)} \\ \exp\left(-\sum_{j=1}^m \gamma_j(\tau_i) h_j(x)\right) & \text{(Post-Measurement)}. \end{cases} \quad (69)$$

Note that the Yau algorithm is a recursive algorithm as it does not need any previous observation data. Furthermore, the YYe is independent of data and so can be computed off-line, and that the YYe is much simpler than the robust DMZ equation. Finally, note that the output of the Yau algorithm is the desired unnormalized conditional probability density.

Note that while there are several other possible real-time solutions to the continuous-continuous nonlinear filtering problem, all of them assume that the signal and measurement model drifts are bounded (see remarks in [26] and [28]). As a result, those algorithms cannot provide a reliable solution to many real-world problems.

Acknowledgements

The author thanks the Editor and three referees for constructive comments and suggestions that helped in significantly improving the paper. This work was supported by a Defence Research and Development Canada under a Technology Investment Fund.

References

1. Jazwinski AH: *Stochastic Processes and Filtering Theory* Dover Publications; 2007.
2. Kalman RE: *Trans ASME J Basic Eng* 1960, **82D**:35-45.
3. Kalman RE, Bucy RS: *Trans ASME J Basic Eng* 1961, **83**:95-108.
4. Kushner HJ: *SIAM Journal of Control* 1964, **2**:106-119.
5. Kushner HJ: *SIAM Journal of Control* 1964, **2**:106-119.
6. Stratanovich RL: *Theor Prob Appl* 1960, **5**:156-178.
7. Bucy RS: *IEEE Transactions on Automatic Control* 1965, **10(2)**:198.
8. Duncan TE: **Probability densities for diffusion processes with applications to nonlinear filtering theory**. In *PhD thesis* Stanford University; 1967.
9. Mortensen RE: **Optimal control of continuous time stochastic systems**. In *PhD thesis* University of California, Berkeley; 1966.
10. Zakai M: *Z Wahrsch Verw Geb* 1969, **11**:230-243.
11. Davis MHA: *Z Wahrsch Vers Gebiete* 1980, **54(2)**:125-139.
12. Rozovskii BL: *Uspekhi Math Nauk* 1972, **27**:213-214.
13. Yau ST, Yau SST: *Mathematical Research Letters* 2000, **7**:671-693 [<http://www.mrlonline.org/mr/2000-007-006/2000-007-006-002.pdf>].
14. Yau ST, Yau SST: *Applied Mathematics and Optimization* 1996, **34(3)**:231-266.
15. Yau ST, Yau SST: *Advances in Mathematics* 1998, **140**:156-189.
16. Balaji B: *Journal of Statistical Mechanics: Theory and Experiment* 2008, **2008(01)**:P01014. (17pp).
17. Siegel W: 1999 [<http://www.arxiv.org/abs/hep-th/9912205>]. <http://www.citebase.org/abstract?id=oai:arXiv.org:hep-th/9912205>
18. Zinn-Justin J: *Quantum Field Theory and Critical Phenomena* International Series in Monographs on Physics, Oxford University Press; 2002.

19. Risken H: *The Fokker-Planck Equation: Methods of Solution and applications* 2nd edition. Springer-Verlag; 1999.
20. Balaji B: submitted to *IEEE Transactions on Aerospace and Electronic Systems* 2006 [<http://www.citebase.org/abstract?id=oai:arXiv.org:0708.0354>].
21. Balaji B: **Continuous-Discrete Filtering using the Dirac-Feynman Algorithm**. *IEEE Radar Conference, Rome, Italy* 2008.
22. Lepage PG: 2005 [<http://arxiv.org/abs/hep-lat/0506036>].
23. Fung CPA: **New Numerical algorithms for Nonlinear Filtering**. In *PhD thesis* University of Southern California; 1995.
24. Lototsky S, Mikulevicius R, Rozovskii BL: *SIAM Journal of Control and Optimization* 1997, **35**(2435-461) [<http://citeseer.ist.psu.edu/lototsky97nonlinear.html>].
25. Yan C, Yau SST: **A New suboptimal filter and numerical solutions for the cubic sensor problem**. *Proceedings of the 2006 IEEE International Conference on Networking, Sensing and Control* 2006.
26. Yau ST, Yau SST: *SIAM Journal on Control and Optimization* 2008, **47**:163-195 [<http://link.aip.org/link/?SJC/47/163/1>].
27. Yau SST, Yau ST: **Real Time Algorithm for nonlinear filtering problem**. In *Proceedings of the 40th IEEE Conference on Decision and Control* Orlando, FL, USA; 2001.
28. Yau SST, Yau ST: *SIAM Journal of Control and Optimization* 2005, **44**(3):1019-1039.