

Article

## Continuous-Discrete Path Integral Filtering

Bhashyam Balaji

Radar Systems Section, Defence Research and Development Canada, Ottawa, 3701 Carling Avenue, Ottawa, ON, K1A 0Z4, Canada; E-Mail: Bhashyam.Balaji@drdc-rddc.gc.ca; Tel.: +1-613-998-2215; Fax: +1-613-990-8906

Received: 19 February 2009 / Accepted: 6 August 2009 / Published: 17 August 2009

---

**Abstract:** A summary of the relationship between the Langevin equation, Fokker-Planck-Kolmogorov forward equation (FPKfe) and the Feynman path integral descriptions of stochastic processes relevant for the solution of the continuous-discrete filtering problem is provided in this paper. The practical utility of the path integral formula is demonstrated via some nontrivial examples. Specifically, it is shown that the simplest approximation of the path integral formula for the fundamental solution of the FPKfe can be applied to solve nonlinear continuous-discrete filtering problems quite accurately. The Dirac-Feynman path integral filtering algorithm is quite simple, and is suitable for real-time implementation.

**Keywords:** Fokker-Planck equation; Kolmogorov equation; universal nonlinear filtering; Feynman path integrals; path integral filtering; data assimilation; tracking; continuous-discrete filters; nonlinear filtering; Dirac-Feynman approximation

---

### 1. Introduction

The following continuous-discrete filtering problem often arises in practice. The time evolution of the state, or signal of interest, is well-described by a continuous-time stochastic process. However, the state process is not directly observable, i.e., the state process is a hidden continuous-time Markov process. Instead, what is measured is a related discrete-time stochastic process termed the measurement process. The continuous-discrete filtering problem is to estimate the state of the system, given the measurements [1].

When the state and measurement processes are linear, excellent performance is often obtained using the Kalman filter [2,3]. However, the Kalman filter merely propagates the conditional mean and covariance, so it is not a universally optimal filter and may be inadequate for some problems with non-Gaussian

characteristics (e.g., multi-modal). When the state and/or measurement processes are nonlinear, a (non-unique) linearization of the problem leads to an extended Kalman filter. If the nonlinearity is benign, it is still very effective. However, for the general case, it cannot provide a robust solution. Simple solutions are also possible for a more general class of filters ([4]), although this is still a limited class of filtering problems.

The complete solution of the filtering problem is the conditional probability density function (pdf) of the state given the observations. It is complete in the Bayesian sense, i.e., it contains all the probabilistic information about the state process that is in the measurements and the initial condition. The solution is termed universal if the initial distribution can be arbitrary. From the conditional probability density, one can compute quantities optimal under various criteria. For instance, the conditional mean is the least mean-squares estimate.

The solution of the continuous-discrete filtering problem requires the solution of a linear, parabolic, partial differential equation (PDE) termed the Fokker-Planck-Kolmogorov forward equation (FPKfe). There are three main techniques to solve the FPKfe type of equations, namely, finite difference methods [5,6], spectral methods [7], and finite/spectral element methods [8]. However, numerical solution of PDEs is not straightforward. For example, the error in a naïve discretization may not vanish as the grid size is reduced, i.e., it may not be convergent. Another possibility is that the method may not be consistent, i.e., it may tend to a different PDE in the limit that the discretization spacing vanishes. Furthermore, the numerical method may be unstable, or there may be severe time step size restrictions. Finally, such methods suffer from the “curse of dimensionality”, i.e., it is not possible to solve higher-dimensional problems.

The fundamental solution of the FPKfe can be represented in terms of a Feynman path integral [9]. The path integral formula can be derived directly from the Langevin equation. A textbook discussion for derivation of the path integral representation of the fundamental solution of the FPKfe corresponding to the Langevin equations for additive and multiplicative noise cases can be found in [10] and [11]. In this paper, it is demonstrated that the simplest approximate path integral formulae lead to a very accurate solution of the nonlinear continuous-discrete filtering problem. In short, we show that the path integral formulation provides a simple and efficient procedure for updating the Fokker-Planck operator required in the prediction step of Bayesian filtering. We demonstrate the utility of this formulation using a grid-based approximation to the conditional density on hidden states. In this approach, we represent conditional probabilities explicitly on a finely sampled grid in state-space. The advantage of this is that one can integrate the density and approximate any arbitrary posterior distribution on the unknown states generating data.

In Section 2., the basic concepts of continuous-discrete filtering theory is reviewed. In Section 3., the path integral formulae for the case of additive and multiplicative noise cases are summarized. In Section 4., an elementary solution of the continuous-discrete filtering problem is presented that is based on the path integral formulae. Some examples illustrating the path integral filtering are presented in the following section. Some remarks on the practical implementation aspects of path integral filtering is presented in Section 6.. The appendix summarizes the path integral results.

## 2. Review of Continuous-Discrete Filtering Theory

### 2.1. Langevin Equation and the FPKfe

The general continuous-time state model is described by the following stochastic differential equation (SDE):

$$d\mathbf{x}(t) = f(\mathbf{x}(t), t)dt + e(\mathbf{x}(t), t)d\mathbf{v}(t) \tag{1}$$

Here  $\mathbf{x}(t)$  and  $f(\mathbf{x}(t), t)$  are  $n$ -dimensional column vectors, the diffusion vielbein  $e(\mathbf{x}(t), t)$  is an  $n \times p_e$  matrix and  $\mathbf{v}(t)$  is a  $p_e$ -dimensional column vector. The noise process  $\mathbf{v}(t)$  is assumed to be Brownian (or Wiener process) with covariance  $Q(t)$  and the quantity  $g \equiv eQe^T$  is termed the diffusion matrix. The term vielbein alludes to the fact that the diffusion matrix in the related Fokker-Planck equation can be viewed as defining a metric in a Riemannian manifold (in Riemannian geometry, the square root of the metric is referred to as the vielbein (or vierbein), and the vielbein proves to be essential for coupling fermions to gravity.). All functions are assumed to be sufficiently smooth. Equation 1 is also referred to as the Langevin equation. It is interpreted in the Itô sense (see Appendix A.). Throughout the paper, bold symbols refer to the stochastic processes while the corresponding plain symbol refers to a sample of the process.

Let  $\sigma_0(x)$  be the initial probability distribution of the state process. Then, the evolution of the probability distribution of the state process described by the Langevin equation,  $p(t, x)$ , is described by the FPKfe, i.e.,

$$\begin{cases} \frac{\partial p}{\partial t}(t, x) = - \sum_{i=1}^n \frac{\partial}{\partial x_i} [f_i(x, t)p(t, x)] + \frac{1}{2} \sum_{i,j=1}^n \frac{\partial^2}{\partial x_i \partial x_j} [g_{ij}(t, x)p(t, x)] \\ p(t_0, x) = \sigma_0(x) \end{cases} \tag{2}$$

### 2.2. Fundamental Solution of the FPKfe

The solution of the FPKfe can be written as an integral equation. To see this, first note that the complete information is in the transition probability density, which also satisfies the FPKfe except with a  $\delta$ -function initial condition. Specifically, let  $t'' > t'$ , and consider the following PDE:

$$\begin{cases} \frac{\partial P}{\partial t}(t, x|t', x') = - \sum_{i=1}^n \frac{\partial}{\partial x_i} [f_i(x, t)P(t, x|t', x')] + \frac{1}{2} \sum_{i,j=1}^n \frac{\partial^2}{\partial x_i \partial x_j} [g_{ij}(t, x)P(t, x|t', x')] \\ P(t', x''|t', x') = \delta^n(x'' - x') \end{cases} \tag{3}$$

Here  $\delta^n(x'' - x')$  is the  $n$ -dimensional Dirac delta function, i.e.,  $\delta(x''_1 - x'_1)\delta(x''_2 - x'_2) \cdots \delta(x''_n - x'_n)$ . Such a solution, i.e.,  $P(t, x|t', x')$ , is also known as the fundamental solution of the FPKfe. From the fundamental solution one can compute the probability at a later time for an arbitrary initial condition as follows:

$$p(t'', x'') = \int P(t'', x''|t', x')p(t', x') \{d^n x'\} \tag{4}$$

In this paper, all integrals are assumed to be from  $-\infty$  to  $+\infty$ , unless otherwise specified. Therefore, in order to solve the FPKfe, it is sufficient to solve for the transition probability density  $P(t, x|t', x')$ . Note that this solution is universal in the sense that the initial distribution can be arbitrary.

### 2.3. Continuous-Discrete Filtering

In this paper, it is assumed that the measurement model is described by the following discrete-time stochastic process:

$$\mathbf{y}(t_k) = h(\mathbf{x}(t_k), \mathbf{w}(t_k), t_k), \quad k = 1, 2, \dots, \quad t_k > t_0 \tag{5}$$

where  $y(t) \in \mathbb{R}^{m \times 1}$ ,  $h \in \mathbb{R}^{m \times 1}$ , and the noise process  $\mathbf{w}(t)$  is assumed to be a white noise process. Note that  $\mathbf{y}(t_0) = 0$ . It is assumed that  $p(y(t_k)|x(t_k))$  is known.

Then, the universal continuous-discrete filtering problem can be solved as follows. Let the initial distribution be  $\sigma_0(x)$  and let the measurements be collected at time instants  $t_1, t_2, \dots, t_k, \dots$ . Let  $p(t_{k-1}, x(t_{k-1})|Y(t_{k-1}))$  be the conditional probability density at time  $t_{k-1}$ , where  $Y(\tau) = \{y(t_l) : t_0 < t_l \leq \tau\}$ . Then the conditional probability density at time  $t_k$ , after incorporating the measurement  $y(t_k)$ , is obtained via the prediction and correction steps:

$$\begin{cases} p(t_k, x|Y(t_{k-1})) = \int P(t_k, x|t_{k-1}, x_{k-1})p(t_{k-1}, x_{k-1}|Y(t_{k-1})) \{d^n x_{k-1}\}, & \text{(Prediction Step)} \\ p(t_k, x|Y(t_k)) = \frac{p(y(t_k)|x)p(t_k, x|Y(t_{k-1}))}{\int p(y(t_k)|\xi)p(t_k, \xi|Y(t_{k-1})) \{d^n \xi\}}, & \text{(Correction Step)} \end{cases} \tag{6}$$

Often (as in this paper), the measurement model is described by an additive Gaussian noise model, i.e.,

$$\mathbf{y}(t_k) = h(\mathbf{x}(t_k), t_k) + \mathbf{w}(t_k), \quad k = 1, 2, \dots, \quad t_k > t_0 \tag{7}$$

with  $\mathbf{w}(t) \sim N(0, R(t))$ , i.e.,

$$p(y(t_k)|x) = \frac{1}{((2\pi)^m \det R(t_k))^{1/2}} \exp \left\{ -\frac{1}{2}(y(t_k) - h(x(t_k), t_k))^T (R(t_k))^{-1} (y(t_k) - h(x(t_k), t_k)) \right\} \tag{8}$$

Observe that, as in the PDE formulations, one may use a convenient set of basis functions. Then, the evolution of each of the basis functions under the FPKfe follows from Equation 4. Since the basis functions are independent of measurements, the computation may be performed off-line. Finally, note that this solution of the filtering problem is universal. In conclusion, the determination of the fundamental solution of the FPKfe is equivalent to the solution of the universal optimal nonlinear filtering problem. A solution for the time independent case with orthogonal diffusion matrix in terms of ordinary integrals was presented in [12]. However, the integrand is complicated and not easily implementable in practice. In the next section, the fundamental solution for the general case in terms of path integrals is summarized. It is shown that it leads to formulae that are simple to implement.

### 3. Path Integral Formulas

In this section, path integral formulae are summarized. It is assumed that  $t'' > t'$ . Details on the formulae are summarized in Appendix A..

### 3.1. Additive Noise

When the diffusion vielbein is independent of the state, i.e.,

$$d\mathbf{x}(t) = f(\mathbf{x}(t), t)dt + e(t)d\mathbf{v}(t) \tag{9}$$

where all quantities are as defined in Section II-C, the noise is said to be additive. The path integral formula for the transition probability density is given by

$$P(t'', x''|t', x') = \int_{x(t')=x'}^{x(t'')=x''} [\mathcal{D}x(t)] \exp\left(-\int_{t'}^{t''} dt L^{(r)}(t, x, \dot{x})\right) \tag{10}$$

where the Lagrangian  $L^{(r)}(t, x, \dot{x})$  is defined as

$$L^{(r)}(t, x, \dot{x}) = \frac{1}{2} \sum_{i=1}^n (\dot{x}_i - f_i(x^{(r)}(t), t)) g_{ij}^{-1}(t) (\dot{x}_j - f_j(x^{(r)}(t), t)) + r \sum_{i=1}^n \frac{\partial f_i}{\partial x_i}(x, t) \tag{11}$$

and

$$g_{ij}(t) = \sum_{a,b=1}^{p_e} e_{ia}(t) Q_{ab}(t) e_{bj}(t) \tag{12}$$

and

$$[\mathcal{D}x(t)] = \frac{1}{\sqrt{(2\pi\epsilon)^n \det g(t')}} \lim_{N \rightarrow \infty} \prod_{k=1}^N \frac{d^n x(t' + k\epsilon)}{\sqrt{(2\pi\epsilon)^n \det g(t' + k\epsilon)}} \tag{13}$$

Here,  $r \in [0, 1]$  specifies the discretization of the SDE (see [13] and Appendix A. for details). The quantity  $S(t'', t') = \int_{t'}^{t''} L^{(r)}(t, x, \dot{x})dt$  is referred to as the action.

### 3.2. Multiplicative Noise

The state model for the general case is given by

$$d\mathbf{x}(t) = f(\mathbf{x}(t), t)dt + e(\mathbf{x}(t), t)d\mathbf{v}(t) \tag{14}$$

As discussed in more detail in Appendix A., the definition of this SDE is ambiguous due to the fact that  $d\mathbf{v}(t) \approx O(\sqrt{dt})$ . The path integral formula for the general discretization is complicated and summarized in Appendix A.. In the simplest Itô case, it reduces to

$$P(t'', x''|t', x') = \int_{x(t')=x'}^{x(t'')=x''} [\mathcal{D}x(t)] \exp\left(-\int_{t'}^{t''} dt L^{(r,0)}(t, x, \dot{x})\right) \tag{15}$$

where the Lagrangian  $L^{(r,0)}(t, x, \dot{x})$  is defined as

$$L^{(r,0)}(t, x, \dot{x}) = \frac{1}{2} \sum_{i=1}^n (\dot{x}_i - f_i(x^{(r)}(t), t)) g_{ij}^{-1}(x^{(0)}(t), t) (\dot{x}_j - f_j(x^{(r)}(t), t)) + r \sum_{i=1}^n \frac{\partial f_i}{\partial x_i}(x^{(r)}, t) \tag{16}$$

and

$$g_{ij}(x^{(0)}(t), t) = \sum_{a,b=1}^{p_e} e_{ia}(x^{(0)}(t), t) Q_{ab}(t) e_{bj}(x^{(0)}(t), t) \tag{17}$$

A nice feature of the Itô interpretation is that the formula is the same as that for the simpler additive noise case (with some obvious changes).

Note that it is always possible to convert from a SDE defined in any sense (say, Stratanovich or  $s = 0$ ) to the corresponding Itô SDE. Therefore, this can be considered to be the result for the general case.

#### 4. Dirac-Feynman Path Integral Filtering

The path integral is formally defined as the  $N \rightarrow \infty$  limit of a  $N$  multi-dimensional integrals and yields the correct answer for arbitrary time step size. In this section, an algorithm for continuous-discrete filtering using the simplest approximation to the path integral formula, termed the Dirac-Feynman approximation, is derived.

##### 4.1. Dirac-Feynman Approximation

Consider first the additive noise case. When the time step  $\epsilon \equiv t'' - t'$  is infinitesimal, the path integral is given by

$$P(t' + \epsilon, x'' | t', x') = \frac{1}{\sqrt{(2\pi\epsilon)^n \det g(t')}} \exp [-\epsilon L^{(r)}(t, x', x'', (x'' - x')/\epsilon)] \tag{18}$$

where the Lagrangian is

$$\begin{aligned} \frac{1}{2} \sum_{i,j=1}^n \left[ \frac{x''_i - x'_i}{\epsilon} - f_i(x' + r(x'' - x'), t) \right] g_{ij}^{-1}(t') \left[ \frac{x''_j - x'_j}{\epsilon} - f_j(x' + r(x'' - x'), t) \right] \\ + r \sum_{i=1}^n \frac{\partial f_i}{\partial x_i}(x' + r(x'' - x'), t) \end{aligned} \tag{19}$$

This leads to a natural approximation for the path integral for small time steps:

$$P(t'', x'' | t', x') = \frac{1}{\sqrt{(2\pi(t'' - t'))^n \det g(t')}} \exp [-\epsilon L^{(r)}(t, x', (x'' - x')/(t'' - t'))] \tag{20}$$

A special case is the one-step pre-point approximate formula

$$\begin{aligned} P(t'', x'' | t', x') = \frac{1}{\sqrt{(2\pi(t'' - t'))^n \det g(t')}} \\ \exp \left( -\frac{(t'' - t')}{2} \sum_{i,j=1}^n \left[ \frac{(x''_i - x'_i)}{(t'' - t')} - f_i(x', t') \right] g_{ij}^{-1}(t') \left[ \frac{(x''_j - x'_j)}{(t'' - t')} - f_j(x', t') \right] \right) \end{aligned} \tag{21}$$

The one-step symmetric approximate path integral formula for the transition probability amplitude (as originally used by Feynman in quantum mechanics [9]) is

$$\begin{aligned} P(t'', x'' | t', x') = \frac{1}{\sqrt{(2\pi(t'' - t'))^n \det g(\bar{t})}} \\ \times \exp \left( -\frac{(t'' - t')}{2} \sum_{i,j=1}^n \left[ \frac{(x''_i - x'_i)}{(t'' - t')} - f_i(\bar{x}, \bar{t}) \right] g_{ij}^{-1}(\bar{t}) \left[ \frac{(x''_j - x'_j)}{(t'' - t')} - f_j(\bar{x}, \bar{t}) \right] - \frac{(t'' - t')}{2} \sum_{i=1}^n \frac{\partial f_i}{\partial x_i}(\bar{x}, \bar{t}) \right) \end{aligned} \tag{22}$$

where  $\bar{x} = \frac{1}{2}(x'' + x')$  and  $\bar{t} = \frac{1}{2}(t' + t'')$ . Note that for the explicit time-dependent case, the time has also been symmetrized in the hope that it will give a more accurate result. Of course, for small time steps and if the time dependence is benign, the error in using this or the end points is small.

Similarly, for the multiplicative noise case in the Itô interpretation/discretization of the state SDE the following approximate formula results:

$$P(t'', x'' | t', x') = \frac{1}{\sqrt{(2\pi(t'' - t'))^n \det g(t')}} \exp \left[ -(t'' - t') L^{(r,0)}(t, x', (x'' - x')/(t'' - t')) \right] \tag{23}$$

where the Lagrangian  $L^{(r,0)}(t, x', x'', (x'' - x')/(t'' - t'))$  is given by

$$L^{(r,0)}(t, x', x'', (x'' - x')/(t'' - t')) = \frac{1}{2} \sum_{i,j=1}^n \left( \frac{(x''_i - x'_i)}{(t'' - t')} - f_i(x' + r(x'' - x'), t') \right) \left( \sum_{a,b=1}^{p_e} e_{ia}(x', t') Q_{ab}(t') e_{jb}(x', t') \right)^{-1} \left( \frac{(x''_j - x'_j)}{(t'' - t')} - f_j(x' + r(x'' - x'), t') \right) + r \sum_{i=1}^n \frac{\partial f_i}{\partial x_i}(x' + r(x'' - x'), t) \tag{24}$$

For the multiplicative noise case, the simplest one-step approximation is the pre-point discretization where  $r = s = 0$ :

$$P(t'', x'' | t', x') = \frac{1}{\sqrt{(2\pi(t'' - t'))^n \det g(x', t')}} \times \exp \left[ -\frac{(t'' - t')}{2} \left( \frac{(x''_i - x'_i)}{(t'' - t')} - f_i(x', t') \right) g_{ij}^{-1}(x', t') \left( \frac{(x''_j - x'_j)}{(t'' - t')} - f_j(x', t') \right) \right] \tag{25}$$

Since  $s = 0$ , this means that we are using the Itô interpretation of the state model Langevin equation. When  $r = 1/2$ , it is termed the Feynman convention, while  $s = 1/2$  corresponds to the Stratanovich interpretation.

#### 4.2. The Dirac-Feynman Algorithm

The one-step formulae discussed in the previous section lead to the simplest path integral filtering algorithm, termed the Dirac-Feynman (DF) algorithm. The steps for DF algorithm may be summarized as follows:

1. From the state model, obtain the expression for the Lagrangian. Specifically,
  - For the additive noise case, the Lagrangian is given by Equation 11;
  - For the multiplicative noise case with Itô discretization the Lagrangian is given by Equation 16, while for the general discretization the action is given in Appendix A..
2. Determine a one-step discretized Lagrangian that depends on  $r \in [0, 1]$  (and  $s \in [0, 1]$  for the multiplicative noise).
3. Compute the transition probability density  $P(t'', x'' | t', x')$  using the appropriate formula (e.g., Equation 22 or 25). The grid spacing should be such that the transition probability tensor and the measurement likelihood is adequately sampled, as discussed below.

4. At time  $t_k$

(a) The prediction step is accomplished by

$$p(t_k, x|Y(t_{k-1})) = \sum_{x'} P(t_k, x|t_{k-1}, x')p(t_{k-1}, x'|Y(t_{k-1})) \{\Delta^n x'\} \quad (26)$$

where  $\{\Delta^n x'\} \equiv \Delta x'_1 \Delta x'_2 \cdots \Delta x'_n$  is the grid measure. Note that  $p(t_0|Y(t_0))$  is simply the initial condition  $p(t_0, x_0) = \sigma_0(x_0)$ .

(b) The measurement at time  $t_k$  are incorporated in the correction step via

$$p(t_k, x|Y(t_k)) = \frac{p(y(t_k)|x)p(t_k, x|Y(t_{k-1}))}{\sum_{\xi} p(y(t_k)|\xi)p(t_k, \xi|Y(t_{k-1})) \{\Delta^n \xi\}} \quad (27)$$

### 4.3. Practical Computational Strategies

The above general filtering algorithm based on the Dirac-Feynman approximation of the path integral formula computes the conditional probability density at grid points. This can be computationally very expensive as the number of grid points can be very large, especially for larger dimensions. Here, a few approximations will be presented that drastically reduces the computational load.

The most crucial property that is exploited is that the transition probability density is an exponential function. Consequently, many elements of the transitional probability tensor are negligibly small, the precise number depending on the grid spacing. A significant computational saving is obtained when the (user-defined) “exponentially small” quantities are simply set to zero. For instance, for the one-dimensional case with  $10^4$  grid points, the transition probability matrix is  $10^4 \times 10^4$  has  $10^8$  elements which places a very large storage and computational requirements. However, if the off-diagonal elements are negligibly small so that only matrix elements satisfying  $|i - j| \leq 1$  are significant, then the number of significant matrix elements is only 0.03% of the number of elements in the full matrix. In the higher dimensional case, the transition probability density is approximated by a sparse tensor, which results in huge savings in memory and computational load.

The next key issue is that of grid spacing. An appropriate grid spacing is one that adequately samples the conditional probability density. Of course, the conditional probability density is not known, but its effective domain (i.e., where it is significant) is clearly a function of the signal and measurement model drifts and noises. For instance, the grid spacing should be of the order of change in state expected in a time step, which is not always easy to determine for a generic model. However, if the measurement noise is small, finer grid spacing is required so as to capture the state information in precise measurements. However, if the measurement noise is large, it may be unnecessary to use a fine grid spacing even if the state model noise is very small since the measurements are not that informative. Alternatively, if the grid spacing is too large compared to the signal model noise vielbein term, replace the diffusion matrix with an “effective diffusion matrix” that is taken to be a constant times the grid spacing, i.e., noise inflation. This additional approximation can still lead to useful results as shown in an example below.

It is also noted that the grid spacing is a function of the time steps. This is analogous to the case of PDE solution via discretization. Thus, when using the one-step DF approximation, there will not be a gain by reducing the grid spacing to smaller values (and at the cost of drastically increasing processing time). It is then more appropriate to use multiple time step approximations to get more accurate results.

Here are some possibilities for practical implementation:

1. Pre-compute the transition probability tensor with pre-determined and fixed grid. The exponential nature of the transition probability density can be used to speed the precomputation step considerably. Specifically, rather than computing to every grid point from a given point, one can omit computation to a final point which is unlikely to be reachable under the assumed dynamics. This is illustrated in the examples. For the correction step, there are two options:
  - (a) Compute the correction at all the grid points;
  - (b) Compute the correction only where the prediction result is significant.

The second of those options is used in the first two examples and the first option in the third example in the paper.

2. Another option is to use a focussed adaptive grid, much as in PDE approaches. Specifically, at each time step:
  - (a) Find where the prediction step result is significant;
  - (b) Find the domain in the state space where the conditional probability density is significant, and possibly interpolate. For the multi-modal case, there would be several disjoint regions;
  - (c) Compute the transition probability tensor with those points as initial points and propagate to points in region suggested by state model.

Thus, the grid is moving. In this case, the grid can be finer than in the previous case, although then the computational advantage of pre-computation is lost.

3. Pre-compute the solution using basis functions. For instance, in many applications wavelets have been known to provide a sparse and compact representation of a wide variety of functions arising in practice. The evolution of the wavelet basis functions can be computed using Equation 4. Then, instead of storing the transition probability tensor, FPKfe solutions with wavelet basis functions can be stored and used for filtering computations.

## 5. Examples

In this section, a couple of two-dimensional examples and one four-dimensional example are presented that illustrate the utility of the DF algorithm. The signal and measurement models are both nonlinear in these examples. Therefore, the Kalman filter is not a reliable solution for these problems. The symmetric discretization formula was used. The MATLAB<sup>®</sup> tensor toolbox developed by Bader and Kolda was used for the computations in the first two examples [14,15]. It is noted that *Mathematica*<sup>®</sup> also has a sparse tensor object; *Mathematica*<sup>®</sup> was used for the third example. The approximation techniques are discussed in Section 4.3. In addition, in order to speed up the pre-computation of the transition probability tensor, it was assumed that  $P(f_1, f_2 | i_1, i_2) = 0$  if  $|f_r - i_r| > r_{\text{ext}}$ , i.e., the “extent”  $r_{\text{ext}}$  of  $P$  was chosen to be 2 for the first two examples and 1 for the third example. Thus, this implementation of the DF algorithm is sub-optimal in many ways.

For comparison, the performance of the sampling importance re-sampling (SIR) particle filter based on the Euler discretization of the state model SDE is also included[16,17]. The MATLAB toolbox PFLib was used in the particle filter simulations[18]. It is noted that there are several particle filtering algorithms possible, such as those based on local linearization, that may yield better performance than the standard SIR-PF. However, the performance is not guaranteed for the general case. The comparison is fair since the assumed model is the same in both cases.

5.1. Example 1

Consider the state model

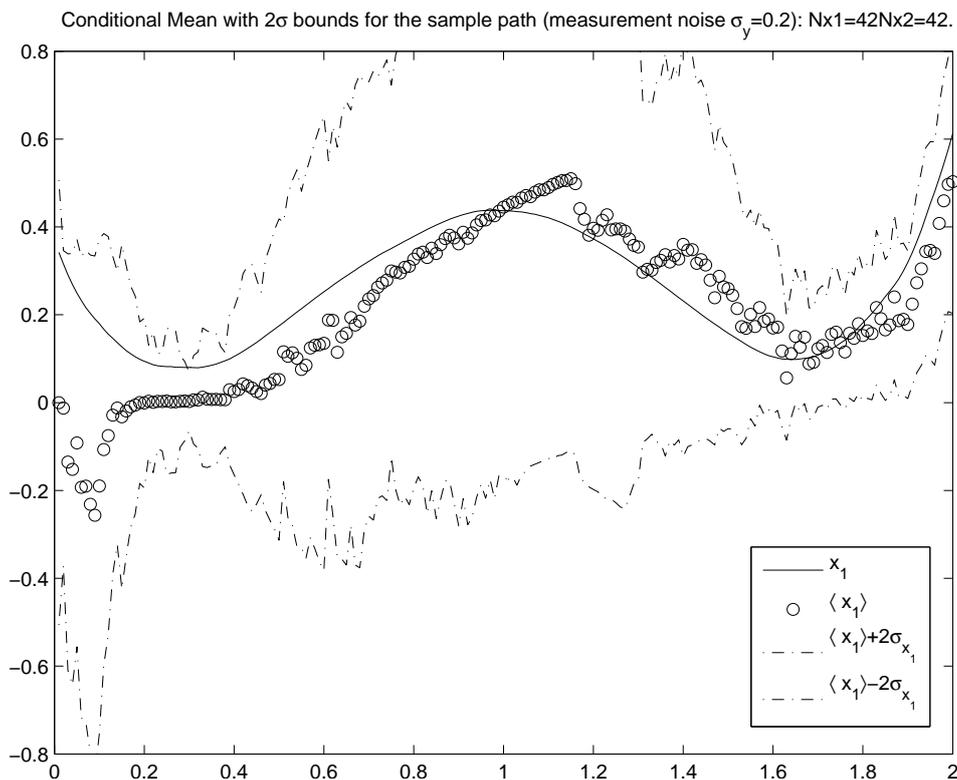
$$\begin{aligned} d\mathbf{x}_1(t) &= (-189\mathbf{x}_2^3(t) + 9.16\mathbf{x}_2(t)) dt + \sigma_{x_1} d\mathbf{v}_1(t), \\ d\mathbf{x}_2(t) &= -\frac{1}{3}dt + \sigma_{x_2} d\mathbf{v}_2(t), \end{aligned} \tag{28}$$

with the nonlinear measurement model

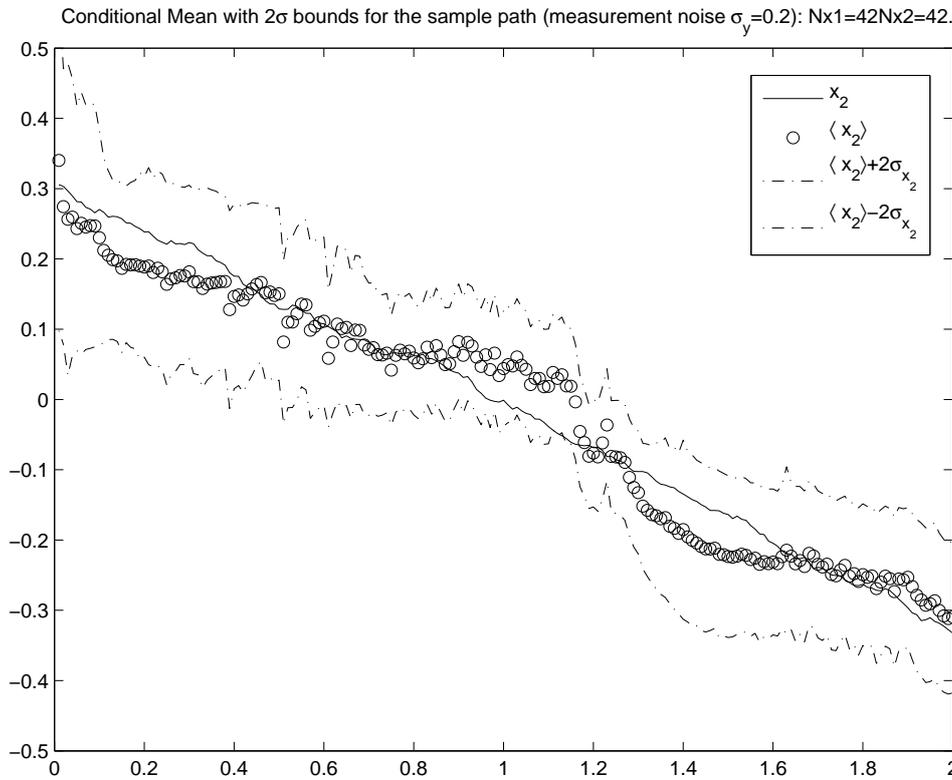
$$\mathbf{y}(t_k) = \sin^{-1} \left( \frac{\mathbf{x}_2(t_k)}{\sqrt{\mathbf{x}_1^2(t_k) + \mathbf{x}_2(t_k)}} \right) + \sigma_y \mathbf{w}(t_k). \tag{29}$$

Here  $[\sigma_{x_1} \ \sigma_{x_2}] = [0.001 \ 0.03]$  and we consider two values for  $\sigma_y$ , namely  $\sigma_y = 0.2$  and  $\sigma_y = 2$ . This example was studied in [19] and the extended Kalman is known to fail for this model.

**Figure 1.** Conditional mean  $\langle \mathbf{x}_1(t) \rangle$  computed for a measurement sample.



**Figure 2.** Conditional mean  $\langle \mathbf{x}_2(t) \rangle$  computed for a measurement sample.



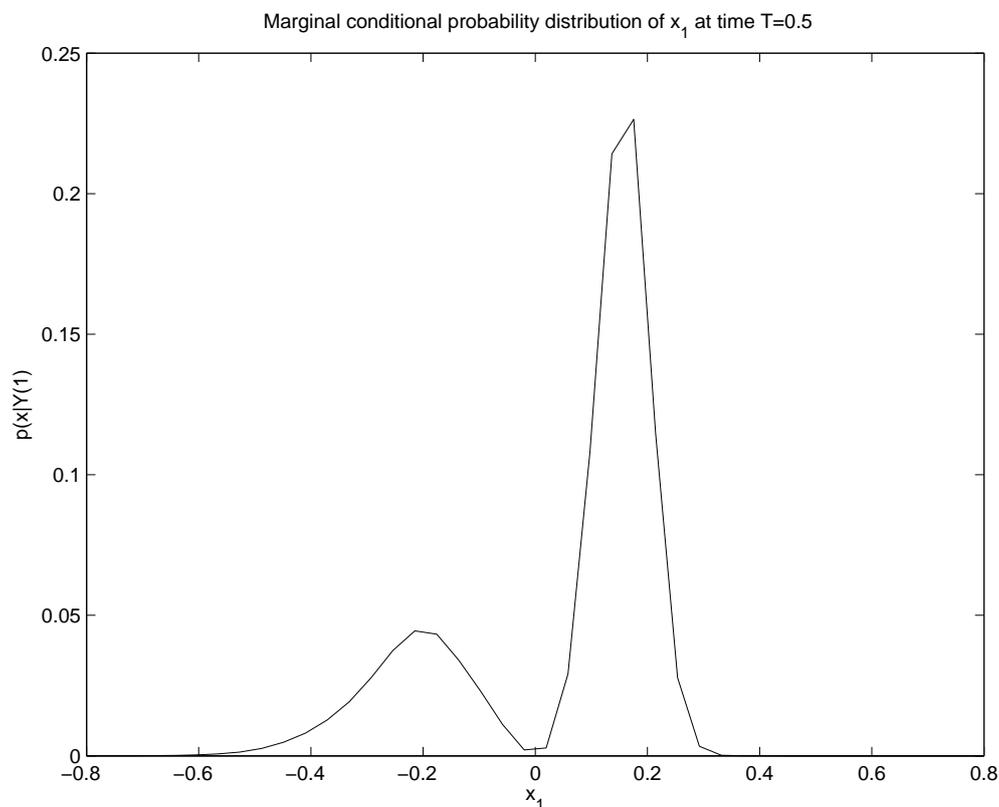
The Lagrangian for this model is easily seen to be given by

$$\frac{1}{2\sigma_{x_1}^2} (\dot{x}_1(t) + 189x_2^3(t) - 9.16x_2(t))^2 + \frac{1}{2\sigma_{x_2}^2} \left( \dot{x}_2(t) + \frac{1}{3} \right)^2 \tag{30}$$

Consider first the  $\sigma_y = 0.2$  case. The time step size is 0.01 and the number of time steps is 200. The spatial interval  $[-0.8, 0.8] \times [-0.8, 0.8]$  is subdivided into  $42 \times 42$  equal intervals. The signal model noise is very small requiring much finer grid spacing. Instead, as discussed in Section 4.3., the effective  $\sigma$ 's were taken to be  $\alpha \times \begin{bmatrix} \Delta x_1 & \Delta x_2 \end{bmatrix}$  with  $\alpha = 1$ . The initial distribution is taken to be uniform.

The conditional mean for  $\mathbf{x}_1(t)$  and  $\mathbf{x}_2(t)$  are plotted in Figures 1 and 2. The standard deviation in the figures were based on the estimated conditional probability density and therefore correspond to a conditional dispersion or precision. The RMS error was found to be 0.1180 and the time taken was 40 seconds. Observe that the conditional mean is within a standard deviation of the true state almost all of the time, which confirms that the tracking quality is good. It is noted that the variance is larger for  $\mathbf{x}_1(t)$ . The reason can be understood from Figure 3, which plots the marginal conditional probability density of the state variable  $\mathbf{x}_1(t)$ —it is bi-modal. The bi-modal nature (for a significant fraction of the time) is the reason the EKF will fail in this instance. The performance is seen to be similar to that reported in [19], which was obtained using considerably more involved techniques and finer grid spacing.

This example was also investigated using SIR-PF. The SIR-PF implemented with 5000 particles took about 155 seconds and the RMS error was found to be 0.165 even when initiated about the true state, i.e., initial distribution was chosen to be Gaussian with mean  $[0.37, 0.31]$  and variance  $I_2$ , where  $I_2$  is the  $2 \times 2$  identity matrix. When the variance was reduced to  $10^{-2} \times I$ , it resulted in an RMS error of only 0.022.

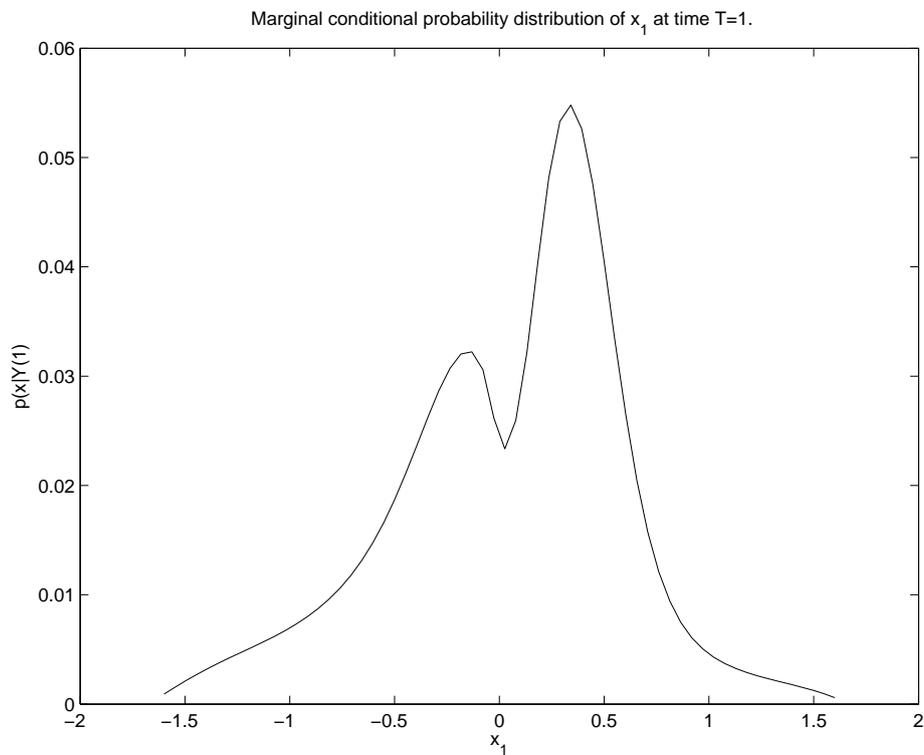
**Figure 3.** Marginal conditional probability density for  $\mathbf{x}_1(t)$ .

Thus, the performance of the SIR-PF depends crucially on the initial condition. It is also noted that no bi-modality of the marginal pdf of state  $\mathbf{x}_1(t)$  at  $T = 1$  was observed for the SIR-PF simulations when the number of particles was 5000. Upon increasing the number of particles to 10,000, the bi-modality was noted, although the RMSE was not significantly smaller.

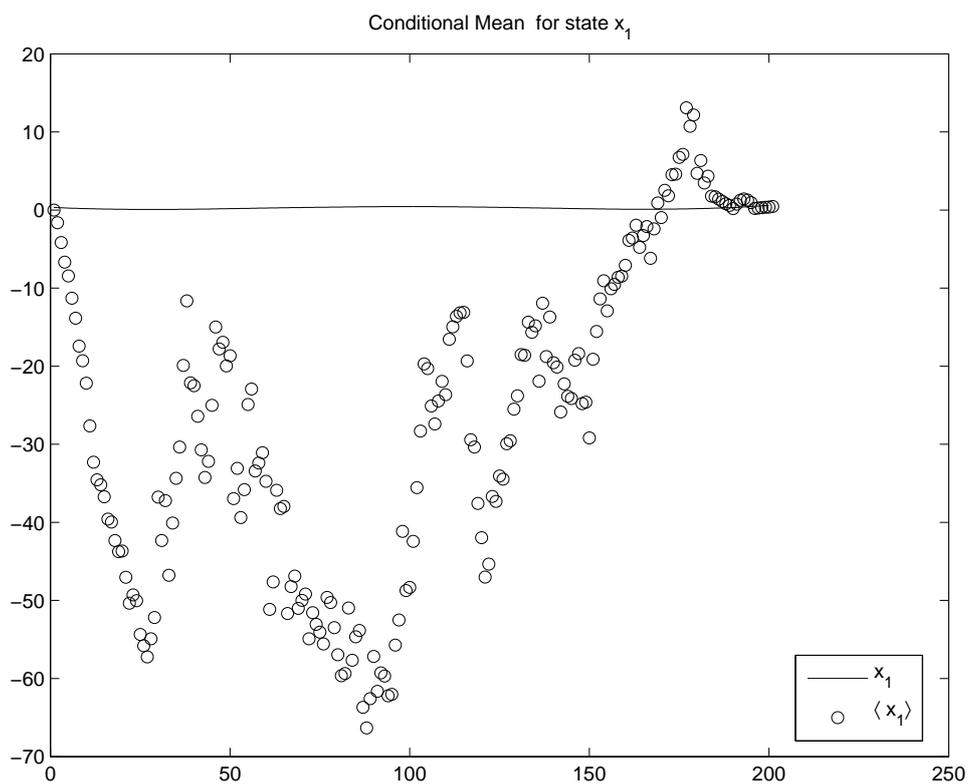
Next consider the larger measurement noise case. The spatial interval  $[-1.6, 1.6] \times [-1, 1]$  was subdivided into  $62 \times 62$  equispaced grid points. The RMS error was found to be 0.128 and the time taken was about 110 seconds. The bimodality at  $T = 1$  is evident in this case as well (see Figure 4).

The SIR-PF was also implemented. When initialized as Gaussian with zero mean and unit variance, the tracking performance of the SIR-PF failed; the RMSE was found to be 25.34 when using 5000 particles (taking about 110 seconds). A sample performance is shown in Figures 5 and 6; it is clear that the state  $\mathbf{x}_1$  is poorly tracked. Even when using 50,000 particles, a sample run that took 25 minutes, resulted in RMS error of 16.53.

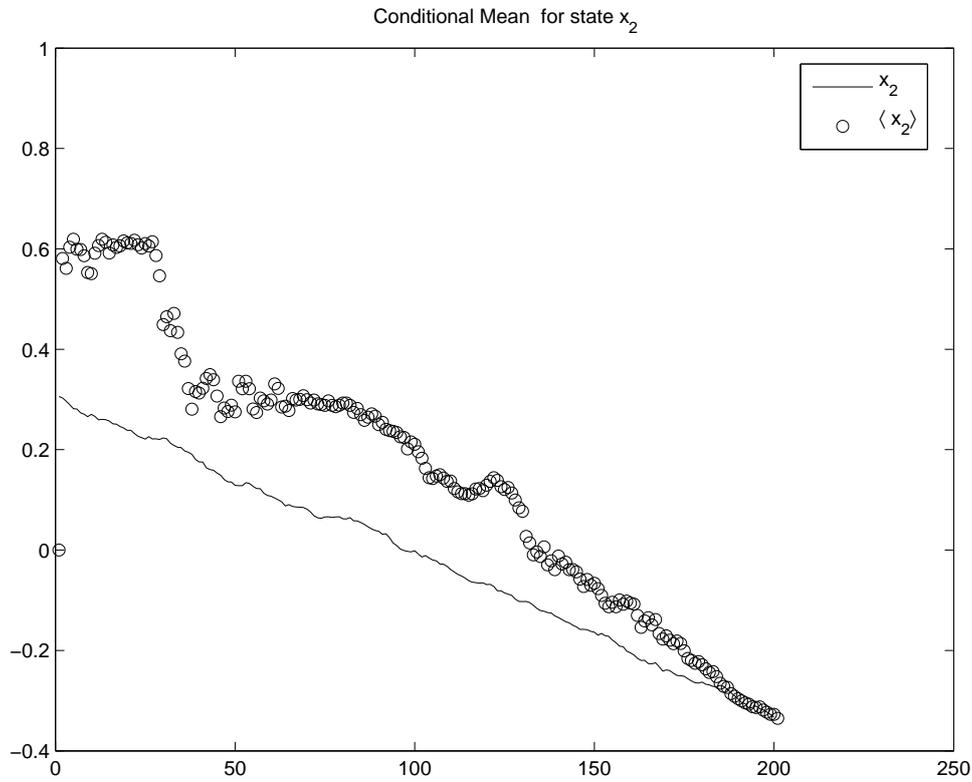
**Figure 4.** Marginal conditional probability density for  $x_1(t)$  for the larger measurement noise case.



**Figure 5.** Conditional mean for state  $x_1(t)$  computed using 5000 particles for  $\sigma_y = 2$ .



**Figure 6.** Conditional mean for state  $\mathbf{x}_2(t)$  computed using 5000 particles for  $\sigma_y = 2$ .



5.2. Example 2

Consider the state model

$$\begin{aligned} d\mathbf{x}_1(t) &= (-\mathbf{x}_2(t) + \cos(\mathbf{x}_1(t)))dt + d\mathbf{v}_1(t), \\ d\mathbf{x}_2(t) &= (\mathbf{x}_1(t) + \sin(\mathbf{x}_2(t)))dt + d\mathbf{v}_2(t), \end{aligned} \tag{31}$$

and the measurement model

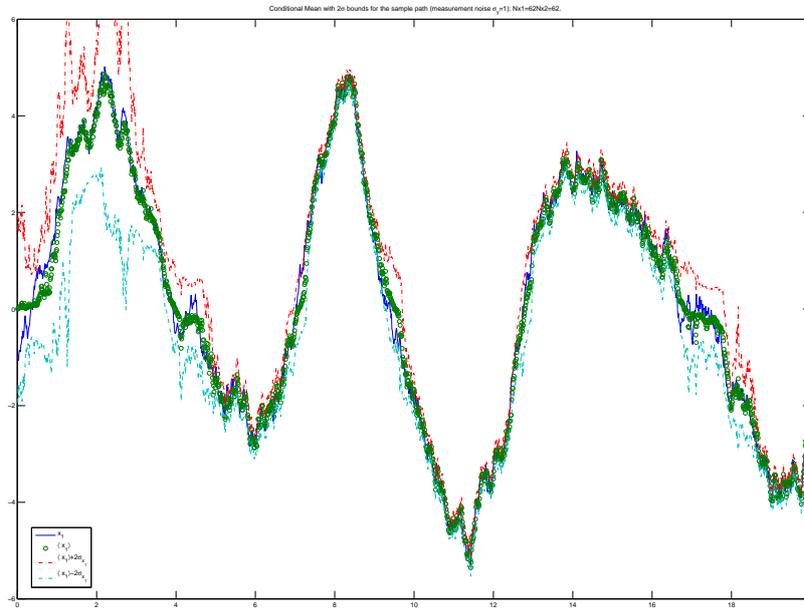
$$\begin{aligned} \mathbf{y}_1(t_k) &= \mathbf{x}_1^2(t_k) + \mathbf{w}_1(t_k), \\ \mathbf{y}_2(t_k) &= \mathbf{x}_2^2(t_k) + \mathbf{w}_2(t_k). \end{aligned} \tag{32}$$

Here  $d\mathbf{v}_i(t)$  are uncorrelated standard Wiener processes, and  $w_i(t_k) \sim N(0, 1)$ . The discretization time step is 0.01.

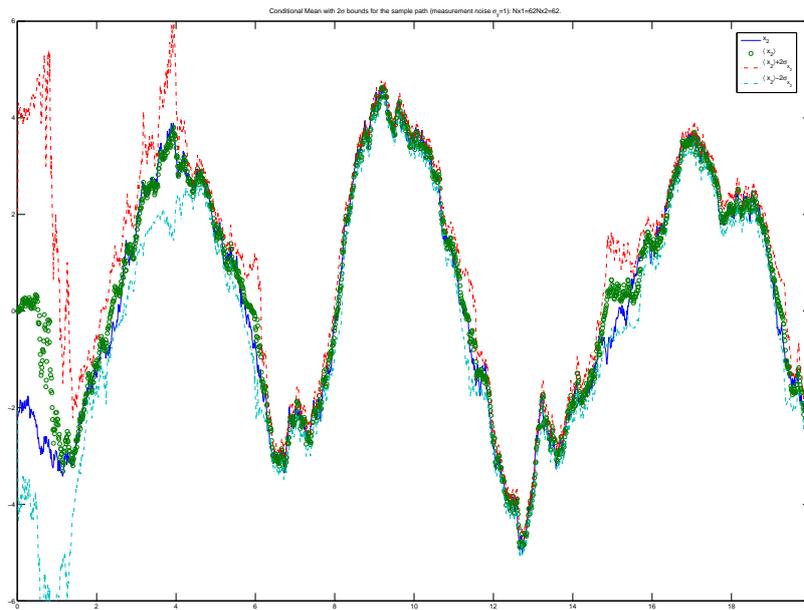
The initial distribution is taken to be a Gaussian with zero mean and a variance of 10. Figures 7 and 8 plots the conditional mean for the state. It is seen that the tracking is quite good despite the error at the start; the RMS error was found to be 0.54. The interval  $[-6, 6]$  was uniformly divided into 62 grid points and the extent of the transition probability tensor was 2. The 2000 time steps took about 8 minutes.

The SIR particle filter was also implemented with the same initial condition and with 5000 particles. The RMS error was found to be 1.48. Each run took about 10 minutes.

**Figure 7.** Conditional mean for state  $x_1(t)$  computed for a measurement sample and with initial distribution  $N(0, 10)$ .

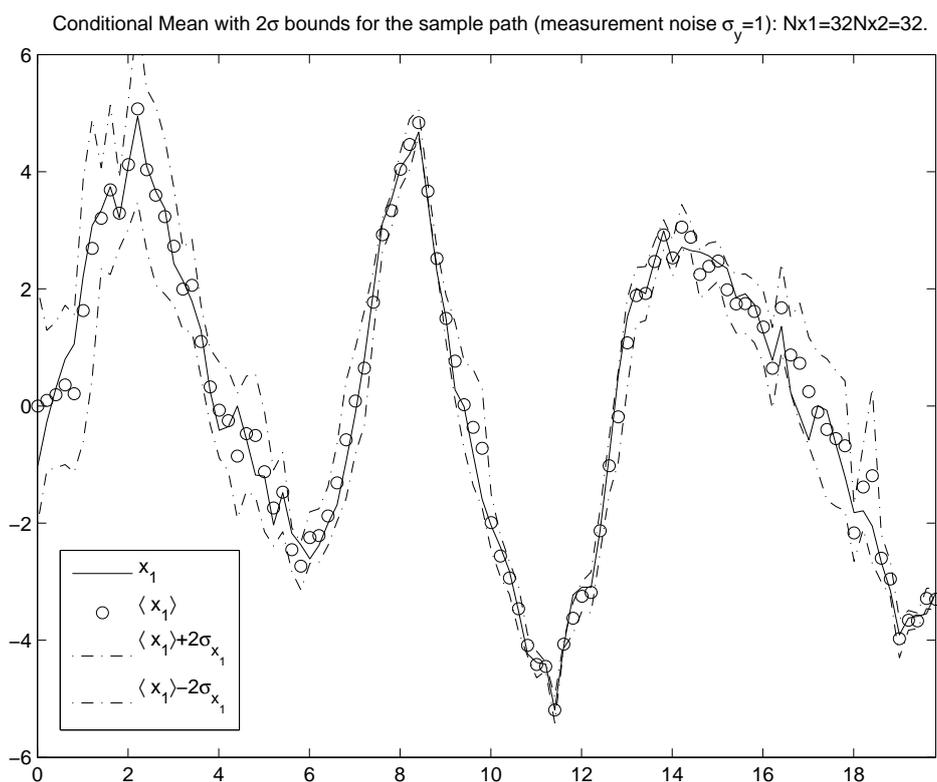


**Figure 8.** Conditional mean for state  $x_2(t)$  computed for a measurement sample and with initial distribution  $N(0, 10)$ .



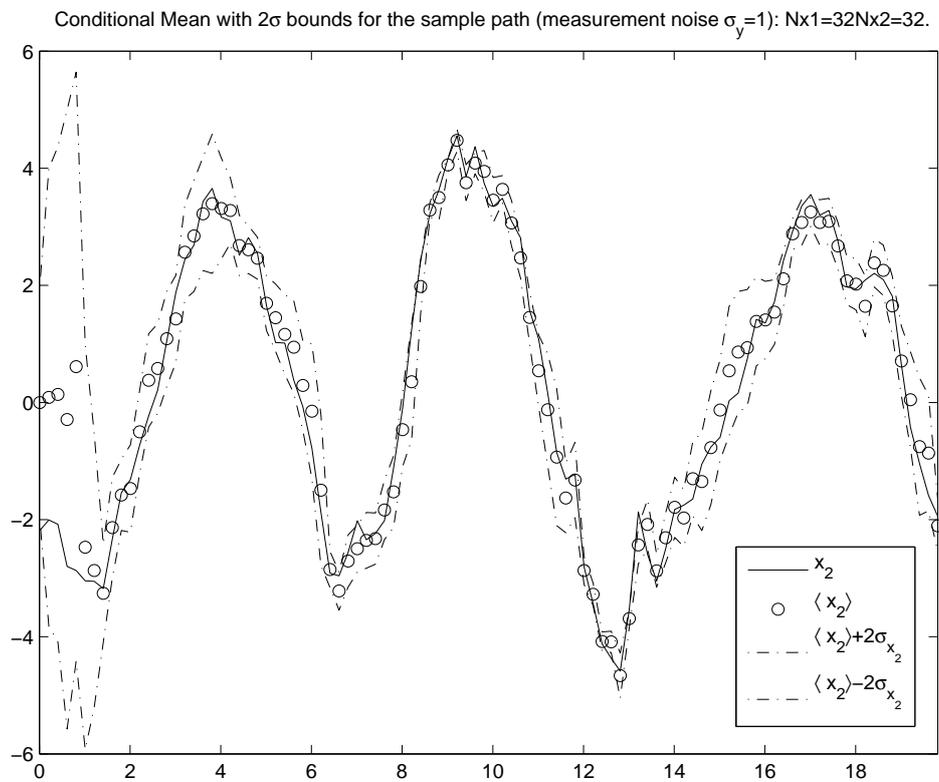
Next, consider time step of 0.2, i.e., only every twentieth measurement sample is assumed given. Figures 9 and 10 show the conditional means of the states. The number of grid points is smaller; the grid spacing is chosen to be twice the previous instance. Consequently, the computational effort is less, requiring only about 14 seconds. It is noted that the tracking performance is very good and the error estimated from the conditional probability density using this approximation is reliable. Now the RMS error is found to be 0.69, and only 0.31 if the first few errors are ignored.

**Figure 9.** Conditional mean for state  $x_1(t)$  when measurement interval is 0.2.

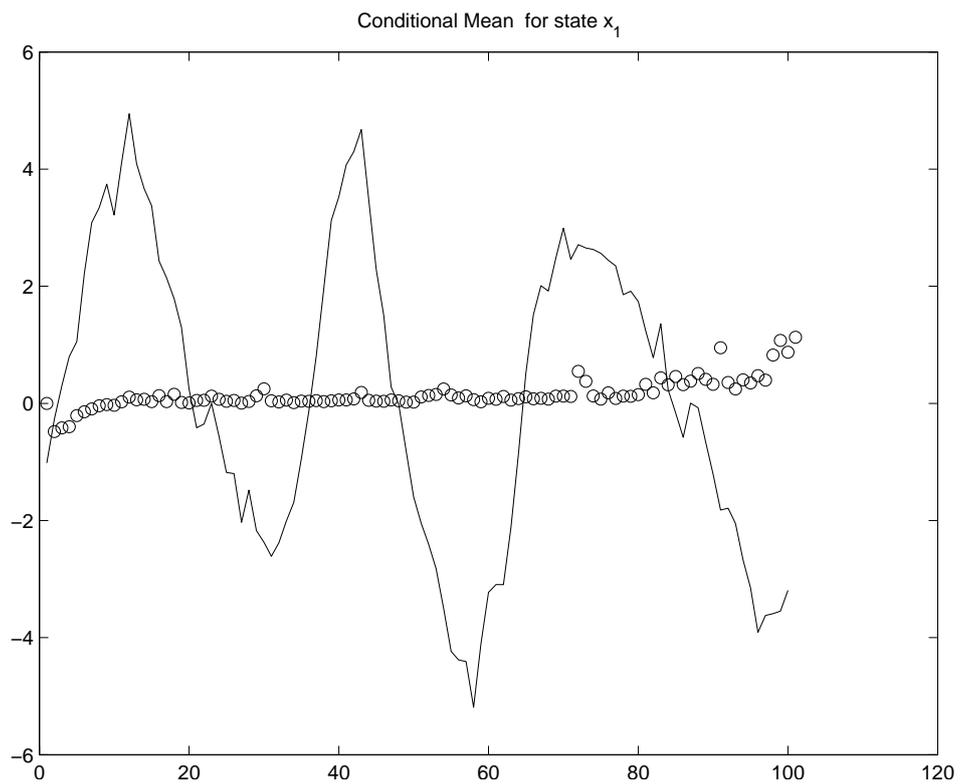


In contrast, the error using the SIR-PF (with 2000 particles that took 37 seconds) is found to fail with an RMS error of 3.68 (see Figures 11 and 12 for typical results). The results for increasing the number of particles to 50,000 (14 minutes execution time) did not improve the situation significantly (RMS error of 3.34); it would be better to divide the time step into several time steps and do the SIR-PF. For the purposes of this paper, it is sufficient to note that in this instance, a single-step DF algorithm succeeds where the one-step SIR-PF fails.

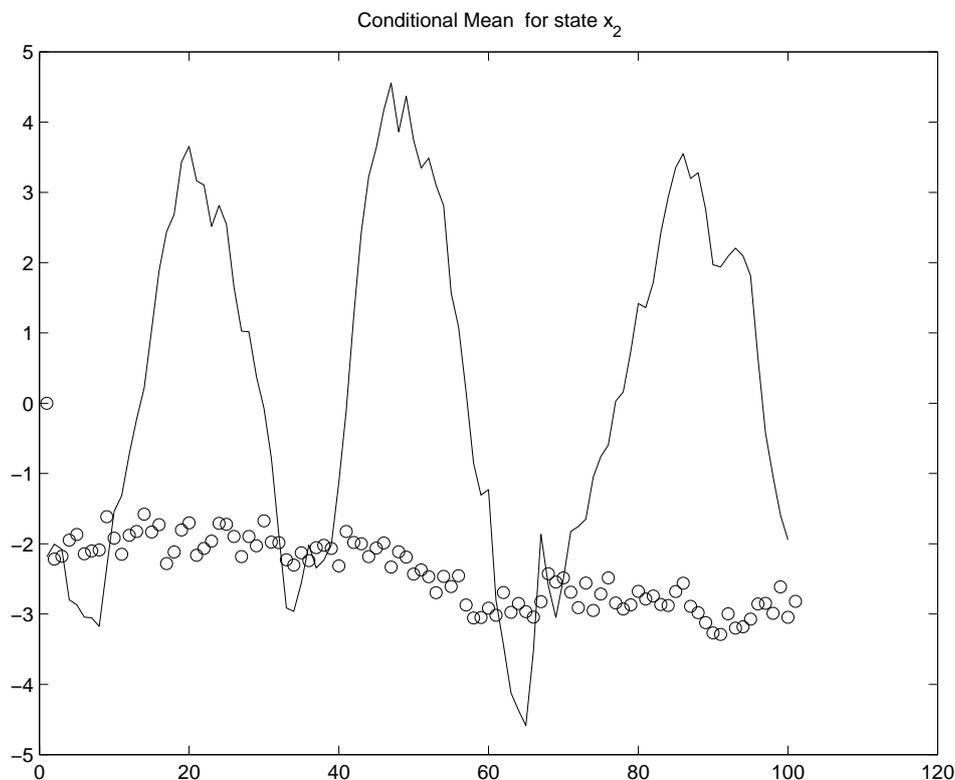
**Figure 10.** Conditional mean for state  $x_2(t)$  when measurement interval is 0.2.



**Figure 11.** Conditional mean for state  $x_1(t)$  computed using the SIR-PF when measurements are every 0.2 seconds.



**Figure 12.** Conditional mean for state  $\mathbf{x}_2(t)$  computed using the SIR-PF when measurements are every 0.2 seconds.



5.3. Example 3

In this Section, we consider a higher-dimensional example—the state model is four-dimensional with

$$\begin{aligned}
 f(\mathbf{x}) &= \begin{bmatrix} \mathbf{x}_2(t) - \mathbf{x}_1(t) \\ \mathbf{x}_1(t)(1 - \mathbf{x}_3(t)) - \mathbf{x}_2(t) \\ \mathbf{x}_1(t)\mathbf{x}_2(t) - \mathbf{x}_3(t) \\ (1 - \mathbf{x}_2(t))\mathbf{x}_3(t) - \mathbf{x}_4(t) \end{bmatrix} & (33) \\
 e(\mathbf{x}(t)) &= \begin{bmatrix} d_1 & \sin(\mathbf{x}_2(t)) & \cos(\mathbf{x}_3(t)) & \sin(\mathbf{x}_4(t)) \\ \sin(\mathbf{x}_4(t)) & d_2 & \cos(\mathbf{x}_3(t)) & \sin(\mathbf{x}_1(t)) \\ \cos(\mathbf{x}_1(t)) & \sin(\mathbf{x}_4(t)) & d_3 & \cos(\mathbf{x}_3(t)) \\ \sin(\mathbf{x}_1(t)) & \cos(\mathbf{x}_2(t)) & \sin(\mathbf{x}_3(t)) & d_4 \end{bmatrix} \\
 \begin{bmatrix} d_1 \\ d_2 \\ d_3 \\ d_4 \end{bmatrix} &= \begin{bmatrix} 5 + 0.1\tanh(\mathbf{x}_1(t) + \mathbf{x}_2(t)) \\ 5 + 0.1\tanh(\mathbf{x}_2(t) - \mathbf{x}_3^2(t)) \\ 5 + 0.1\tanh(\mathbf{x}_3(t) - \mathbf{x}_4(t)) \\ 5 + 0.1\tanh(\mathbf{x}_1(t) - \mathbf{x}_4(t)) \end{bmatrix}
 \end{aligned}$$

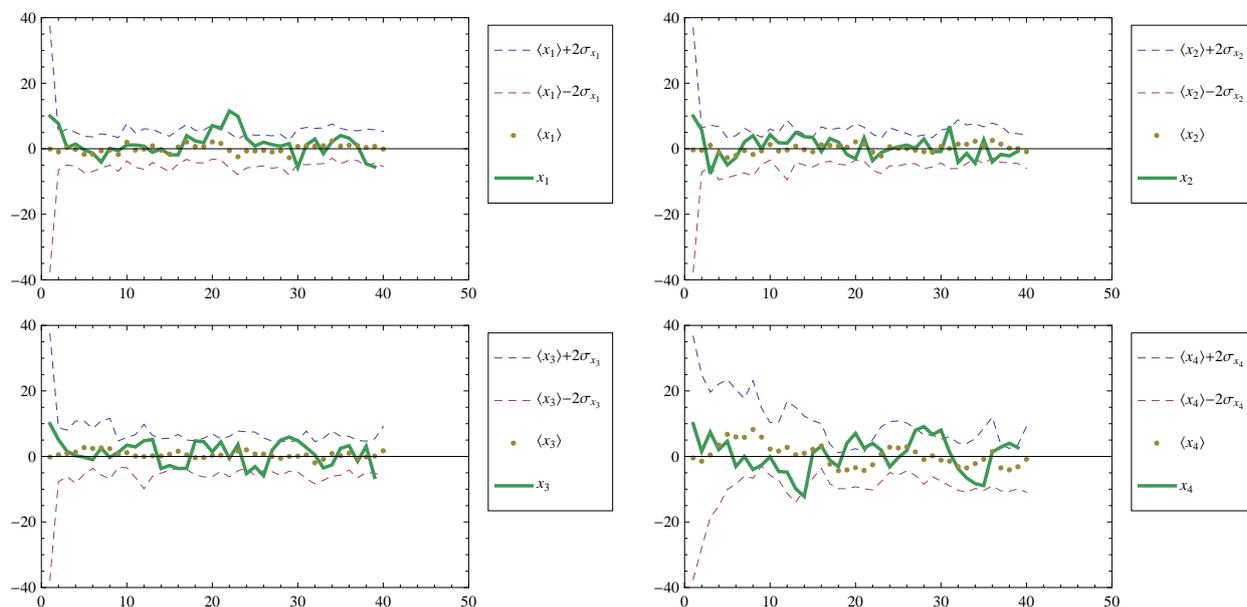
Note that the state model drift is nonlinear. In addition, observe that the diffusion vielbein is not a diagonal matrix; in fact, it is also state-dependent. Finally, note that the model is fully coupled, i.e., the state model drift and diffusion vielbein cannot be written as the direct sum of lower-dimensional objects.

The measurement model was chosen to be nonlinear as well with  $R(t) = I_{4 \times 4}$ :

$$\begin{aligned}
 \mathbf{y}_1(t) &= \sin(\mathbf{x}_1(t) + \mathbf{x}_2(t)) + \mathbf{w}_1(t) \\
 \mathbf{y}_2(t) &= \cos(\mathbf{x}_2(t) - \mathbf{x}_3(t)) + \mathbf{w}_2(t) \\
 \mathbf{y}_3(t) &= \sin(\mathbf{x}_3(t) + \mathbf{x}_4(t)) + \mathbf{w}_3(t) \\
 \mathbf{y}_4(t) &= \sin(\mathbf{x}_4(t) - \mathbf{x}_1(t)) + \mathbf{w}_4(t)
 \end{aligned}
 \tag{34}$$

Filtering was carried out using the DF algorithm. The  $r = 1/2, s = 0$  DF Lagrangian was used in the algorithm. In order to reduce the computational burden, the grid extent was chosen to be 1 (rather than 2) and the grid spacing was set at 5. The measurement time interval was 0.5, although the state was simulated at a much lower time interval. The results are shown in Figure 13. It is notable that the filtering performance is quite good. It is especially interesting since a PDE implementation would be considerably more involved with severe time-step restrictions. Observe that the measurement noise is quite large.

**Figure 13.** Filtering performance for a state sample path generated using Equation 33 and with measurement model given by Equation 34 and measurement time interval 0.5.



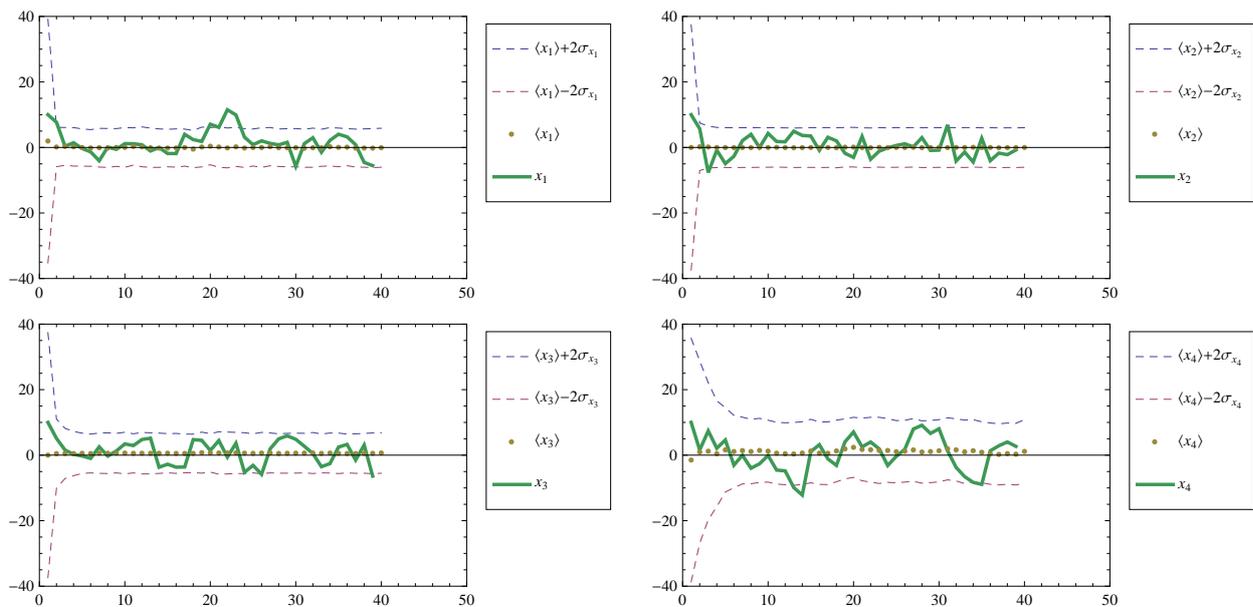
It is also possible to re-use the pre-computed transition probability tensor to solve a filtering problem with a different measurement model. This is illustrated by considering the following measurement model:

$$\begin{aligned}
 \mathbf{y}_1(t) &= c \tan^{-1} \left( \sqrt{\mathbf{x}_1^2(t) + \mathbf{x}_3^2(t)} \right) + \mathbf{w}_1(t) \\
 \mathbf{y}_2(t) &= c \cos^{-1} \left( \frac{\mathbf{x}_1(t)}{\sqrt{1 + \mathbf{x}_1^2(t) + \mathbf{x}_4^2(t)}} \right) + \mathbf{w}_2(t) \\
 \mathbf{y}_3(t) &= c \sin^{-1} \left( \frac{\mathbf{x}_4(t)}{\sqrt{1 + \mathbf{x}_3^2(t) + \mathbf{x}_4^2(t)}} \right) + \mathbf{w}_3(t) \\
 \mathbf{y}_4(t) &= c \tan^{-1} \left( 1 + \mathbf{x}_4^2(t) + \sin^2(\mathbf{x}_1(t)) \right) + \mathbf{w}_4(t)
 \end{aligned}
 \tag{35}$$

The results for the case  $c = 0.1$  for a measurement sample history are shown in Figure 14. Once again, the filter performance is quite good. Since the measurement model is different, the filtered output is different.

Of course, the same transition probability tensor cannot be used for all measurement models, such as those with large  $c$ . This is because then the measurement likelihood becomes more peaked and the grid used here becomes too coarse for adequate sampling. The resolution is to compute the transition probability tensor using a finer grid.

**Figure 14.** Filtering performance for a state sample path generated using Equation 33 and with measurement model given by Equation 35 and measurement time interval 0.5.



## 6. Additional Remarks

### 6.1. Additional Comments

It is remarkable to note that the simplest approximations to the path integral formulae leads to very accurate results. Note that the time steps are small, but not infinitesimal. Such time step sizes are not unrealistic in real world applications.

It is particularly noteworthy since it was found that SIR-PF was not a reliable solution to the studied problems. Note that the rigorous results for MC type of techniques assume that the drifts in state are bounded[20]. If that is not the case, as here, the SIR-PF is not guaranteed to work well. In any case, the speed of convergence to the correct solution is not specified for a general filtering problem, as emphasized in [21]; PFs need to be “tuned” to the problem to get desired level of performance. In fact, for discrete-time and continuous-discrete filtering problems, excellent performance also follows from a well-chosen grid using sparse tensor techniques [22,23]. Clearly, it is not axiomatic that a generic particle filter will lead to significant computation savings (or performance) over a well-chosen sparse grid method for smaller dimensional problems.

Observe also that the DF path integral filtering formulae have a simple and clear physical interpretation. Specifically, when the signal model noise is small, the transition probability is significant only near trajectories satisfying the noiseless equation. The noise variance quantifies the extent to which the state may deviate from the noiseless trajectory.

The following additional observations can be made on the (conceptually, but not necessarily computationally) simplest path integral-based filtering method proposed in this paper:

1. In an accurate solution of the universal nonlinear filtering problem, the standard deviation computed from the conditional probability density is a reliable measure of the filter performance. This is not the case for suboptimal methods like the EKF, even when the state estimate is very good. In the examples studied, it was found that the conditional standard deviation did give a more reliable indication of the actual performance of the filter.
2. The major source of computational savings follows from noting that the transition probability is given in terms of an exponential function. This implies that  $P(t'', x'' | t', x')$  is non-negligible only in a very small region, or the transition probability density tensor is sparse. The sparsity property is crucial for storage and computation speed.
3. In the examples studied, only the simplest one-step approximate formulae for the path integral expression were applied. There is a large body of work on the more accurate one-step formulae that could be used to get better results if the formulae used in this paper are not accurate enough (see, for instance, [11]).
4. Observe that higher accuracy (than the DF approximation) is attained by approximating the path integral with a finite-dimensional integral. The most efficient technique for evaluating such integrals would be to use Monte Carlo or quasi Monte Carlo methods. Another possibility is to use Monte Carlo based techniques for computing path integrals [24]. Observe that this is different from particle filtering.

5. Observe that even with coarse sampling, the computed conditional probability density is “smooth”. It seems apparent that a finer spatial grid spacing (with the same temporal grid spacing) will yield essentially the same result (using the DF approximation) at significantly higher computational cost. This was observed in the examples studied in this paper. Of course, a multiple time step approximation would be more accurate.
6. In the example studied in Section 5.1., the grid spacing was larger than the noise. Since the grid must be able to sample the probability density, the effective noise vielbein was taken to be a constant (1 in our example) times the grid spacing, i.e., the signal model noise term is “inflated”. Of course, this means that the result is not as accurate as the solution that uses the smaller values for the noise. However, it may still lead to acceptable results (as in the first example) at significantly lower computational effort.
7. Also note that the conditional mean estimation is quite good, i.e., of the order of the grid spacing, even for the coarser resolutions. This confirms the view that the conditional probability density calculated at grid points approximates very well the true value at those grid points (provided the computations are accurate). Alternatively, an interpolated version of the fundamental solution at coarser grid is close to the actual value. This suggests that a practical way of verifying the validity of the approximation is to note if the variation in the statistics with grid spacing, such as the conditional mean, is minimal.
8. It is also noted that the PDE-based methods are considerably more complicated for general two- or higher-dimensional problems. Specifically, the non-diagonal diffusion matrix case is no harder to tackle using path integral methods than the diagonal case. This is in sharp contrast to the PDE approach which for higher-dimensional problems are typically based on operator splitting approaches. The operator splitting approaches are not reliable approximations in general.
9. Observe that the one-step approximation of the path integral can be stored more compactly. Compact representation of the transition probability density, especially in the Itô case where it is of the Gaussian form. Even for the general case, the transition probability density from a certain initial point and given time step can be stored in terms of a few points with the rest obtainable via interpolation.
10. Observe that the prediction step computation was sped up considerably by restricting calculation only in areas with significant conditional probability mass (or more accurately, in the union of the region of significant probability mass of  $p(y(t_k)|x)$  and  $p(x|Y(t_{k-1}))$ ).
11. It is noted that when the DF approximation is used with larger time steps, a coarser grid is more appropriate, which requires far fewer computations. Thus, a quasi-real-time implementation could use the coarse-grid approximation with larger time steps to identify local regions where the conditional pdf is significant so that a more accurate computation can then be carried out.

12. When the step size is too large, the approximation will not be adequate. In fact, except for the  $r = 0, s = 0$  case, the positivity of the action is not guaranteed. However, unlike some PDE discretization schemes, the degradation in performance is more graceful for the  $r = s = 0$  case. For instance, positivity is always maintained since the transition probability density is manifestly positive. It is also significant to note that in physics, path integral methods are used to compute quantities where  $t' \rightarrow -\infty$  and  $t'' \rightarrow +\infty$  (see, for instance, [24]). This is not possible by simple discretization of the corresponding PDE due to time step restrictions (note that implicit schemes are not as accurate).
13. For the multiplicative noise case, the choice of  $s \neq 0$  leads to a more complicated form of the Lagrangian. The accuracy of the one-step approximation depends on  $s$  in addition to  $r$  and will be model-dependent.
14. Note that, unlike the result of S-T. Yau and Stephen Yau in [12], there is no rigorous bound on errors obtained for the Dirac-Feynman path integral formulae studied in the examples. It is known rigorously for a large class of problems that the continuum path integral formula converges to the correct solution[25].
15. It has been shown that the Feynman path integral filtering techniques also leads to new insights and reliable, practical algorithms for the general continuous-continuous nonlinear filtering problem [26,27].
16. In some instances, the fundamental solution may be computed exactly. In particular, there exists an equivalence between nonlinear filtering and Euclidean quantum mechanics that may be exploited to arrive at the exact fundamental solution valid for arbitrary time step size[28]. In that case, the only approximation is the sparse grid integration.

## 6.2. Limitations

Notwithstanding the very good performance of the proposed DF algorithm in the examples presented in the paper, it is important to note the various shortcomings:

- The approximation for the correct path integral formula with the DF approximation, which is in fact the poorest possible approximation of the path integral;
- The replacement of the integrals in the prediction step by a summation;
- Approximation of the true infinite-dimensional pdf with a finite set of grid points.

It is clear that the DF algorithm cannot be applied to the large-dimensional problems even with the use of the sparse tensors.

However, it is still important to note that the kernel of the DF algorithm is based on the Feynman path integral formula that is rigorously proven to be correct[25]. This is in contrast to some other sub-optimal algorithms (such as those based on analytical or statistical linearization and/or Gaussian assumption for the form of the conditional pdf).

### 6.3. Some Related Work

There has been prior work that can be viewed as the application of path integrals and action to filtering (e.g., [29] and [30]). Essentially, it follows from noting that the Euler approximation of the state model directly leads to the  $r = 0, s = 0$  DF approximation.

In this paper we focus on grid-based representations of the posterior. However, there are other approaches that incorporate the DF approximation in different ways and that can address the shortcomings of the DF algorithm for solving larger dimensional filtering problems.

An interesting alternative approach is Variational Filtering, which represents this density by the sample density of particles that perform a gradient descent on the Lagrangian, in generalized coordinates of motion, while being dispersed by standard Weiner fluctuations. In fact, Variational Filtering dispenses with the distinction between a prediction and correction step by absorbing the implicit likelihood term in the correction step into the Lagrangian (see [31] and [32] for details). A related scheme, called Dynamic Expectation Maximization, assumes that the conditional density is Gaussian[33]. This means that it is sufficient to estimate the trajectory of the mode, which corresponds to the path of stationary action. This provides a computationally efficient scheme but cannot represent free-form or multimodal densities.

As has been pointed out in [31], Variational Filtering offers a more robust solution to the nonlinear filtering problem than particle filtering. Since, it is also based on the Feynman path integral results and is more computationally efficient than the proposed grid-based DF algorithm, it is likely to be the algorithm of choice for many practical applications.

Another approach based on the action formed from the likelihood of the state and measurement models is studied in [34,35]. Strictly speaking, they do not do filtering, which requires the additional steps of computing the transition probability density and integration; they merely sample from the exponential of the (negative of the) action formed from the likelihoods of the state and measurement models. However, it is encouraging to note that very good performance was obtained using standard Monte Carlo methods, i.e., it was possible to generate samples efficiently that were close to the actual state.

## 7. Conclusions

In this paper, a new approach for solving the continuous-discrete filtering problem is presented. It is based on the Feynman path integral, which has been spectacularly successful in many areas of theoretical physics. The application of path integral methods to quantum field theory has also given striking insights to large areas of pure mathematics. The path integral methods has been shown to offer deep insight into the solution of the continuous-discrete filtering problem that has potentially useful practical implications. In particular, it is demonstrated via non-trivial examples that the simplest approximations suggested by the path integral formulation can yield a very accurate solution of the filtering problem. The proposed Dirac-Feynman path integral filtering algorithm is very simple, easy to implement and practical for modest size problems such as those in target tracking applications. Such formulae are also especially suitable from a real-time implementation point of view since it enables us to focus computation only on domains of significant probability mass. Furthermore, the kernel of the DF algorithm, namely the DF approximation of the path integral formula for the transition probability density, forms the basis of other elegant and potentially more computationally efficient algorithms, such as Variational Filtering.

The application of path integral based filtering algorithms for tracking problems, especially those with significant nonlinearity in the state model, will be investigated in subsequent papers.

**Acknowledgements**

The author thanks the referees for their comments and criticisms that helped improve the paper. The author is especially grateful to one of the referees for pointing out interesting references that are also related to the notions of path integrals and action. The author is grateful to Defence Research and Development Canada (DRDC) Ottawa for supporting this work under a Technology Investment Fund.

**Appendix**

**A. Summary of Path Integral Formulas**

*A.1. Additive Noise*

The additive noise model

$$d\mathbf{x}(t) = f(\mathbf{x}(t), t)dt + e(t)d\mathbf{v}(t) \tag{36}$$

is interpreted as the continuum limit of

$$\Delta\mathbf{x}(t) = f(\mathbf{x}^{(r)}(t), t)\Delta t + e(t)\Delta\mathbf{v}(t) \tag{37}$$

where

$$\mathbf{x}^{(r)}(t) = \mathbf{x}(t - \Delta t) + r(\mathbf{x}(t) - \mathbf{x}(t - \Delta t)) \tag{38}$$

Observe that any  $r \in [0, 1]$  leads to the same continuum expression.

The transition probability density for the additive noise case is given by

$$P(t'', x'' | t', x') = \int_{x(t')=x'}^{x(t'')=x''} [\mathcal{D}x(t)] \exp\left(-\int_{t'}^{t''} dt L^{(r)}(t, x, \dot{x})\right) \tag{39}$$

where the Lagrangian  $L^{(r)}(t, x, \dot{x})$  is

$$L^{(r)}(t, x, \dot{x}) = \frac{1}{2} \sum_{i=1}^n [\dot{x}_i - f_i(x^{(r)}(t), t)] g_{ij}^{-1}(t) [\dot{x}_j - f_j(x^{(r)}(t), t)] + r \sum_{i=1}^n \frac{\partial f_i}{\partial x_i}(x^{(r)}(t), t) \tag{40}$$

and  $g_{ij}(t) = \sum_{a,b=1}^{p_e} e_{ia}(t)Q_{ab}(t)e_{jb}(t)$ , and

$$[\mathcal{D}x(t)] = \frac{1}{\sqrt{(2\pi\epsilon)^n \det g(t')}} \lim_{N \rightarrow \infty} \prod_{k=1}^N \frac{d^n x(t' + k\epsilon)}{\sqrt{(2\pi\epsilon)^n \det g(t' + k\epsilon)}} \tag{41}$$

This formal path integral expression is defined as the continuum limit of

$$\frac{1}{\sqrt{(2\pi\epsilon)^n \det g(t'')}} \int \prod_{k=1}^N \left[ d^n x(t' + k\epsilon) \frac{1}{\sqrt{(2\pi\epsilon)^n \det g(t' + k\epsilon)}} \right] \exp(-S_\epsilon^{(r)}(t'', t')) \tag{42}$$

where the discretized action  $S_\epsilon^{(r)}(t'', t')$  is defined as

$$\frac{1}{2\epsilon} \sum_{k=1}^{N+1} \left[ \sum_{i,j=1}^n (x_i(t_k) - x_i(t_{k-1}) - \epsilon f_i(x^{(r)}(t_k), t_k)) g_{ij}^{-1} (x_j(t_k) - x_j(t_{k-1} + \epsilon f_j(x^{(r)}(t_k), t_k))) \right] \quad (43)$$

$$+ \sum_{k=1}^{N+1} \left[ r \sum_{i=1}^n \frac{\partial f_i}{\partial x_i}(x^{(r)}(t_k), t_k) \right]$$

and where

$$x^{(r)}(t_k) = x(t_{k-1}) + r(x(t_k) - x(t_{k-1})) \quad (44)$$

### A.2. Multiplicative Noise

Consider the evolution of the stochastic process in the time interval  $[t', t'']$ . Divide the time interval into  $N + 1$  equi-spaced points and define  $\epsilon$  by  $t' + (N + 1)\epsilon = t''$ , or  $\epsilon = \frac{t''-t'}{N+1}$ . Then, in discrete-time, the most general discretization of the Langevin equation is

$$\mathbf{x}_i(t_p) - \mathbf{x}_i(t_{p-1}) = \epsilon f_i(\mathbf{x}^{(r)}(t_p), t_p) + e_{ia}(\mathbf{x}^{(s)}(t_p), t_p)(\mathbf{v}_a(t_p) - \mathbf{v}_a(t_{p-1})) \quad (45)$$

where  $p = 1, 2, \dots, N + 1, 0 \leq r, s \leq 1$ , and

$$\begin{aligned} \mathbf{x}_i^{(r)}(t_p) &= \mathbf{x}_i(t_{p-1}) + r\Delta\mathbf{x}_i(t_p), & \mathbf{x}_i^{(s)}(t_p) &= \mathbf{x}_i(t_{p-1}) + s\Delta\mathbf{x}_i(t_p), \\ &= \mathbf{x}_i(t_{p-1}) + r(\mathbf{x}_i(t_p) - \mathbf{x}_i(t_{p-1})), & &= \mathbf{x}_i(t_{p-1}) + s(\mathbf{x}_i(t_p) - \mathbf{x}_i(t_{p-1})) \end{aligned} \quad (46)$$

In this section, the Einstein summation convention is adopted, i.e., all repeated indices are assumed to be summed over, so that  $e_{ia}d\mathbf{v}_a = \sum_{a=1}^p e_{ia}d\mathbf{v}_a$ . Also,  $\frac{\partial}{\partial x_i^{(r)}}$  is written as  $\partial_i^{(r)}$ .

Note that the change in Equation 45 when  $f_i(\mathbf{x}^{(r)}(t_p), t_p)$  and  $e_{ia}(\mathbf{x}^{(s)}(t_p), t_p)$  are replaced with  $f_i(\mathbf{x}^{(r)}(t_p), t_{p-1})$  and  $e_{ia}(\mathbf{x}^{(s)}(t_p), t_{p-1})$  is of  $O(\epsilon^2)$  and  $O(\epsilon^{3/2})$  respectively. Hence, it may be ignored in the continuum limit as it is of order higher than  $O(\epsilon)$ .

In summary, there are infinitely many possible discretizations parameterized by two reals  $r, s \in [0, 1]$ . In the continuum limit, i.e.,  $\epsilon \rightarrow 0$ , observe that the stochastic process depends on  $s$ , but not on  $r$ . When  $s = 0$ , the limiting equation is said to be interpreted as an Itô SDE, while when  $s = \frac{1}{2}$ , the equation is said to be interpreted in the Stratanovich sense.

Similarly, for the general multiplicative noise case

$$P(t'', x'' | t', x') = \int_{x(t')=x'}^{x(t'')=x''} [\mathcal{D}x(t)] \exp(-S^{(r,s)}) \quad (47)$$

where the action  $S^{(r,s)}$  is given by

$$S^{(r,s)} = \int_{t'}^{t''} dt \left[ \frac{1}{2} J_i^{(r,s)} (g^{-1})_{ij} J_j^{(r,s)} + r \partial_i^{(r)} f_i + \frac{s^2}{2} \left[ (\partial_i^{(s)} e_{ja}) Q_{ab}(t_p) (\partial_j^{(s)}) e_{ia} - (\partial_i^{(s)} e_{ia}) Q_{ab}(t) (\partial_j^{(s)} e_{ja}) \right] \right] \quad (48)$$

where

$$g_{ij} = \sum_{a,b=1}^{p_e} e_{ia}(x^{(s)}(t), t) Q_{ab}(t) e_{jb}(x^{(s)}(t), t) \quad (49)$$

and

$$J_i^{(r,s)} = \left( \frac{dx_i}{dt}(t) - f_i(x^{(r)}(t), t) - s \sum_{a,b=1}^{p_e} \sum_{i'=1}^n e_{ia}(x^{(s)}(t), t) Q_{ab}(t) \frac{\partial e_{i'b}}{\partial x_{i'}^{(s)}}(x^{(s)}(t), t) \right) \quad (50)$$

and the probability measure  $[\mathcal{D}x(t)]$  is given by

$$\frac{1}{\sqrt{(2\pi\epsilon)^n \det g(x^{(s)}(t''), t'')}} \left[ \prod_{p=1}^N \left\{ \frac{d^n x(t_p)}{\sqrt{(2\pi\epsilon)^n \det g(x^{(s)}(t_p), t_p)}} \right\} \right] \quad (51)$$

The discretized expression for the general case is complicated but can be written down from these results in a straightforward manner.

## References

1. Jazwinski, A.H. *Stochastic Processes and Filtering Theory*. Dover Publications: New York, NY, USA, 1970.
2. Kalman, R.E. A new approach to linear filtering and prediction problems. *Trans. ASME, J. Basic Eng.* **1960**, *82*, 35–45.
3. Kalman, R.E.; Bucy, R.S. New results in linear filtering and prediction problems. *Trans. ASME, J. Basic Eng.* **1961**, *83*, 95–108.
4. Daum, F. Exact finite-dimensional nonlinear filters. *Automatic Control, IEEE Transactions on* **1986**, *31*, 616–622.
5. Thomée, V. Handbook of Numerical Analysis. Ciarlet, P.G., Lions, J.L., Eds., North Holland: Amsterdam, The Netherlands, 1990, Vol. 1, Chapter Finite difference methods for linear parabolic equations, pp. 5–196.
6. Marchuk, G.I. Handbook of Numerical Analysis. Ciarlet, P.G., Lions, J.L., Eds., North Holland: Amsterdam, The Netherlands, 1990, Vol. 1, Chapter Splitting and Alternating Direction Methods, pp. 197–462.
7. Canuto, C.; Hussaini, M.Y.; Quarteroni, A.; Zang, T.A. *Spectral Methods: Fundamentals in Single Domains*. Springer-Verlag: Berlin, Germany, 2006.
8. Thomée, V. *Galerkin Finite Element Methods for Parabolic Problems*, Vol. 25, *Springer Series in Computational Mathematics*. Springer-Verlag: Berlin, Germany, 1997.
9. Feynman, R.P.; Hibbs, A.R. *Quantum Mechanics and Path Integrals*. McGraw-Hill Book Company: New York, NY, USA, 1965.
10. Zinn-Justin, J. *Quantum Field Theory and Critical Phenomena*. International Series in Monographs on Physics. Oxford University Press: New York, NY, USA, 2002.
11. Langouche, F.; Roekaerts, D.; Tirapegui, E. *Functional Integration and Semiclassical Expansions*. Reidel: Dordrecht, The Netherlands, 1982.
12. Yau, S.T.; Yau, S.S.T. Explicit Solution of a Kolmogorov Equation. *Appl. Math. Opt.* **1996**, *34*, 231–266.
13. Balaji, B. Feynman Path Integrals and continuous-discrete filtering: The additive noise case. Technical Report TM 2008-343, 2009.

14. Bader, B.W.; Kolda, T.G. Algorithm 862: MATLAB tensor classes for fast algorithm prototyping. *ACM Trans. Math. Software* **2006**, *32*, 635–653.
15. Bader, B.W.; Kolda, T.G. Efficient MATLAB computations with sparse and factored tensors. Technical Report SAND2006-7592, 2006.
16. Gordon, N.; Salmond, D.; Smith, A. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proc.-F Rad. Sig. Proc.* **1993**, *140*, 107–113.
17. Budhiraja, A.; Chen, L.; Lee, C. A survey of numerical methods for nonlinear filtering problems. *Physica D* **2007**, *230*, 27–36.
18. Chen, L.; Lee, C.; Bidhiraja, A.; Mehra, R.K. PFLib —An object oriented MATLAB toolbox for particle filtering. In *Proceedings to SPIE Signal Processing, Sensor Fusion, and Target Recognition XVI*, 2007, Vol. 6567.
19. Lototsky, S.V.; Rozovskii, B.L. Recursive nonlinear filter for a continuous discrete-time model: separation of parameters and observations. *IEEE Transactions on Automatic Control* **1998**, *43*, 1154–1158.
20. Crisan, D.; Doucet, A. A survey of convergence results on particle filtering methods for practitioners. *Signal Processing, IEEE Transactions on [see also Acoustics, Speech, and Signal Processing, IEEE Transactions on]* **2002**, *50*, 736–746.
21. Daum, F.; Huang, J. Curse of dimensionality and particle filters. In *IEEE Aerospace Conference*, 2003, Vol. 4.
22. Balaji, B. Sparse Tensors and Discrete-Time nonlinear filtering. In *IEEE Radar Conference*. 2008, pp. 1785–1790.
23. Balaji, B. Continuous-Discrete Filtering using the Dirac-Feynman Algorithm. In *IEEE Radar Conference*, Rome, Italy, 2008.
24. Montváy, I.; Münster, G. *Quantum Fields on a Lattice*. Cambridge Monographs on Mathematical Physics. Cambridge University Press: New York, USA, 1997.
25. Alicki, R.; Makowiec, D. Functional integrals for parabolic differential equations. *J. Phys. A: Math. Gen.* **1985**, *18*, 3319–3325.
26. Balaji, B. Universal Nonlinear Filtering Using Path Integrals II: The Continuous-Continuous Model with Additive Noise. *PMC Physics A* **2009**.
27. Balaji, B. Estimation of indirectly observable Langevin states: path integral solution using statistical physics methods. *J. Stat. Mech.-Theory Exp.* **2008**, *2008*, P01014:1–P01014:17.
28. Balaji, B. Euclidean Quantum Mechanics and Universal Nonlinear Filtering. *Entropy* **2009**, *11*, 42–58.
29. Whittle, P. Likelihood and Cost as Path Integrals. *J. Roy. Stat. Soc. Ser. B* **1991**, *53*, 505–538.
30. Archambeau, C.; Cornford, D.; Lawrence, D.; Schwaighofer, A.; Quinero, J. Gaussian process approximations of stochastic differential equations. In *Journal of Machine Learning Research Workshop and Conference Proceedings*, 2007.
31. Friston, K. Variational filtering. *NeuroImage* **2008**, *41*, 747–766.
32. Friston, K.; Mattout, J.; Trujillo-Barreto, N.; Ashburner, J.; Penny, W. Variational free energy and the Laplace approximation. *NeuroImage* **2007**, *34*, 220–234.
33. Friston, K.; Trujillo-Barreto, N.; Daunizeau, J. DEM: A variational treatment of dynamic systems. *NeuroImage* **2008**, *41*, 849–885.

34. Alexander, F.J.; Eynik, G.L.; Restrepo, J.M. Accelerated Monte Carlo for Optimal Estimation of Time Series. *J. Statist. Phys.* **2005**, *119*, 1331–1345.
35. Restrepo, J.M. A path integral method for data assimilation. *Physica D* **2008**, *237*, 14–27.

© 2009 by the author; licensee Molecular Diversity Preservation International, Basel, Switzerland. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license <http://creativecommons.org/licenses/by/3.0/>.