Universal Nonlinear Filtering using Feynman Path Integrals I: The Continuous-Discrete Model with Additive Noise

BHASHYAM BALAJI
Defence Research and Development Canada

The continuous-discrete filtering problem requires the solution of a partial differential equation known as the Fokker-Planck-Kolmogorov forward equation (FPKfe). The path integral formula for the fundamental solution of the FPKfe is derived and verified for the general additive noise case (i.e., explicitly time-dependent state model and with state-independent rectangular diffusion vielbein). The solution is universal in the sense that the initial distribution may be arbitrary. The practical utility is demonstrated via some examples.

I. INTRODUCTION

In a wide variety of applications, the evolution of a state, or a signal of interest, is described by a stochastic dynamical model. That is, the state of the system is described by a noisy version of a deterministic dynamical system termed the state model. If the state process is continuous (discrete), then the dynamics are governed by a system of first-order differential equations (difference equations) in the state variable ($\mathbf{x}(t)$) with an additional contribution due to random noise ($\nu(t)$). The noise in the state model is referred to as the signal noise.

However, in many applications the signal model cannot be directly observed. Instead, what is measured is a nonlinearly related stochastic process ($\mathbf{y}(t)$) called the measurement process. The measurement process can often by modelled as yet another stochastic dynamical system called the measurement model. That is, the observations, or measurements, are drawn from a different system of noisy first-order differential (difference) equations. The noise in the measurement dynamical system is referred to as measurement noise. The nonlinear filtering problem is to estimate the state of a stochastic dynamical system, given the observations of a related stochastic measurement process. For an excellent discussion of the subject, see Jazwinski [1].

The conditional probability density function of the state vector, given the observations, is the complete solution of the filtering problem. It contains all the probabilistic information about the state process that is in the measurements and the initial condition. Given the conditional probability density, optimality may be defined under various criteria. 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.

The signal and measurement processes may be discrete-time or continuous-time stochastic processes and there is a different type of filtering problem for each possible combination of continuous (discrete) signal and measurement processes. In this paper the measurement process is described by a discrete-time stochastic process and the underlying signal process is a continuous-time stochastic process.

The linear/Gaussian system was investigated by Kalman and Bucy [2], [3]. For a modern review of the subject, see [4]. The Kalman filter has been successfully applied to a large number of problems.

In spite of its enormous success, the Kalman filter suffers from some major limitations. From a Bayesian perspective, the simplicity of the Kalman filter for the linear state model is because it is based on the assumption that the system is linear and Gaussian. Then, the probability distribution function is

completely characterized by the mean and covariance matrices.[1] The Kalman filter is still optimal if the initial distribution is not Gaussian only under certain criteria, such as minimum variance, but not under an arbitrary loss function. In other words, the Kalman filter is not a universal optimal filter even when the filtering problem is linear. In the general nonlinear case, the filter state is infinite dimensional, i.e., the state cannot be specified in terms of a finite number of parameters. The Kalman filter cannot be optimal and will fail, although it may still work quite well if the nonlinearity is mild enough.

There are other methods, such as the unscented Kalman filter that improve upon the extended Kalman filter (EKF) and can result in acceptable performance in some cases [5]. However, they cannot be considered to be reliable in solving the general nonlinear filtering problem since it also is based on the first two moments of the distribution. It is also noted that sometimes it may be possible to reduce the solution of partial differential equations (PDEs) to that of the solution of ordinary differential equations (ODEs). This is the case for the "finite-dimensional filters," where the conditional probability density can be expressed in terms of a finite number of parameters [6, 7]. The Kalman filter is a special such filter. Again, this covers a very limited class of problems.

It is well known that the evolution of the probability distribution of the state variable is governed by the Fokker-Planck-Kolmogorov equation (FPKfe). The solution of the continuous-discrete filtering problem also requires the solution of an FPKfe, since the probability density between measurements evolves according to the FPKfe (see, for instance, [1]). The FPKfe is a linear, parabolic, PDE.[2] The literature on methods to solve the FPKfe type of equations is vast. In a broad sense, there are three main types of methods: finite difference methods [8, 9], spectral methods [10], and finite element methods [11]. In fact, some papers in continuous-discrete filtering theory have already studied use of such methods [12, 13].

However, there are several difficulties in solving PDEs. 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, it may be that the numerical method is unstable. Finally, since the solution of an FPKfe is a probability density, it must be positive. This may not be guaranteed by the discretization of the FPKfe.

Another technique for tackling the filtering problem is the particle filter which is a simulation-based, computationally efficient method investigated in recent years [14].The simplest variant is quite easy to implement and often shown to work satisfactorily in some cases. However, there is no precise quantitative understanding of the number of particles needed to attain a certain level of performance. In fact, the error bounds derived to date depend on the boundedness of the functions in the model. It is also highly sensitive to the choice of proposal distribution [15]; sampling from the optimal proposal distribution is possible only for a few cases [16].

A different approach was taken by Shing-Tung Yau and Stephen Yau in [17]. A more general equation than the FPKfe, one which arises in the continuous-continuous nonlinear filtering problem with additive noise (and with no explicit time dependence), was studied in [17]. It was shown that the formal solution may be written as an ordinary, but somewhat complicated, multi-dimensional integral, where the integrand is an infinite series. In addition, they also presented an estimate of the time needed for the solution to converge to the true solution. Since this FPKfe is independent of the measurements, it can be solved off-line.

In this paper it is shown that the fundamental solution of the FPKfe, and hence the continuous-discrete filtering problem, can be solved in terms of a Feynman path integral for the general additive noise state model. Specifically, the transition probability probability density is expressed as a Feynman path integral. Thus, the solution of the FPKfe is reduced to the computation of a path integral. Although the path integral is the limit of a large-dimensional integral, it is seen that the path integrand is considerably simpler than the integrand in [17] (e.g., no convolution or infinite series).

Path integrals first arose in the study of quantum physics. Inspired by an observation/remark of Dirac [18,] [19], Feynman discovered the path integral representation of quantum physics [20], [21]. Until the mid 1960s, path integrals were thought to be merely a curiosity, and not a serious alternative to the operator methods in quantum physics. The importance of path integrals grew when Fadeev and Popov used the path integrals to derive Feynman rules for the non-Abelian gauge theories. Its status was further enhanced when Hooft and Veltman used path integral methods to prove that the Yang-Mills theories were renormalizable. Since then, the path integral formulation of quantum field theory has led to extraordinary insights into quantum field theory, such as renormalization and renormalization group, anomalies, skyrmions, monopoles, instantons, and supersymmetry (see, for instance, the freely available text [22]). Much of the advance in the

---

[1] Such a simple characterization of the probability distribution and solution is possible only for a limited class of filtering problems called finite-dimensional filters.

[2] Sometimes in the literature the term "nonlinear FPKfe" is used to denote the FPKfe corresponding to a state process with a nonlinear drift. However, this FPKfe is still a linear, parabolic PDE.

past 40 years in particle theory would not have been possible without the path integral. The path integral has also led to numerous insights into mathematics, such as the quantum field theory representation of the knots invariants, instantons, and monopoles in supersymmetric gauge theories and the Donaldson-Seiberg-Witten invariants of four-manifolds, and more recently, S-duality in supersymmetric gauge theories and the Laglands program in mathematics. For a rigorous account of some aspects of path integral methods, see [23].

The path integral representation is especially interesting from a computational point of view. A classic example is lattice quantum chromodynamics (QCD) which is the lattice formulation of the quantum field theory of strong interactions. Observables in quantum field theory require path integration over quantum fields whose dynamics are described by highly nonlinear infinite-dimensional systems. The path integral method has been used to perform the nonperturbative computation of quantities in QCD (see, for instance, [24]). The comparison to experimental results, even at coarse resolutions, with lattice computations is excellent.

The path integral solution of the FPKfe has been studied in various branches of theoretical physics for different purposes. This is because the FPKfe (and the Langevin equation) is a fundamental equation of statistical physics. A textbook discussion of the FPKfe is presented in [25], [26]. The results derived in this paper generalize some of the results in [26]. Specifically, the path integral formula for the fundamental solution is derived and independently verified for the case when the drift in the state model is explicitly time dependent, the diffusion vielbein is an explicitly time-dependent rectangular matrix, and when the noise process is colored with a time-dependent covariance matrix. In addition, the discussion here improves upon the discussion in [26] by clarifying important and subtle aspects of the derivation of the path integral formulas and by providing more details so that this powerful technique is accessible to a wider audience.

The outline of this paper is as follows. In Section II the Langevin equation, the corresponding FPKfe, and the role the FPKfe plays in solving the continuous-discrete filtering problem, are reviewed. A quick and simple derivation of the simplest path integral formula is presented in Section III. In Section IV the meaning of the Gaussian path integral measure is clarified and some important properties of the Langevin equation are reviewed. In Section V two derivations are presented of the path integral formula for the fundamental solution for the simpler case; i.e., where the drift and diffusion vielbein are not explicitly time dependent and the noise is white and additive. This is likely to be the case of most interest in practice. In Section VI this is generalized

to the case when the drift and diffusion vielbein are explicitly time dependent and the additive noise is colored. In Section VII it is independently verified that the derived path integral formula does indeed satisfy the FPKfe. In the following section some numerical examples are presented. This is followed by a discussion of the physical meaning of the path integral formula and a comparison to the solution with that obtained by S-T. Yau and S. S.-T. Yau.

In the Appendices the tools used in the derivations are reviewed including Gaussian integration, the imposition of delta function constraints as an integral, and aspects of functional calculus. Although most of these topics are discussed in modern quantum field theory texts, it has been included in order to make the paper completely self-contained and accessible to a wider audience.

II. CONTINUOUS-DISCRETE FILTERING: A REVIEW

A. The Fokker-Planck-Kolmogorov Forward Equation

In continuous-discrete filtering theory, the state model is given by the stochastic differential equation of the form

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

Here $\mathbf{x}(t)$ and $f(\mathbf{x}(t),t)$ are n-dimensional column vectors, $e(\mathbf{x}(t),t)$ is an $n \times p$ matrix and $\mathbf{v}(t)$ is a p-dimensional Wiener process column vector. In physics literature, (1) is written as follows:

$$\frac{d\mathbf{x}(t)}{dt} = f(\mathbf{x}(t),t) + e(\mathbf{x}(t),t)\boldsymbol{\nu}(t). \qquad (2)$$

The stochastic process $\boldsymbol{\nu}(t)$ is assumed to be Gaussian with zero mean and "δ-correlated," i.e.,

$$\langle \boldsymbol{\nu}(t) \rangle = 0$$
$$\langle \boldsymbol{\nu}(t)\boldsymbol{\nu}^T(t') \rangle = Q(t)\delta(t - t') \qquad (3)$$

where $Q(t)$ is a $p \times p$ covariance matrix. The quantity f is referred to as the (signal model) drift and e as the diffusion vielbein, and the quantity eQe^T is referred to as the diffusion matrix. The stochastic processes are represented in bold and one of its samples by its corresponding plain symbol. In the interest of clarity, the tensor indices are suppressed in this subsection. The expectation with respect to Gaussian noise is denoted by angular brackets. Equation (2) is also referred to as the Langevin equation with multiplicative noise and δ-correlated Langevin force. When $e = e(t)$, i.e., the diffusion vielbein is independent of the state variable, the noise in the Langevin equation is termed additive noise.

Consider an ensemble of dynamical systems with state variables evolving according to the Langevin equation. Due to random noise, each system leads to a different vector $x(t)$ that depends on time. Although only one realization of the stochastic process is observed, it is meaningful to consider an ensemble

average. For fixed times $t = t_i$, $i = 1, 2, \ldots, r$, the probability density of finding the random vector $\mathbf{x}(t)$ in the (n-dimensional) interval $x_i \le \mathbf{x}(t_i) \le x_i + dx_i (1 \le i \le r)$ is given by

$$W_r(t_r, x_r; \ldots; t_1, x_1) = \left\langle \prod_{i=1}^{r} \delta^n(\mathbf{x}(t_i) - x_i) \right\rangle \quad (4)$$

where x_i is an n-dimensional column vector. 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[3]

$$P(t_r, x_r \mid t_{r-1}, x_{r-1}, \ldots; t_1, x_1)$$
$$= \langle \delta^n(\mathbf{x}(t_r) - x(t_r)) \rangle |_{x(t_{r-1}) = x_{r-1}, \ldots, x(t_1) = x_1}, \quad x(t_r) \equiv x_r$$
$$= \frac{W_r(t_r, x_r; \ldots; t_1, x_1)}{\int W_r(t_r, x_r; \ldots; t_1, x_1) d^n x_r}. \quad (5)$$

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:

$$P(t_r, x_r \mid t_{r-1}, x_{r-1}; \ldots; t_1, x_1) = P(t_n, x_n \mid t_{n-1}, x_{n-1}). \quad (6)$$

This implies that

$$W_r(t_r, x_r; \ldots; t_1, x_1) = \left[\prod_{i=1}^{r-1} P(t_{i+1}, x_{i+1} \mid t_i, x_i) \right] W_1(t_1, x_1). \quad (7)$$

Hence, the complete information for a Markov process is contained in the transition probability densities

$$P(t_2, x_2 \mid t_1, x_1) = \frac{W_2(t_2, x_2; t_1, x_1)}{W_1(t_1, x_1)}. \quad (8)$$

A fundamental property is the Chapman-Kolmogorov semi-group property of the transition probability density:

$$P(t_3, x_3 \mid t_1, x_1) = \int P(t_3, x_3 \mid t_2, x_2) P(t_2, x_2 \mid t_1, x_1) \{d^n x_2\}. \quad (9)$$

This property plays a fundamental role in arriving at the path integral formula.

It can be shown that the state probability distribution function $p(t, x) = \int P(t, x \mid t'x') p(t', x') \{d^n x'\}$ satisfies the FPKfe (see for instance [25]):

$$\frac{\partial p}{\partial t}(t, x) = -\sum_{i=1}^{n} \frac{\partial}{\partial x_i}[f_i(x(t), t)p(t, x)] + \frac{1}{2} \sum_{i,k=1}^{n} \sum_{b=1}^{p} \frac{\partial}{\partial x_i}$$
$$\times \left[(e(x(t), t)Q(t))_{ib} \frac{\partial}{\partial x_k}(e_{bk}^T(x(t), t)p(t, x)) \right]$$
$$\equiv \mathcal{L} p(t, x). \quad (10)$$

---

[3]Unless otherwise specified, all integration variables are from $-\infty$ to ∞.

The diffusion matrix is clearly a symmetric matrix and is also positive semi-definite: $D_{ij} a_i a_j \ge 0$. Often, D_{ij} is assumed to be positive-definite, in which case the inverse of the diffusion matrix exists. An important property of the FPKfe is that it is a continuity equation, i.e., the right-hand side a total divergence.

B. Continuous-Discrete Filtering: Model and Solution

In this paper it is assumed that the dynamics are given by a continuous-time process and observations are samples of a discrete-time process. The continuous-time state model is described by (1). The measurement model is described by the following discrete-time stochastic process

$$\mathbf{y}(t_k) = h(\mathbf{x}(t_k), \mathbf{w}(t_k), t_k), \quad k = 1, 2, \ldots, \quad t_{k+1} > t_k \ge 0 \quad (11)$$

where $y(t) \in \mathbb{R}^{m \times 1}$, $h \in \mathbb{R}^{m \times 1}$ and the noise process is described by $\mathbf{w}(t) \sim N(0, R(t))$. It is assumed that $p(y(t_k) \mid x)$ is known.

Let the initial distribution be $\sigma_0(x)$ and let the measurements be collected at time instants $t_1, t_2, \ldots, t_k, \ldots$. We use the notation $Y(\tau) = \{y(t_l) : t_0 < t_l \le \tau\}$. Prior to incorporating the measurements, the state evolves according to the FPKfe, i.e.,

$$\frac{\partial p}{\partial t}(t, x \mid Y(t_0)) = \mathcal{L}(p(t, x \mid Y(t_0))), \quad t_0 < t \le t_1$$
$$p(t_0, x \mid Y(t_0)) = \sigma_0(x). \quad (12)$$

This is the "prediction" step.

From Bayes' rule (and using $p(t_i, x \mid Y(t_i)) = p(t_i, x \mid y(t_i), Y(t_{i-1}))$) at observation t_1, the "corrected" conditional density at time t_1 is

$$p(t_1, x \mid Y(t_1)) = \frac{p(y(t_1) \mid x) p(t_1, x \mid Y(t_0))}{\int p(y(t_1) \mid \xi) p(t_1, \xi \mid Y(t_0)) \{d^n \xi\}}. \quad (13)$$

This is then the initial condition of the FPKfe for the next prediction step which results in

$$\frac{\partial p}{\partial t}(t, x \mid Y(t_1)) = \mathcal{L}(p(t, x \mid Y(t_1))), \quad t_1 < t \le t_2,$$

(Prediction Step) (14)

$$p(t_2, x \mid Y(t_2)) = \frac{p(y(t_2) \mid x) p(t_2, x \mid Y(t_1))}{\int p(y(t_2) \mid \xi) p(t_2, \xi \mid Y(t_1)) \{d^n \xi\}}$$

(Correction Step)

and so on. Note that the measurement model need not be Gaussian.

Thus, the nontrivial part of continuous-discrete filtering is the solution of an FPKfe. Note that the FPKfe is independent of measurements, and hence can be computed off-line. Furthermore, observe that the complete information is in the transition probability density which also satisfies the FPKfe except with a

δ-function initial condition (known as the fundamental solution, or kernel, of the FPKfe):

$$\frac{\partial}{\partial t}P(t,x\mid\tau,y) = \mathcal{L}(P(t,x\mid\tau,y)), \qquad t > \tau$$
$$P(t,x\mid t,y) = \delta(x-y).$$
(15)

This is because from the transition probability density the probability for an arbitrary initial condition can be computed as follows:

$$p(t_n,x) = \int P(t_n,x\mid t_{n-1},x')p(t_{n-1},x')\{d^n x'\}. \quad (16)$$

Thus, it is sufficient to solve the FPKfe for $P(t,x\mid\tau,y)$ and use the above to compute $p(t_k,x\mid Y(t_{k-1}))$ using

$$p(t_k,x\mid Y(t_{k-1}))$$
$$= \int P(t_k,x\mid t_{k-1},x')p(t_{k-1},x'\mid Y(t_{k-1}))\{d^n x'\}.$$
(17)

Alternatively, a convenient set of basis functions may be used. Then, the evolution of each of the basis functions under the FPKfe follows from (16). Since the basis functions are independent of measurements, the computation may be performed off-line [27]. Note that the solution is universal, i.e., the initial distribution can be arbitrary.

In conclusion, the solution of (15) is equivalent to the solution of the universal optimal nonlinear filtering problem. A solution in terms of ordinary integrals was presented in [17]. Here, the solution in terms of Feynman path integrals is presented.

III. FEYNMAN PATH INTEGRAL FORMULA. A PHYSICAL DERIVATION

In this section a physical derivation of the simplest form of the path integral formula is presented. The more general form is derived and verified in greater detail in later sections.

For simplicity consider the following model:

$$d\mathbf{x}(t) = f(\mathbf{x}(t),t)dt + d\mathbf{v}(t). \tag{18}$$

We are interested in the quantity $P(t,x\mid t_0,x_0)$. To this end, consider the quantity $P(t_0+\Delta t,x(t_0+\Delta t)\mid t_0,x_0)$. Now from the state model it follows that for "small" Δt

$$x(t_0+\Delta t) \approx x(t_0) + f(x(t_0),t_0)\Delta t + \Delta\mathbf{v}(t).$$

Now $\mathbf{v}(t)$ is a standard Wiener process, which means that $\Delta\mathbf{v}(t)$ is a Gaussian noise process $N(0,\Delta t)$. This implies that

$$P(t_0+\Delta t_0,x(t_0+\Delta t)\mid t_0,x(t_0))$$
$$\approx \frac{1}{\sqrt{2\pi\Delta t}}\exp\left(-\frac{1}{2\Delta t}(x(t_0+\Delta t)-x(t_0)-\Delta t f(x(t_0),t_0))^2\right).$$
(19)

Next, consider the quantity $P(t_0+2\Delta t,x(t_0+\Delta t)\mid t_0, x(t_0))$. Using the Chapman-Kolmogorov property, it follows that

$$P(t_0+2\Delta t,x(t_0+2\Delta t)\mid t_0,x(t_0))$$
$$\approx \int \{d^n x(t_0+\Delta t)\}$$
$$\times P(t_0+2\Delta t,x(t_0+2\Delta t)\mid t_0+\Delta t,x(t_0+\Delta t))$$
$$\times P(t_0+\Delta t_0,x(t_0+\Delta t)\mid x(t_0+\Delta t_0)\mid t_0,x(t_0))$$
$$\approx \frac{1}{\sqrt{(2\pi\Delta t)^2}}\int \{d^n x(t_0+\Delta t)\}$$
$$\times \exp\left(-\frac{1}{2\Delta t}\sum_{i=1}^{2}(x(t_0+i\Delta t)-x(t_0+(i-1)\Delta t)\right.$$
$$\left. -\Delta t f(x(t_0+(i-1)\Delta t)))^2\right)$$
$$\approx \frac{1}{\sqrt{(2\pi\Delta t)^2}}\int \{d^n x(t_0+\Delta t)\}$$
$$\times \exp\left(-\frac{1}{2}\Delta t\sum_{i=1}^{2}\left(\frac{x(t_0+i\Delta t)-x(t_0+(i-1)\Delta t)}{\Delta t}\right.\right.$$
$$\left.\left. -f(x(t_0+(i-1)\Delta t))\right)^2\right).$$
(20)

Likewise, it follows that if $t_0+N\Delta t\equiv t$, then

$$P(t,x(t)\mid t_0,x(t_0))$$
$$= \frac{1}{\sqrt{(2\pi\Delta t)^N}}\int\left\{\prod_{i=1}^{N-1}d^n x(t_0+i\Delta t)\right\}$$
$$\times \exp\left(-\frac{1}{2}\Delta t\sum_{i=1}^{N}\left(\frac{x(t_0+i\Delta t)-x(t_0+(i-1)\Delta t)}{\Delta t}\right.\right.$$
$$\left.\left. -f(x(t_0+(i-1)\Delta t))\right)^2\right).$$
(21)

This approximation leads to the correct answer when $\Delta t \to 0$, and it is formally written as

$$P(t,x(t)\mid t_0,x(t_0))$$
$$= \int_{x(t_0)=x_0}^{x(t)=x}[\mathcal{D}x(t)]\exp\left(-\frac{1}{2}\int_{t_0}^{t}dt(\dot{x}(t)-f(x(t)))^2\right)$$
$$\equiv \int_{x(t_0)=x_0}^{x(t)=x}[\mathcal{D}x(t)]\exp(-S(t,t_0)). \tag{22}$$

This is a Feynman path integral formula for the transition probability density. Recall that in the derivation above the Chapman-Kolmogorov semigroup property played a fundamental role.

Following Feynman, there is an alternative beautiful physical picture that is as follows. The Feynman path integral formula is formed by the sum over all paths satisfying the initial and final condition with each path weighted by the action $S(t, t_0)$. There is a subtlety in that it is the nondifferentiable paths that contribute since $(\Delta x)^2 \approx \Delta t$. However, this appealing physical picture has proven to be useful in obtaining new results in many areas of theoretical physics.

In the subsequent sections, we note the following.

1) There is a more general expression of the Feynman path integral formula parametrized by a real r. The expression derived here corresponds to the case $r = 0$.

2) The case $r = 1/2$ is special in the sense that it can be derived direclty using the so-called functional methods that are reminiscent of ordinary calculus techniques.

3) It is verified that the derived path integral formula is the fundamental solution of the FPKfe.

IV. SOME PRELIMINARY REMARKS

The reader may refer to the Appendices for the derivation of the results used here.

A. Gaussian Path Integral Measure

The Gaussian noise process $\nu(t)$, defined in (3), may be represented by a path integral measure

$$[d\rho(\nu(t))] \equiv [\mathcal{D}\nu(t)] \exp\left[-\frac{1}{2}\int dt \sum_{a,b=1}^{p} \nu_a(t)(Q^{-1}(t))_{ab}\nu_b(t)\right],$$

$$\nu \in \mathbb{R}^{p \times 1} \quad (23)$$

where $\nu(t)$ is a real vector for each t. The meaning of (23) is explained next.

Recall that the probability measure of an n-dimensional Gaussian vector ν with zero mean and covariance matrix Q is

$$\left\{\frac{1}{\sqrt{\det Q}}\left(\frac{1}{2\pi}\right)^{p/2}\{d^p\nu\}\right\}\exp\left(-\frac{1}{2}\sum_{a,b=1}^{p}\nu_a Q^{-1}_{ab}\nu_b\right).$$

$$(24)$$

For a sequence of random vectors uncorrelated with each other at different times

$$\int\left\{\prod_{k=1}^{N}\frac{1}{\sqrt{(2\pi)^p\det(Q(t_k))}}\{d^p\nu(t_k)\}\right\}\nu_c(t_k)\nu_d(t_l)$$

$$\times \exp\left(-\frac{1}{2}\sum_{k=1}^{N}\sum_{a,b=1}^{p}\nu_a(t_k)Q^{-1}_{ab}(t_k)\nu_b(t_k)\right)$$

$$= Q_{ab}(t_k)\delta_{kl}. \quad (25)$$

Note that the noise may be colored. It is also important to note that had uncorrelatedness at different times not been assumed, the exponent would not be a sum of quadratic terms local in time.

For the continuum limit, the following expression is of interest (in the notation of Appendix I, $\alpha(t) = \frac{1}{2}$, $\beta = \Delta t$)

$$\int\left\{\prod_{k=1}^{N}\frac{1}{\sqrt{(2\pi\Delta t)^p\det(Q(t_k))}}\{d^px(t_k)\}\right\}x_c(t_k)x_d(t_l)$$

$$\times \exp\left(-\frac{\Delta t}{2}\sum_{k=1}^{N}\sum_{a,b=1}^{p}x_a(Q^{-1}(t_k))_{ab}x_b\right)$$

$$= \frac{1}{\Delta t}Q_{ab}(t_k)\delta_{kl}. \quad (26)$$

The measure in the continuum limit (i.e., $\Delta t \to 0$) is formally written as

$$[d\rho(\nu(t))]$$

$$\equiv [\mathcal{D}\nu(t)]\exp\left(-\frac{1}{2}\sum_{a,b=1}^{p}\int \nu_a(t)(Q^{-1}(t))_{ab}\nu_b(t)dt\right)$$

$$(27)$$

where

$$[\mathcal{D}\nu(t)] \equiv \lim_{N\to\infty}\prod_{k=1}^{N}\left\{\frac{1}{\sqrt{(2\pi\Delta t)^p\det Q(t_k)}}\{d^p\nu(t_k)\}\right\}.$$

$$(28)$$

Hence, in the continuum limit, the mean is easily seen to be 0 and (26) becomes

$$\langle\nu_a(t)\nu_b(t')\rangle = Q_{ab}(t)\delta(t-t'). \quad (29)$$

These results may also be directly derived using the methods of functional calculus (see Appendix III). In particular it is straightforward to verify that the mean is zero due to odd symmetry (setting $\alpha = 1/2$)

$$\langle\nu_c(t)\rangle = \int[d\rho(\nu)]\nu_c(t)$$

$$= \int[\mathcal{D}\nu(t)]\nu_c(t)$$

$$\times \exp\left(-\frac{1}{2}\int dt\sum_{a,b=1}^{p}\nu_a(t)(Q^{-1}(t))_{ab}\nu_b(t)\right)$$

$$= 0 \quad (30)$$

and the variance is $Q(t)$:

$$\langle \boldsymbol{\nu}_c(t)\boldsymbol{\nu}_d(t')\rangle = \int [d\rho(\nu)]\nu_c(t)\nu_d(t')$$

$$= \int [d\nu(t)]\nu_c(t)\nu_d(t')\exp\left(-\frac{1}{2}\int dt \sum_{a,b=1}^{p} \nu_a(t)(Q^{-1}(t))_{ab}\nu_b(t)\right)$$

$$= \frac{\delta}{\delta b_c(t)}\frac{\delta}{\delta b_d(t)}\exp\left(-\frac{1}{2}\int dt' \sum_{a,b=1}^{p} b_a(t')Q_{ab}(t')b_b(t')\right)\Bigg|_{b=0}$$

$$= Q_{cd}(t)\delta(t-t') \tag{31}$$

and higher order moments can easily be written down using Wick's theorem. The results are in accord with the expectation that the measure represents a Gaussian process with these first two moments.

B. Basic Properties of the Langevin Equation

We conclude this section with some fundamental properties of the Langevin equation.

1) *Time translation invariance:* When the drift does not depend on time explicitly, i.e., $f(\mathbf{x}(t))$, then the process is time-translation invariant, i.e., it does not depend on the origin of time. Alternatively, the time translation operator $\partial/\partial t$ commutes with the FPK operator $\mathcal{L}$:

$$\left[\frac{\partial}{\partial t},\mathcal{L}\right]\Phi = \frac{\partial}{\partial t}\mathcal{L}\Phi - \mathcal{L}\frac{\partial}{\partial t}\Phi = 0 \tag{32}$$

where Φ is an arbitrary differentiable function.

2) *Locality:* Langevin equation is a differential equation specifying the time evolution as a system of first-order stochastic differential equations with all terms evaluated at the same time. In other words, the Langevin equation is local in time.

3) *Noise uncorrelated at different times:* This is an important assumption which leads to the δ-correlated in time property of the noise process.

From the locality of the Langevin equation and the uncorrelatedness of noise at different times, it immediately follows that P satisfies the Chapman-Kolmogorov semi-group property. This implies that the distribution P is completely determined from its knowledge for small time intervals. It is important to note that if the Langevin equation was not local and/or the noise was not uncorrelated at different times, the Chapman-Kolmogorov semi-group property would not be valid.[4]

---

[4]However, if noise is correlated at different times, some non-Markovian processes (for instance, exponentially correlated) may be reduced to Markovian processes.

V. PATH INTEGRAL FORMULA I. IMPLICIT TIME DEPENDENCE

In this section, the noise is assumed to be additive, white, and Gaussian distributed. The state model is given by

$$d\mathbf{x}(t) = f(\mathbf{x}(t))dt + d\mathbf{v}(t), \qquad \mathbf{x}(t), f(\mathbf{x}(t)), \mathbf{v}(t) \in \mathbb{R}^n,$$

$$Q_{ij} = \hbar_\nu \delta_{ij}. \tag{33}$$

Observe that the drift is not explicitly time dependent, and the diffusion vielbein is proportional to the identity matrix with proportionality constant $\hbar_\nu$. In the additive noise case, the Itô and Stratanovich forms are the same. A sample of this stochastic process is the differential equation

$$\dot{x}_i(t) = f_i(x(t)) + \nu_i(t), \qquad i = 1,2,\ldots,n \tag{34}$$

where ν is the sample of a Gaussian noise process with δ-correlated covariance matrix $\hbar_\nu I$.

A. Formal Derivation using Functional Calculus

The transition probability density $\langle\delta[\mathbf{x}(t)-x]\rangle_\nu|_{x(t_0)=x_0}$ is now derived using functional methods for state model given by the Langevin equation given by (33). From the path integral representation of the Gaussian noise in (23), it follows that

$$\langle\delta^n[\mathbf{x}(t)-x]\rangle_\nu|_{x(t_0)=x_0} = \int [d\rho(\nu)]\delta^n[x(t)-x]|_{x(t_0)=x_0},$$

$$\dot{x}(t) = f(x(t)) + \nu(t). \tag{35}$$

The condition that $x(t)$ in the integrand satisfies the Langevin equation needs to be imposed. This can be done using the identity (173) as follows:

$$P(t,x\,|\,t_0,x_0) = \int [\mathcal{D}\nu(t)]\exp\left(-\frac{1}{2\hbar_\nu}\sum_{i=1}^{n}\int_{t_0}^{t}\nu_i^2(t)dt\right)$$

$$\times \int [\mathcal{D}\dot{x}(t')]J\delta^n(\dot{x}(t)-f(x(t))-\nu(t))$$

$$\times \delta^n[x(t)-x]|_{x(t_0)=x_0} \tag{36}$$

where

$$J = \prod_{i=1}^{n} \det\left(\frac{\delta}{\delta x_j(t')}[\dot{x}_i(t) - f_i(x(t)) - \nu_i(t)]\right). \quad (37)$$

Interchanging the order of integration and then performing the trivial integration over $\nu(t)$ leads to

$$P(t,x \mid t_0, x_0)$$

$$= \int [\mathcal{D}\dot{x}(t)] \exp\left(-\frac{1}{2\hbar_\nu} \sum_{i=1}^{n} \int_{t_0}^{t} (\dot{x}_i(t) - f_i(x(t)))^2 dt\right)$$

$$\times J \delta^n[x(t) - x]|_{x(t_0)=x_0}. \quad (38)$$

It remains to compute the Jacobian and convert the path integral to one over $x(t)$. The Jacobian J follows from noting that

$$\frac{\delta}{\delta x_j(t')}[\dot{x}_i(t) - f_i(x(t)) - \nu_i(t)]$$

$$= \left[\delta_{ij}\frac{d}{dt} - \frac{\partial f_i}{\partial x_j}(x(t'))\right]\delta(t - t')$$

$$= \frac{d}{dt}\left[\delta_{ij}\delta(t - t') - \theta(t - t')\frac{\partial f_i}{\partial x_j}(x(t'))\right] \quad (39)$$

because

$$\frac{d}{dt}\theta(t - t') = \delta(t - t') \quad \text{and} \quad \frac{\delta x_i(t)}{\delta x_j(t')} = \delta_{ij}\delta(t - t'). \quad (40)$$

The Jacobian of the transformation is then given by

$$J \equiv \det\left(\frac{d}{dt}\left[\delta_{ij}\delta(t - t') - \theta(t - t')\frac{\partial f_i}{\partial x_j}(x(t'))\right]\right)$$

$$= \det\left(\frac{d}{dt}\right)\det\left(\delta_{ij}\delta(t - t') - \theta(t - t')\frac{\partial f_i}{\partial x_j}(x(t'))\right)$$

$$= \mathcal{N}\det\left(\delta_{ij}\delta(t - t') - \theta(t - t')\frac{\partial f_i}{\partial x_j}(x(t'))\right) \quad (41)$$

where $\mathcal{N}$ is an irrelevant constant. From the identity $\det A = \exp(\mathrm{tr}\ln A)$, and using the identity in (175), it follows that

$$\ln J = \ln\det\left[\delta_{ij}\delta(t - t') - \theta(t - t')\frac{\partial f_i}{\partial x_j}(x(t'))\right]$$

$$= -\theta(0)\sum_{i=1}^{n}\int_{t_0}^{t} dt\frac{\partial f_i}{\partial x_i}(x(t))$$

$$= -\frac{1}{2}\sum_{i=1}^{n}\int_{t_0}^{t}\frac{\partial f_i}{\partial x_i}(x(t))dt. \quad (42)$$

Since the symmetric or Feynman convention is used here,[5] $\theta(0) = \frac{1}{2}$. It also ensures that the operations of averaging and differentiation commute (see, for instance, [26]). Also,

$$\int [\mathcal{D}\dot{x}(t)] \to \int [\mathcal{D}x(t)]\det\left(\frac{\delta\dot{x}(t)}{\delta x(t)}\right)$$

$$= \det\left(\frac{d}{dt}\delta(t - t')\right)\int [\mathcal{D}x(t)] \to \int [\mathcal{D}x(t)]$$

$$(43)$$

where the irrelevant constant $\det(d/dt)\delta(t - t')$ has been absorbed into the measure.

Combining these results leads to the following path integral formula:

$$P(t,x \mid t_0, x_0)$$

$$= \int_{x(t_0)=x_0}^{x(t)=x} [\mathcal{D}x(t)]$$

$$\times \exp\left(-\frac{1}{2\hbar_\nu}\sum_{i=1}^{n}\int_{t_0}^{t} dt\right.$$

$$\left.\times \left[(\dot{x}_i(t) - f_i(x(t)))^2 + \hbar_\nu\frac{\partial f_i}{\partial x_i}(x(t))\right]\right). \quad (44)$$

B. Derivation using the Chapman-Kolmogorov Semi-Group Property

Consider the time evolution of the state in an infinitesimal time interval. Integrating the Langevin equation (33) in an infinitesimal time interval $(t, t + \epsilon)$ yields

$$\mathbf{x}_i(t + \epsilon) = \mathbf{x}_i(t) + \epsilon f_i(\mathbf{x}(t)) + \int_t^{t+\epsilon}\nu_i(\tau)d\tau + O(\epsilon^{3/2}). \quad (45)$$

To order ϵ, the continuum Langevin equation is equivalent to the discretized Langevin equation:

$$\mathbf{x}_i(t + \epsilon) = \mathbf{x}_i(t) + \epsilon f_i(\mathbf{x}(t)) + \sqrt{\epsilon}\bar{\nu}_i(t) \quad (46)$$

where

$$\sqrt{\epsilon}\bar{\nu}_i(t) \equiv \int_t^{t+\epsilon}\nu_i(\tau)d\tau. \quad (47)$$

From the properties of $\nu_i(t)$, it follows that

$$\langle\sqrt{\epsilon}\bar{\nu}_i(t)\rangle = \int_t^{t+\epsilon}\langle\nu_i(\tau)\rangle d\tau$$

$$= 0 \quad (48)$$

---

[5]The Stratanovich convention symmetrizes the argument of the diffusion vielbein, while the Feynman convention symmetrizes the argument of the drift.

and

$$\epsilon \langle \bar{\nu}_i(t) \bar{\nu}_j(t') \rangle = \left\langle \int_t^{t+\epsilon} \int_{t'}^{t'+\epsilon} \nu_i(\tau) \nu_j(\tau') d\tau d\tau' \right\rangle$$

$$= \int_t^{t+\epsilon} \int_{t'}^{t'+\epsilon} \langle \nu_i(\tau) \nu_j(\tau') \rangle d\tau d\tau'$$

$$= \int_t^{t+\epsilon} \int_{t'}^{t'+\epsilon} \hbar_\nu \delta_{ij} \delta(\tau - \tau') d\tau d\tau'$$

$$= \begin{cases} 0 & t \neq t' \\ \epsilon \hbar_\nu \delta_{ij} & t = t' \end{cases}. \tag{49}$$

Therefore,

$$\langle \bar{\nu}_i(t) \rangle = 0$$
$$\langle \bar{\nu}_i(t) \bar{\nu}_j(t') \rangle = \hbar_\nu \delta_{ij} \delta_{tt'} \tag{50}$$

which implies that $\bar{\nu}(t)$ is also a (discrete-time) Gaussian process.

Consider the quantity $P(t + \epsilon, x \mid t, x')$. The Fourier transform $\tilde{P}$ of P with respect to x is

$$\tilde{P}(t + \epsilon, p \mid t, x')$$

$$= \int \{d^n x\} \exp\left(-i \sum_{i=1}^n p_i x_i\right) P(t + \epsilon, x; t, x')$$

$$= \int \{d^n x\} \exp\left(-i \sum_{i=1}^n p_i x_i\right) \langle \delta^n(\mathbf{x}(t + \epsilon) - x) \rangle|_{\mathbf{x}(t) = x'}. \tag{51}$$

The order of integration over x and averaging are interchanged to yield

$$\tilde{P}(t + \epsilon, p \mid t, x') = \left\langle \exp\left(-i \sum_{i=1}^n p_i \mathbf{x}_i(t + \epsilon)\right) \right\rangle$$

$$= \exp\left[-i \sum_{i=1}^n p_i(x_i' + \epsilon f_i(x'))\right]$$

$$\times \left\langle \exp\left[-i \sqrt{\epsilon} \sum_{i=1}^n p_i \bar{\nu}_i(t)\right] \right\rangle. \tag{52}$$

Now,

$$\left\langle \exp\left(-i \sqrt{\epsilon} \sum_{i=1}^n p_i \bar{\nu}_i(t)\right) \right\rangle$$

$$= \int \frac{1}{\sqrt{(2\pi\hbar_\nu)^n}} \{d^n \bar{\nu}(t)\} \exp\left(-i \sqrt{\epsilon} \sum_{i=1}^n p_i \bar{\nu}_i(t)\right)$$

$$\times \exp\left(-\frac{1}{2\hbar_\nu} \sum_{i,j=1}^n \bar{\nu}_i(t) \delta_{ij}(t) \bar{\nu}_j(t)\right)$$

$$= \exp\left(-\frac{\epsilon \hbar_\nu}{2} \sum_{i,j=1}^n p_i^2\right). \tag{53}$$

Inverting the Fourier transform, it follows that

$$P(t + \epsilon, x \mid t, x')$$

$$\approx \frac{1}{(2\pi)^n} \int \{d^n p\}$$

$$\times \exp\left(i \sum_{i=1}^n p_i(x_i - x_i' - \epsilon f_i(x')) - \epsilon \frac{\hbar_\nu}{2} \sum_{i=1}^n p_i^2\right)$$

$$= \frac{1}{\sqrt{(2\pi\hbar_\nu \epsilon)^n}} \exp\left[-\frac{1}{2\hbar_\nu \epsilon} \sum_{i=1}^n (x_i - x_i' - \epsilon f_i(x'))^2\right]. \tag{54}$$

Partition the time interval $[t_0, t]$ into N equi-spaced time intervals so that $t_i = t_0 + i\epsilon$ where $\epsilon = (t - t_0)/N$. Then, from the Chapman-Kolmogorov semi-group property it follows that

$$P(t, x \mid t_0, x_0) = \int \{d^n x(t_1) \cdots d^n x(t_{N-1})\}$$

$$\times P(t, x \mid t_{N-1}, x(t_{N-1})) \cdots P(t_1, x(t_1) \mid t_0, x_0)$$

$$= \int \left\{ \prod_{i=1}^{N-1} d^n x(t_i) \right\} \left[\prod_{i=1}^N P(t_i, x(t_i) \mid t_{i-1}, x(t_{i-1})) \right] \tag{55}$$

where $x(t_0) = x_0$ and $x(t_N) = x$. Therefore,

$$P(t, x \mid t_0 x_0) = \int_{x(t_0) = x_0}^{x(t) = x} [\mathcal{D}x(t)] \exp\left[-\frac{1}{\hbar_\nu} S_I(t_0, t)\right] \tag{56}$$

where

$$S_I(t_0, t) = \lim_{\epsilon \to 0} \frac{1}{2\epsilon} \sum_{k=1}^N \sum_{i=1}^n (x_i(t_k) - x_i(t_{k-1}) - \epsilon f(x_i(t_{k-1})))^2 \tag{57}$$

$$[\mathcal{D}x(t)] = \lim_{\epsilon \to 0} \frac{1}{\sqrt{(2\pi\hbar_\nu \epsilon)^n}} \prod_{i=1}^{N-1} \left\{ \frac{d^n x(t_i)}{\sqrt{(2\pi\hbar_\nu \epsilon)^n}} \right\}, \qquad \epsilon = \frac{t - t_0}{N}.$$

Observe that the extra factor of $1/\sqrt{(2\pi\hbar_\nu \epsilon)^n}$ is due to N terms in the square bracketed term in (55). In the continuum notation this is

$$P(t, x \mid t_0, x_0) = \int_{x(t_0) = x_0}^{x(t) = x} [\mathcal{D}x(t)]$$

$$\times \exp\left[-\frac{1}{2\hbar_\nu} \int_{t_0}^t dt \sum_{i=1}^n (\dot{x}_i(t) - f_i(x^{(0)}(t)))^2\right] \tag{58}$$

where the superscript (0) is to remind us that the drift is evaluated at the earlier time. This is termed the "prepoint" version.

This formula clearly differs from the formula (44) obtained using functional methods. In fact, another important difference is that in the (58), the argument of f is to be evaluated at the prepoint, rather than the mid-point (as in the case of (44)). This

discrepancy is resolved once the prepoint form of the proportional-integral (PI) formula is converted to the symmetric form.

In order to obtain the result in the Feynman convention, the argument of f needs to be symmetrized. It is straightforward to obtain a more general expression as follows. As evident from (45) and (50), typical values of $x - x' \sim O(\epsilon^{1/2})$. Let $x^{(r)} \equiv x' + r(x - x')$. Then, it follows that the leading order term in $x_i - x_i' - \epsilon f_i(x')$ is given by

$$
x_i - x_i' - \epsilon f_i(x') = x_i - x_i' - \epsilon f_i(x^{(r)})
$$
$$
+ r\epsilon \sum_{i'=1}^{n} (x_{i'} - x_{i'}') \frac{\partial f_i}{\partial x_{i'}}(x^{(r)}) + O(\epsilon^2)(\bar{x}).
$$
(59)

The last term may be eliminated by introducing a new variable $\tilde{x}_i$ defined as

$$
\tilde{x}_i = x_i + r\epsilon \sum_{i'=1}^{n} (x_{i'} - x_{i'}') \frac{\partial f_i}{\partial x_{i'}}(x^{(r)}).
$$
(60)

The Jacobian under this transformation is $(d^n\tilde{x} = \det(\partial \tilde{x}_i)/\partial x_i)d^n x)$

$$
\det\left[\frac{\partial}{\partial x_i}\left(x_j + r\epsilon \sum_{l=1}^{n} (x_l - x_l') \frac{\partial f_j}{\partial x_l}(x^{(r)}) \right) \right]^{-1}
$$
$$
= \det\left(\delta_{ij} + r\epsilon \sum_{l=1}^{n} \delta_{li} \frac{\partial f_j}{\partial x_l}(x^{(r)}) \right)
$$
$$
= \exp\left[-r\hbar_\nu \epsilon \sum_{i=1}^{n} \frac{\partial f_i}{\partial x_i}(x^{(r)}) \right]
$$
(61)

where the identity $\det A = \exp(\mathrm{tr}\ln A)$ has been used. Note that the transformation in (60) does not change the $x - x'$ term in $S_\epsilon(t_0, t)$ since the difference is $O(\epsilon^{3/2})$. Thus the overall exponent is

$$
S_\epsilon(t_0, t) = \lim_{\epsilon \to 0} \sum_{k=1}^{N} \sum_{i=1}^{n} \left[\frac{1}{2\epsilon}(x_i(t_k) - x_i(t_{k-1}) - \epsilon f_i(x^{(r)}(t_k)))^2 \right.
$$
$$
\left. + r\epsilon \frac{\partial f_i}{\partial x_i}(\bar{x}(t_k)) \right].
$$
(62)

Therefore, the result for the transition probability density is

$$
P(t, x \mid t_0, x_0) = \int_{x(t_0)=x_0}^{x(t)=x} [\mathcal{D}x(t)] \exp\left(-\frac{1}{\hbar_\nu} S(t_0, t) \right)
$$

where
(63)

$$
S(t_0, t)
$$
$$
= \sum_{i=1}^{n} \int_{t_0}^{t} \left[\frac{1}{2}(\dot{x}_i(t) - f_i(x^{(r)}(t)))^2 + \hbar_\nu r \frac{\partial f_i}{\partial x_i}(x^{(r)}(t)) \right] d\tau.
$$
(64)

When $r = \frac{1}{2}$, or the mid-point or Feynman discretization, agreement with the result obtained by functional methods is seen. The $r = 0$ case is, of course, the prepoint version.

VI. PATH INTEGRAL FORMULA II. EXPLICIT TIME DEPENDENCE

In this section, the general additive noise model is considered: noise is colored and the diffusion vielbein is a time-dependent rectangular matrix. Since the diffusion vielbein is no longer invertible, the method used in the previous section is not applicable.

The state model is given by

$$
d\mathbf{x}(t) = f(\mathbf{x}(t), t)dt + e(t)d\boldsymbol{\nu}(t)
$$
(65)

where $\mathbf{x}(t), f(\mathbf{x}(t)) \in \mathbb{R}^n$, $\boldsymbol{\nu}(t) \in \mathbb{R}^p$, and $e(t) \in \mathbb{R}^{n \times p}$. Observe that drift is now an explicit function of time, and the diffusion vielbein, apart from being a rectangular matrix, has an explicit time dependence as well. However, the diffusion term is assumed to be independent of the state vector; the more general state-dependent, or multiplicative noise, case is more subtle and will be studied in a future paper. Again, the Itô and Stratanovich forms are the same for this equation. A sample of this stochastic process satisfies the following differential equation

$$
\dot{x}_i(t) = f_i(x(t), t) + \sum_{a=1}^{p} e_{ia}(t)\nu_a(t), \qquad i = 1, 2, \ldots, n
$$
(66)

where the ν is a sample of a Gaussian noise process with δ-correlated covariance matrix Q.

It is assumed that the diffusion matrix is invertible; if not, it is physically plausible that results for the singular case are close to the "nearly singular" case, which is obtained by adding a small quantitity to render it invertible.

A. Formal Derivation using Functional Calculus

The discussion here is similar to that in Section V-A. However, imposing a delta function constraint requires a different method (as in Section II) since $e(t)$ is not a square matrix; and hence is not invertible. The path integral expression for the transition probability density is

$$
\langle \delta^n[\mathbf{x}(t) - x] \rangle |_{x(t_0)=x_0}
$$
$$
= \int [\mathcal{D}\rho(\nu)] \delta^n[x(t) - x] |_{x(t_0)=x_0}
$$
$$
= \int [\mathcal{D}\nu(t)] \exp\left(-\frac{1}{2} \sum_{a,b=1}^{p} \int \nu_a(t)(Q^{-1}(t))_{ab}\nu_b(t)dt \right)
$$
$$
\times \delta[x(t) - x] |_{x(t_0)=x_0}
$$
(67)

where $x(t)$ satisfies the Langevin equation

$$
\dot{x}_i(t) = f_i(x(t), t) + \sum_{a=1}^{p} e_{ia}(t)\nu_a(t).
$$
(68)

The Langevin equation condition can be imposed using (173) to yield

$$P(t, x \mid t_0, x_0) = \int [\mathcal{D}\nu(t)] \exp\left(-\frac{1}{2} \sum_{a,b=1}^{p} \int \nu_a(t)(Q^{-1}(t))_{ab}\nu_b(t)dt\right)$$

$$\times \int [\mathcal{D}\dot{x}(t')] J \delta^n \left[\dot{x}_i(t) - f_i(x(t),t) - \sum_{a=1}^{p} e_{ia}\nu_a(t)\right] \delta[x(t) - x]\Big|_{x(t_0)=x_0}. \tag{69}$$

Upon evaluating the functional derivative of the Langevin equation

$$\frac{\delta}{\delta x_j(t')} \left[\dot{x}_i(t) - f_i(x(t),t) - \sum_{a=1}^{p} e_{ia}(t)\nu_a(t)\right] = \left[\delta_{ij}\frac{d}{dt} - \frac{\partial f_i}{\partial x_j}(x(t'),t')\right] \delta(t - t')$$

$$= \frac{d}{dt}\left[\delta_{ij}\delta(t - t') - \theta(t - t')\frac{\partial f_i}{\partial x_j}(x(t'),t')\right] \tag{70}$$

it follows that the Jacobian J is given by

$$J = \det\left(\frac{d}{dt}\right) \det\left(\delta_{ij}\delta(t - t') - \theta(t - t')\frac{\partial f_i}{\partial x_j}(x(t'),t')\right)$$

$$= \mathcal{N} \det\left(\delta_{ij}\delta(t - t') - \theta(t - t')\frac{\partial f_i}{\partial x_j}(x(t'),t'))\right) \tag{71}$$

where $\mathcal{N}$ is an irrelevant constant that can be ignored (or absorbed into the measure). Hence,

$$\ln J = \ln \det\left[\delta_{ij}\delta(t - t') - \theta(t - t')\frac{\partial f_i}{\partial x_j}(x(t'),t')\right]$$

$$= -\sum_{i=1}^{n} \frac{1}{2} \int \frac{\partial f_i}{\partial x_i}(x(t),t)dt \tag{72}$$

Also, following arguments in (43) the measure $[\mathcal{D}\dot{x}(t)]$ can be replaced by $[\mathcal{D}x(t)]$. Thus, so far,

$$P(t, x \mid t_0, x_0) = \int_{x(t_0)=x_0}^{x(t)=x} [\mathcal{D}x(t)][\mathcal{D}\nu(t)] \exp\left(-\frac{1}{2}\int \sum_{i,j=1}^{n}\sum_{a,b=1}^{p} \nu_a(t)(Q^{-1}(t))_{ab}\nu_b(t)dt\right)$$

$$\times \delta^n\left(\dot{x}_i(t) - f_i(x(t),t) - \sum_{a=1}^{p} e_{ia}(t)\nu_a(t)\right) \exp\left(-\frac{1}{2}\int \sum_{i=1}^{n}\frac{\partial f_i}{\partial x_i}(x(t),t)dt\right). \tag{73}$$

Using the Fourier integral version of the delta function, the transition probability density becomes

$$P(t, x \mid t_0, x_0) = \int_{x(t_0)=x_0}^{x(t)=x} [\mathcal{D}x(t)][\mathcal{D}\nu(t)][\mathcal{D}\lambda(t)] \exp\left(i\int \sum_{i=1}^{n}\lambda_i(t)(\dot{x}_i(t) - f_i(x(t),t) - \sum_{a=1}^{p} e_{ia}(t)\nu_a(t))\right)$$

$$\times \exp\left(-\frac{1}{2}\int \sum_{a,b=1}^{p} \nu_a(t)(Q^{-1}(t))_{ab}\nu_b(t)dt\right) \exp\left(-\frac{1}{2}\int \sum_{i=1}^{n}\frac{\partial f_i}{\partial x_i}(x(t),t)dt\right). \tag{74}$$

Integration over $\nu(t)$ results in

$$P(t,x \mid t_0,x_0) = \int_{x(t_0)=x_0}^{x(t)=x} [\mathcal{D}x(t)][\mathcal{D}\lambda(t)] \exp\left(-\int dt \left[\frac{1}{2}\sum_{i=1}^{n} \frac{\partial f_i}{\partial x_i}(x(t),t)\right]\right)$$

$$\times \exp\left(-\int dt \left[\sum_{i,j=1}^{n} \lambda_i(t)g_{ij}(t)\lambda_j(t) - i\sum_{i-1}^{n} \lambda_i(t)(\dot{x}_i(t) - f_i(x(t),t))\right]\right). \tag{75}$$

Finally, integrating over $\lambda_i(t)$ leads to

$$P(t,x \mid t_0,x_0) = \int_{x(t_0)=x_0}^{x(t)=x} [\mathcal{D}x(t)]\exp(-S) \tag{76}$$

where the "action" S is

$$S = \frac{1}{2}\int dt \left[\sum_{i,j=1}^{n} (\dot{x}_i(t) - f_i(x(t),t))g_{ij}^{-1}(t)(\dot{x}_j(t) - f_j(x(t),t)) + \sum_{i=1}^{n} \frac{\partial f_i}{\partial x_i}(x(t),t)\right]. \tag{77}$$

B. Derivation using the Chapman-Kolmogorov Semi-Group Property

Integrating the (65) over an infinitesimal time interval $(t, t + \epsilon)$ yields

$$\mathbf{x}_i(t + \epsilon) = \mathbf{x}_i(t) + \epsilon f_i(\mathbf{x}(t),t)$$

$$+ \sum_{a=1}^{p} \int_t^{t+\epsilon} e_{ia}(t)\nu_a(t) + O(\epsilon^{3/2}). \tag{78}$$

Note, in particular, that replacing t with $t + \alpha\epsilon$, in $e(t)$ and $f(\mathbf{x}(t),t)$ (where $\alpha \in (0,1)$) leads to errors of order higher than ϵ $(O(\epsilon^{3/2}))$ and hence are irrelevant in the continuum limit. The continuum Langevin equation (65) then becomes the following discretized Langevin equation:

$$\mathbf{x}_i(t + \epsilon) = \mathbf{x}_i(t) + \epsilon f_i(\mathbf{x}(t),t) + \sqrt{\epsilon}\sum_{a=1}^{p} e_{ia}\bar{\nu}_a(t). \tag{79}$$

As in the previous discussion, the following identification has been made

$$\sqrt{\epsilon}\sum_{a=1}^{p} e_{ia}(t)\bar{\nu}_a(t) \equiv \sum_{a=1}^{p}\int_t^{t+\epsilon} e_{ia}(\tau)\nu_a(\tau)d\tau. \tag{80}$$

Since

$$\langle\nu_a(t)\rangle = 0$$
$$\langle\nu_a(t)\nu_b(t')\rangle = Q_{ab}(t)\delta(t - t') \tag{81}$$

it is clear that

$$\epsilon\sum_{a,b=1}^{p}\langle e_{ia}(t)\bar{\nu}_a(t)\bar{\nu}_b(t')e_{bj}^T(t')\rangle$$

$$= \epsilon\sum_{a,b=1}^{p} e_{ia}(t)\langle\bar{\nu}_a(t)\bar{\nu}_b(t')\rangle e_{bj}^T(t)$$

$$= \int_t^{t+\epsilon}\int_{t'}^{t'+\epsilon}\sum_{a,b=1}^{p} e_{ia}(\tau)\langle\nu_a(\tau)\nu_b(\tau')\rangle e_{bj}^T(\tau')d\tau d\tau'$$

$$= \int_t^{t+\epsilon}\int_{t'}^{t'+\epsilon}\sum_{a,b=1}^{p} e_{ia}(\tau)Q_{ab}(\tau)\delta(\tau - \tau')e_{bj}^T(\tau')d\tau d\tau'$$

$$= \begin{cases} \int_t^{t+\epsilon}\sum_{a,b=1}^{p} e_{ia}(\tau)Q_{ab}(\tau)e_{bj}^T d\tau, & t = t' \\ 0, & t \neq t' \end{cases}$$

$$= \begin{cases} \epsilon\sum_{a,b=1}^{p} e_{ia}(\tau)Q_{ab}(\tau)e_{bj}^T(\tau), & t = t' \\ 0, & t \neq t' \end{cases}. \tag{82}$$

Hence, it follows that

$$\langle\bar{\nu}_a(t)\rangle = 0$$
$$\langle\bar{\nu}_a(t)\bar{\nu}_b(t')\rangle = Q_{ab}(t)\delta_{tt'} \tag{83}$$

and $\bar{\nu}(t)$ is a discrete-time Gaussian process.

The Fourier transform of $P(t + \epsilon, x; t, x')$ with respect to x is

$$\tilde{P}(t + \epsilon, p \mid t, x')$$

$$= \int d^n x \exp\left(-i\sum_{i=1}^{n} p_i x_i\right) P(t + \epsilon, x; t, x'). \tag{84}$$

Interchanging the order of averaging and integration over x leads to

$$\tilde{P}(t + \epsilon, p \mid t, x')$$

$$= \left\langle\int d^n x \exp\left(-i\sum_{i=1}^{n} p_i x_i\right)\delta^n(\mathbf{x}(t + \epsilon) - x)\right\rangle$$

$$= \left\langle\exp\left(-i\sum_{i=1}^{n} p_i\mathbf{x}_i(t + \epsilon)\right)\right\rangle. \tag{85}$$

Substituting the value of $\mathbf{x}(t+\epsilon)$ in (79),

$$\tilde{P}(t+\epsilon, p \mid t, x')$$

$$= \left\langle \exp\left(-i\sum_{i=1}^{n} p_i \mathbf{x}_i(t+\epsilon)\right) \right\rangle$$

$$= \exp\left(-i\sum_{i=1}^{n} p_i(x_i' + \epsilon f_i(x',t))\right)$$

$$\times \left\langle \exp\left(-i\sqrt{\epsilon}\sum_{i=1}^{n}\sum_{a=1}^{p} p_i e_{ia}(t)\bar{\nu}_a(t)\right) \right\rangle.$$

(86)

Now,

$$\left\langle \exp\left(-i\sqrt{\epsilon}\sum_{i=1}^{n}\sum_{a=1}^{p} p_i e_{ia}(t)\bar{\nu}_a(t)\right) \right\rangle$$

$$= \int \frac{1}{\sqrt{(2\pi\epsilon)^n \det Q(t)}}\{d^n\bar{\nu}(t)\}$$

$$\times \exp\left(-i\sqrt{\epsilon}\sum_{i=1}^{n}\sum_{a=1}^{p} p_i e_{ia}\bar{\nu}_a(t)\right.$$

$$\left. -\frac{\epsilon}{2}\sum_{a,b=1}^{p} \bar{\nu}_a(t)Q_{ab}^{-1}(t)\bar{\nu}_b(t)\right)$$

$$= \exp\left(-\frac{\epsilon}{2}\sum_{i,j=1}^{n}\sum_{a,b=1}^{p} p_i e_{ia}(t)Q_{ab}(t)e_{bj}^{T}(t)p_j\right). \quad (87)$$

Combining (87) and (86), it is evident that

$$\tilde{P}(t+\epsilon, p \mid t, x') = \exp\left(-\frac{\epsilon}{2}\sum_{i,j=1}^{n}\sum_{a,b=1}^{p} p_i e_{ia}(t)Q_{ab}(t)e_{bj}^{T}p_j\right.$$

$$\left. -i\sum_{i=1}^{n} p_i(x_i' + \epsilon f_i(x',t))\right). \quad (88)$$

Inverting the Fourier transform,

From this result, the finite time result can be calculated using the Chapman-Kolmogorov semi-group property of the transition probability density. Specifically, divide the time interval (t,t_0) into N parts, i.e., $\epsilon = (t-t_0)/N$, and use (55) to get

$$P(t,x \mid t_0, x_0) = \lim_{\epsilon \to 0} \int \frac{1}{\sqrt{(2\pi\epsilon)^n \det g(t_0)}} \prod_{i=1}^{N-1}$$

$$\times \left\{\frac{d^n x(t_0 + i\epsilon)}{\sqrt{(2\pi\epsilon)^n \det g(t_0 + i\epsilon)}}\right\} \exp-S_\epsilon(t_0,t)$$

(91)

where

$$S_I(t,t_0) = \lim_{\epsilon \to 0}\frac{1}{2\epsilon}\sum_{k=1}^{N}\sum_{i,j=1}^{n}$$

$$\times [(x_i(t_k) - x_i(t_{k-1}) - \epsilon f_i(x(t_{k-1}),t_{k-1}))$$

$$\times g_{ij}^{-1}(t_k)(x_j(t_k) - x_j(t_{k-1}) - \epsilon f_j(x(t_{k-1}),t_{k-1}))]$$

(92)

and

$$g_{ij}(t) = \left(\sum_{a,b=1}^{p} e_{ia}(t)Q_{ab}(t)e_{bj}^{T}(t)\right). \quad (93)$$

When taking the continuum limit, the argument of f needs to be symmetrized in the Feynman form. Note that it suffices to symmetrize x in the argument since $x - x' \approx O(\sqrt{\epsilon})$. Symmetrization with respect to t, i.e., $t_{k-1} \to \frac{1}{2}(t_k + t_{k-1})$, leads to an error higher than $O(\epsilon)$ and vanishes in the continuum limit.

More generally, setting $x^{(r)}(t_k) = x(t_{k-1}) + r(x(t_k) - x(t_{k-1}))$,

$$x_i(t_k) - x_i(t_{k-1}) - \epsilon f_i(x(t_{k-1}),t_k)$$

$$= x_i(t_k) - x_i(t_{k-1}) - \epsilon f_i(x^{(r)}(t_k) - r[x(t_k) - x(t_{k-1})],t_k)$$

$$= x_i(t_k) - x_i(t_{k-1}) - \epsilon f_i(x^{(r)}(t_k),t_k))$$

$$+ r\epsilon\sum_{j=1}^{n}(x_j(t_k) - x_j(t_{k-1}))\frac{\partial f_i}{\partial x_j^{(r)}}(x^{(r)},t). \quad (94)$$

$$P(t+\epsilon, x \mid t, x') = \frac{1}{(2\pi)^n}\int \{d^n p\}\exp\left(i\sum_{i=1}^{n} p_i x_i\right)\tilde{P}(t+\epsilon, p; t, x')$$

$$= \frac{1}{(2\pi)^n}\int \{d^n p\}\exp\left(-\frac{\epsilon}{2}\sum_{i,j=1}^{n}\sum_{a,b=1}^{p} p_i e_{ia}(t)Q_{ab}(t)e_{bj}^{T}(t)p_j\right)\exp\left(+i\sum_{i=1}^{n} p_i(x_i - x_i' - \epsilon f_i(x',t))\right)$$

$$= \frac{1}{\sqrt{(2\pi\epsilon)^n \det e(t)Q(t)e^T(t)}}\exp(-S_\epsilon(t+\epsilon,t)) \quad (89)$$

where

$$S_\epsilon(t+\epsilon,t) = +\frac{1}{2\epsilon}\sum_{i,j=1}^{n}\left[(x_i - x_i' - \epsilon f_i(x',t))\left[\sum_{a,b=1}^{p} e_{ia}(t)Q_{ab}(t)e_{bj}^{T}(t)\right]^{-1}(x_j - x_j' - \epsilon f_j(x',t))\right]. \quad (90)$$

Define $\tilde{x}$ as follows:

$$\tilde{x}_i = x_i + r\epsilon \sum_{j=1}^{n}(x-x')_j \frac{\partial f_i}{\partial x_j}(x^{(r)},t). \qquad (95)$$

The Jacobian under this change of variables (from x to $\tilde{x}$) is

$$\det\left[\frac{\partial}{\partial x_i}\left(x_j + r\epsilon\sum_{l=1}^{n}(x-x')_l\frac{\partial f_j}{\partial x_l}\right)\right]^{-1} = \det\left[\delta_{ij} + r\epsilon\frac{\partial f_j}{\partial x_i}\right]^{-1}$$

$$= \exp\left(-r\epsilon\sum_{i=1}^{n}\frac{\partial f_i}{\partial x_i}\right)$$

$$(96)$$

where the identity $\det(1+\epsilon A) \approx \exp\epsilon\mathrm{tr}A$ has been used. Hence, the overall action is

$$\sum_{k=1}^{N}\sum_{i,j=1}^{n}\left[\frac{1}{2\epsilon}(x_i(t_k)-x_i(t_{k-1})-\epsilon f_i(x_{i,k}^{(r)},t_k))g_{ij}^{-1}(t_k)(x_j(t_k)\right.$$

$$\left. -x_j(t_{k-1}) + \epsilon f_j(x_k^{(r)},t_k)) + r\epsilon\frac{\partial f_i}{\partial x_i}(x^{(r)}(t_k),t_k)\right]. $$

$$(97)$$

Therefore, in the continuum limit, the transition probability density is given by

$$P(t,x\,|\,t_0,x_0) = \int_{x(t_0)=x_0}^{x(t)=x}[\mathcal{D}x(t)]\exp(-S_F(t_0,t))$$

$$(98)$$

where the action is

$$S^{(r)}(t,t_0) = \sum_{i=1}^{n}\int_{t_0}^{t}\left[\frac{1}{2}(\dot{x}_i(t)-f_i(x^{(r)}(t),t))g_{ij}^{-1}(t)(\dot{x}_j(t)\right.$$

$$\left. -f_j(x^{(r)}(t),t)) + r\frac{\partial f_i}{\partial x_i}(x^{(r)}(t),t)\right]$$

$$(99)$$

and

$$[\mathcal{D}x(t)] = \lim_{\epsilon\to 0}\frac{1}{\sqrt{(2\pi\epsilon)^n \det g(t_0)}}\prod_{i=1}^{N-1}$$

$$\times\left\{d^nx(t_0+i\epsilon)\frac{1}{\sqrt{(2\pi\epsilon)^n \det g(t_0+i\epsilon)}}\right\}. $$

$$(100)$$

Naïvely, it seems that the path integral formula depends on r, which contradicts the fact that the FPKfe and the Langevin equation do not depend

on r. This is resolved in Section VII where it is demonstrated that the r dependence of the Feynman path integral formula is illusory, i.e., r dependence cancels.

VII. VERIFICATION OF THE PATH INTEGRAL FORMULA

In this section it is independently verified that the path integral formulas derived in the previous sections are the fundamental solutions of the FPKfe. Clearly, it is sufficient to prove the path integral formula derived for the explicit time-dependent case and for arbitrary discretization. The method used closely follows the techniques first used by Feynman to verify the path integral formula he derived for the fundamental solution of the Schrödinger equation (see [20]).

From the Chapman-Kolmogorov semi-group property, it follows that

$$P(t+\epsilon,x\,|\,t,x_0) = \int P(t+\epsilon,x\,|\,t,x')P(t,x'\,|\,t_0,x_0)\{d^nx'\}. $$

$$(101)$$

According to (89) and (99)

$$P(t+\epsilon,x\,|\,t,x') = \frac{1}{\sqrt{(2\pi\epsilon)^n \det g(t)}}\exp\left(-\frac{1}{2\epsilon}\sum_{i,j=1}^{n}[x_i-x_i'-\epsilon f_i(x^{(r)},t)]g_{ij}^{-1}(t)[x_j-x_j'-\epsilon f_j(x^{(r)},t)] - \epsilon\frac{1}{2}\sum_{i=1}^{n}\frac{\partial f_i}{\partial x_i}(x^{(r)},t)\right). $$

$$(102)$$

Here $x^{(r)} = x' + r(x-x')$.

The quantity $P(t+\epsilon,x\,|\,t,x')$ is large only if

$$x-x' \approx \epsilon f(x^{(r)},t). \qquad (103)$$

We may then write

$$x = x' + \epsilon f(x^{(r)},t) + \eta \qquad (104)$$

or

$$x' = x - \epsilon f(x^{(r)},t) - \eta \qquad (105)$$

where the equalities are valid to $O(\epsilon)$. Substituting this into (102),

$$P(t+\epsilon,x\,|\,t_0,x_0)$$

$$= \int_{-\infty}^{\infty}\exp\left(-\frac{1}{2\epsilon}\sum_{i,j=1}^{n}\eta_i g_{ij}^{-1}(t)\eta_j\right) \qquad (106)$$

$$\times\left(1-r\epsilon\sum_{i=1}^{n}\frac{\partial f_i}{\partial x_i}(\bar{x},t)\right)P(t,x'\,|\,t_0,x_0)\{d^nx'\}$$

$$\times\frac{1}{\sqrt{(2\pi\epsilon)^n \det g}}. \qquad (107)$$

Next, the variable of integration is changed from x' to η. Since

$$\eta = x' - x + \epsilon f(x' + r(x-x')t) \qquad (108)$$

so that

$$\frac{\partial \eta_i}{\partial x'_j} = \delta_{ij} + (1-r)\epsilon \frac{\partial f_i}{\partial x'_j}(x^{(r)},t) \quad (109)$$

and the Jacobian of the transformation from x' to η is

$$\left(1 - (1-r)\epsilon \sum_{i=1}^{n} \frac{\partial f_i}{\partial x_i}\right). \quad (110)$$

Combining these results leads to

$$P(t+\epsilon, x \mid t_0, x_0)$$

$$= \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\epsilon}\sum_{i,j=1}^{n} \eta_i g_{ij}^{-1}(t)\eta_j\right)$$

$$\times \left(1 - \frac{\epsilon}{2}\sum_{i=1}^{n}\frac{\partial f_i}{\partial x_i}\right) P(t,x' \mid t_0,x_0)$$

$$\times \left(1 - (1-r)\epsilon \sum_{i=1}^{n}\frac{\partial f_i}{\partial x_i}\right)\{d^n\eta\}\frac{1}{\sqrt{(2\pi\epsilon)^n \det g}}$$

$$= \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\epsilon}\sum_{i,j=1}^{n}\eta_i g_{ij}^{-1}(t)\eta_j\right)$$

$$\times \left(1 - \epsilon \sum_{i=1}^{n}\frac{\partial f_i}{\partial x_i}\right) P(t,x-\epsilon f(x,t)-\eta \mid t_0,x_0)\{d^n\eta\}$$

$$\times \frac{1}{\sqrt{(2\pi\epsilon)^n \det g}} \quad (111)$$

where the Jacobian of the the transformation from x' to η to $O(\epsilon)$ is included. Observe that the r dependent terms are now absent.

The left hand side of (111) is

$$P(t,x \mid t_0,x_0) + \epsilon\frac{\partial P}{\partial t}(t,x \mid t_0,x_0). \quad (112)$$

Also,

$$P(t,x-\epsilon f(x,t)-\eta \mid t_0,x_0)$$

$$= P(t,x \mid t_0,x_0) - \epsilon(f_i(x,t)+\eta_i)\frac{\partial P}{\partial x_i}(t,x \mid t_0,x_0)$$

$$+ \frac{1}{2}\eta_i\eta_j\frac{\partial^2 P}{\partial x_i \partial x_j}(t,x \mid t_0,x_0) \quad (113)$$

where terms that contribute to the order of ϵ have been retained. Therefore, the integrand

Here, the term linear in η has been dropped since its contribution to $P(t+\epsilon,x \mid t_0,x_0)$ vanishes upon integration over η due to odd symmetry.

Since

$$\int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\epsilon}\sum_{i,j=1}^{n}\eta_i g_{ij}^{-1}(t)\eta_j\right)\{d^n\eta\}$$

$$\times \frac{1}{\sqrt{(2\pi\epsilon)^n \det g(t)}} = 1$$

$$\quad (116)$$

$$\int_{-\infty}^{\infty} \eta_i\eta_j \exp\left(-\frac{1}{2\epsilon}\sum_{i',j'=1}^{n}\eta_{i'} g_{i'j'}^{-1}(t)\eta_{j'}\right)\{d^n\eta\}$$

$$\times \frac{1}{\sqrt{(2\pi\epsilon)^n \det g(t)}} = \epsilon g_{ij}(t)$$

the right hand side of (111) is

$$P(t,x \mid t_0,x_0) - \epsilon\sum_{i=1}^{n}\frac{\partial}{\partial x_i}[f_i(x,t)P(t,x \mid t_0,x_0)]$$

$$+ \epsilon\frac{1}{2}\sum_{i,j=1}^{n}g_{ij}(t)\frac{\partial^2 P}{\partial x_i \partial x_j}(t,x \mid t_0,x_0). \quad (117)$$

Finally, note that

$$P(t,x \mid t,x') = \lim_{\epsilon \to 0}\frac{1}{\sqrt{(2\pi\epsilon)^n \det g(t)}}$$

$$\times \exp\left(-\frac{1}{2\epsilon}\sum_{i,j=1}^{n}[(x_i-x'_i)g_{ij}^{-1}(t)(x_j-x'_j)]\right)$$

$$= \delta^n(x-x'). \quad (118)$$

Therefore, from (112), (117), and (118) it follows that $P(t,x \mid t_0,x_0)$ is the fundamental solution of the FPKfe:

$$\frac{\partial P}{\partial t}(t,x \mid t_0,x_0) = -\sum_{i=1}^{n}\frac{\partial}{\partial x_i}[f_i(x,t)P(t,x \mid t_0,x_0)]$$

$$+ \frac{1}{2}\sum_{i,j=1}^{n}\frac{\partial^2}{\partial x_i \partial x_j}[g_{ij}(t)P(t,x \mid t_0,x_0))] \quad (119)$$

$$P(t,x \mid t,x_0) = \delta^n(x-x_0).$$

$$\left(1 - \epsilon\sum_{i=1}^{n}\frac{\partial f_i}{\partial x_i}(x,t)\right)\left(P(t,x \mid t_0,x_0) - \epsilon(f_i(x,t)+\eta_i)\frac{\partial P}{\partial x_i}(t,x \mid t_0,x_0) + \frac{1}{2}\eta_i\eta_j\frac{\partial^2 P}{\partial x_i \partial x_j}(t,x \mid t_0,x_0\right) \quad (114)$$

---

becomes, to $O(\epsilon)$,

$$P(t,x \mid t_0,x_0) - \epsilon\sum_{i=1}^{n}\frac{\partial}{\partial x_i}[f_i(x,t)P(t,x \mid t_0,x_0)]$$

$$+ \frac{1}{2}\eta_i\eta_j\frac{\partial^2 P}{\partial x_i \partial x_j}(t,x \mid t_0,x_0). \quad (115)$$

VIII. EXAMPLES

In this section a few examples are presented illustrating the utility of the path integral formulas derived in this paper. For simplicity, all examples are one-dimensional. For some higher dimensional

examples, see [28]. Note that an accurate determination of the transition probability density automatically implies that the continuous-filtering problem can be solved very accurately. It is therefore sufficient to demonstrate that the transition probability density can be computed accurately. One complete filtering example is also included.

From the discussions in the previous sections, the one-step $r = 0$, or prepoint, approximate formula is

$$P(t'', x'' \mid t', x') = \frac{1}{\sqrt{(2\pi(t'' - t)) \det g(t')}} \exp\left(-\frac{(t'' - t')}{2} \sum_{i,j=1}^{n} \left[\frac{x'' - x'}{t'' - t'} - f_i(x', t') \right] g_{ij}^{-1}(t') \left[\frac{(x'' - x')}{(t'' - t')} - f_j(x', t') \right] \right).$$

(120)

The one-step $r = \frac{1}{2}$, or symmetric, approximate path integral formula for the transition probability amplitude is

$$P(t'', x'' \mid t', x') = \frac{1}{\sqrt{(2\pi(t'' - t)) \det g(\bar{t})}}$$

$$\times \exp\left(-\frac{(t'' - t')}{2} \sum_{i,j=1}^{n} \left[\frac{x'' - x'}{t'' - t'} - f_i(\bar{x}, \bar{t}) \right] g_{ij}^{-1}(\bar{t}) \left[\frac{(x'' - x')}{(t'' - t')} - f_j(\bar{x}, \bar{t}) \right] - \frac{(t'' - t')}{2} \sum_{i=1}^{n} \frac{\partial f_i}{\partial x_i}(\bar{x}, \bar{t}) \right)$$

(121)

where $\bar{x} = \frac{1}{2}(x'' + x')$ and $\bar{t} = \frac{1}{2}(t' + t'')$. This one-step approximation, for any r, is termed the Dirac-Feynman (DF) approximation, as it was first noted by Dirac in the quantum mechanical context (and for $r = \frac{1}{2}$). Using the DF approximation in an algorithm to solve the continuous-discrete filtering problem will be called the DF algorithm.

EXAMPLE 1 (Wiener Process)   The simplest example is the following state model:

$$d\mathbf{x}(t) = \sigma d\mathbf{v}(t). \tag{122}$$

The one-step formula for this case is

$$p(t'', x'' \mid t', x') = \frac{1}{\sqrt{2\pi\sigma^2(t'' - t')}} \exp\left(-\frac{(x'' - x')^2}{2\sigma^2(t'' - t')} \right). \tag{123}$$

This formula is well known to be exact for arbitrary size time step [29]. The extension to the n-dimensional case is straightforward.

EXAMPLE 2 (A Polynomial Nonlinear Model) Consider the model

$$d\mathbf{x}(t) = -4\mathbf{x}(t)(\mathbf{x}^2(t) - 1)dt + 0.1 d\mathbf{v}(t). \tag{124}$$

The results for the transition probability density for $x' = -0.75, -0.35, -0.05, 0.30, 0.70$ are plotted in Fig. 1 along with values obtained from 10^5 simulations. The time step size is 0.1.

EXAMPLE 3 (A Transcendental Nonlinear Model) Consider the model

$$d\mathbf{x}(t) = 1.2 \cos(3\mathbf{x}(t))dt + 0.3 d\mathbf{v}(t). \tag{125}$$

The symmetric one-step approximate path integral formula for the conditional probability density is plotted in Fig. 2. The time step size is 0.2.

EXAMPLE 4 (A Filtering Example)   Consider the following signal and measurement model:

$$d\mathbf{x}(t) = -\sin(\mathbf{x}(t)(\mathbf{x}^2(t) - 1))dt + 3 d\mathbf{v}(t)$$

$$\mathbf{y}(t) = \mathbf{x}^2(t) + \mathbf{w}(t). \tag{126}$$

A state and measurement process sample path is shown in Figs. 3 and 4.

Figure 5 shows the conditional probability density as an intensity plot with the true state overlaid on it. The spatial grid spacing was 0.25, while the temporal spacing was $\Delta t = 0.01$. It is seen that the conditional probability density is bi-modal, and that the state is on (or close to) regions of significant probability mass. Note that the fundamental solution could be precomputed. It is found that it took about 0.5 s to track over 1000 time steps in this instance.

This was also investigated using the basic SIR (sampling importance resampling) particle filter by using the basic Euler discretized version of the state model [30, 31, 32]. Figures 6 and 7 show two markedly different results using the SIR particle filter using 200 particles. It is apparent that the particle filter does not capture the bi-modal nature in this instance. Increasing the number of particles to 1000 and using the auxiliary particle filter did not yield better performance. Note that this is the result for a single state and measurement sample. Also note that the time required for the particle filter solution is almost 2 min, over two orders of magnitude more than for the DF algorithm using grid methods. Note that no precomputation is possible using particle filters.

These examples illustrate that some of the simplest one-step path integral approximation formulas are often very accurate. Clearly, the smaller the time steps, the more accurate is the approximation. What is remarkable is to note that the time step need not
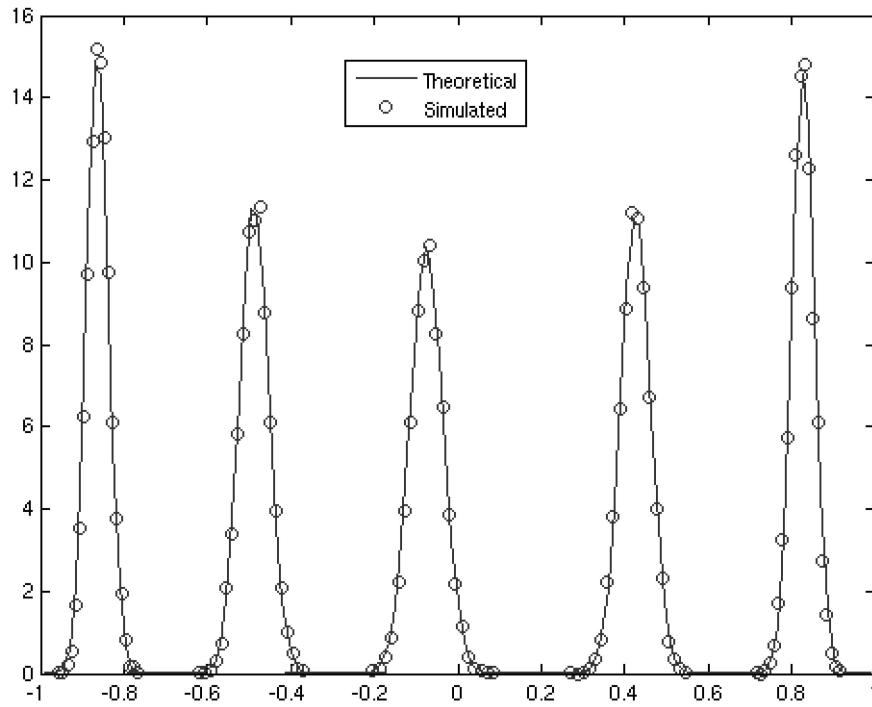
Fig. 1. Transition probability density for nonlinear state model with polynomial drift (124). Here $t'' - t' = 0.1$ and $x' = -0.75, -0.35, -0.05, 0.30, 0.70$.
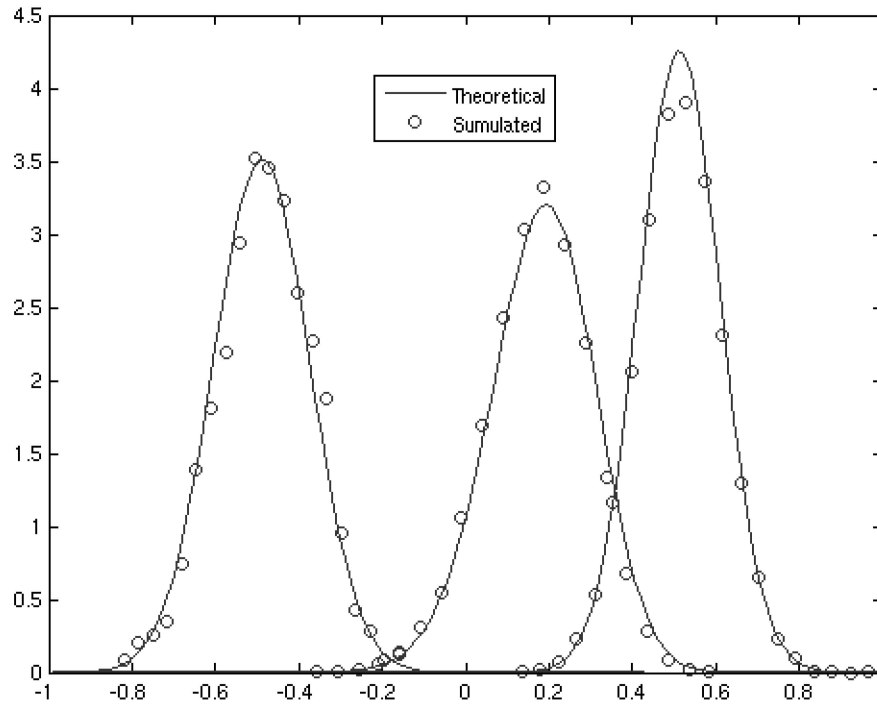

Fig. 2. Transition probability density for a nonlinear state model with polynomial drift (125). Here $t'' - t' = 0.2$ and $x' = -0.5, -0.05, -0.05, 0.5$.

be infinitesimal for extremely accurate results. In practice, the time steps are often small, and the one-step approximations discussed here might be adequate for many applications. For larger time steps, there are better one-step approximations that have been obtained and will yield better performance.

Finally, observe that time dependence is not a problem provided the variation of the diffusion vielbein is small over the time interval $(t'' - t')$. Of course, this means that the transition probability density depends on the actual times (t'' and t'), rather than just their difference (as in the time-independent case).
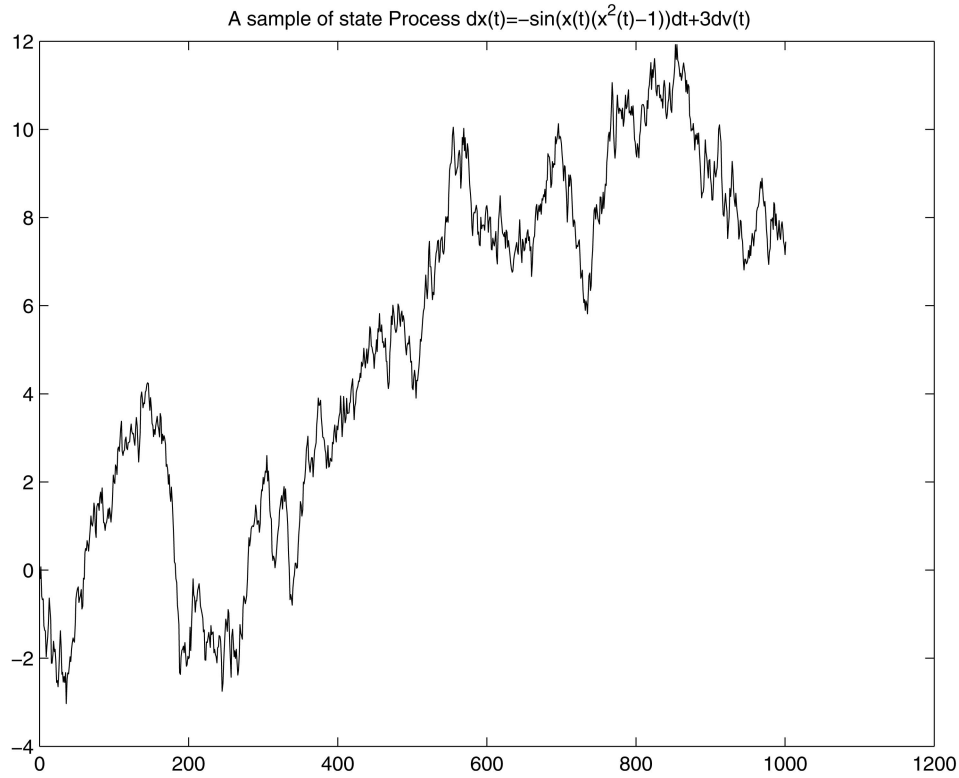
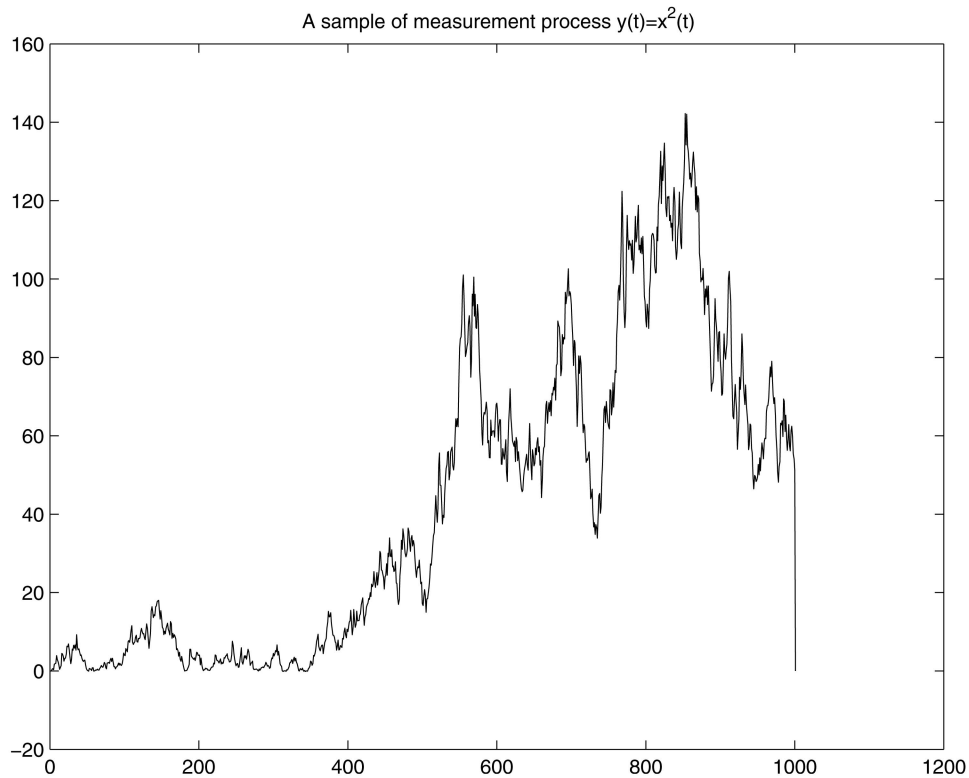Fig. 3.   Sample path of continuous-time state process described by (126).

Fig. 4.   Sample path of measurement process described by (126).

IX.   ADDITIONAL REMARKS

A.   Physical Meaning of the Path Integral Formula

A beautiful discussion of the path integral in the quantum mechanical context is given in [20]. Let us consider the model discussed in Section V. Intuitively, the path integral is a sum of contributions over all paths satisfying the boundary conditions: $x(t_0) = x_0$ and $x(t) = x$. The action is computed for each
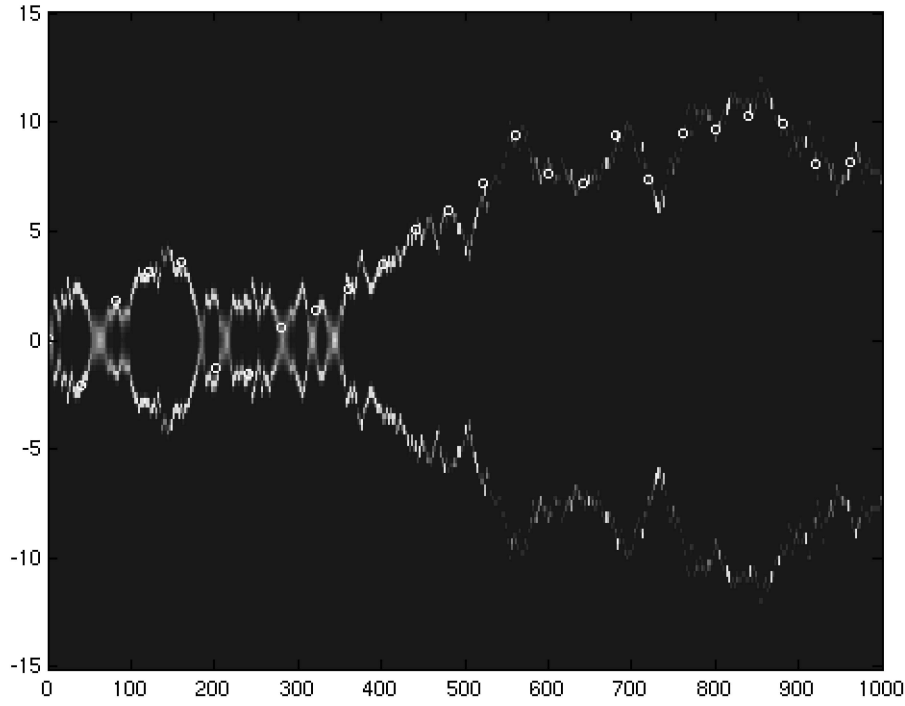
Fig. 5. Conditional probability density for one sample of signal and measurement models described by (126) using the DF algorithm.
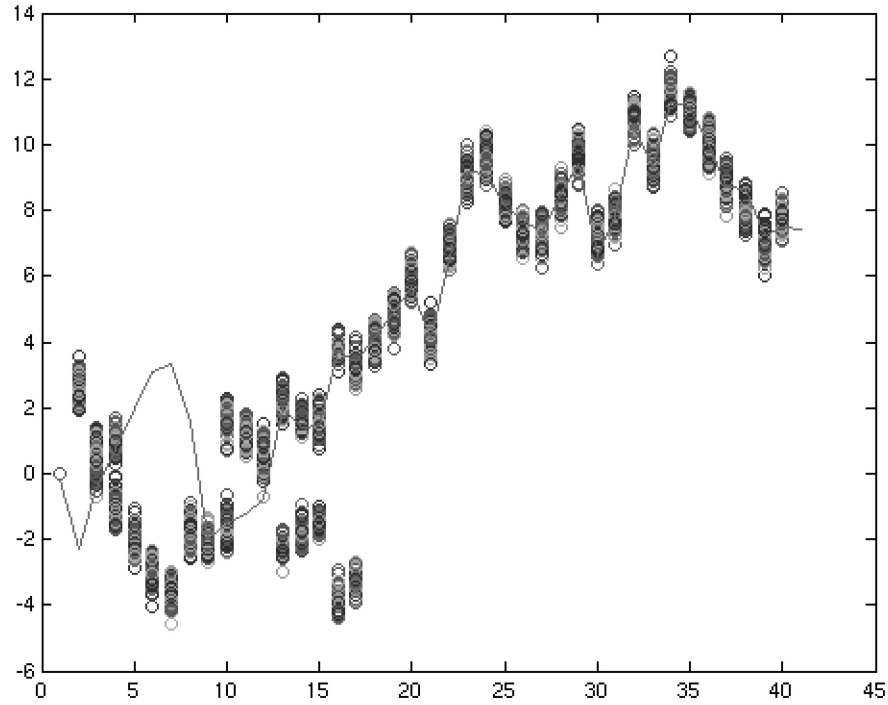

Fig. 6. Particle cloud for sample of signal and measurement models described by (126) using SIR particle filter. It is seen that particle filter is able to pick correct branch in latter part.

path, and the exponential of the action is the contribution from that path. In other words, the contributions of the paths are exponentially weighted according to the action. It is clear that the dominant contribution comes from paths near the path of least action.

This heuristic picture is seen to be consistent with intuition in the limit the noise is small. Let us consider the model discussed in Section V and assume that the noise is absent. Then, the evolution of the state is governed by a deterministic dynamical equation—the state evolves according to the zero noise dynamical law:

$$\dot{x}_i(t) = f_i(x(t)), \qquad i = 1, 2, \ldots, n. \tag{127}$$

According to the path integral formula, (44), when the noise is small, action is small only if the ordinary
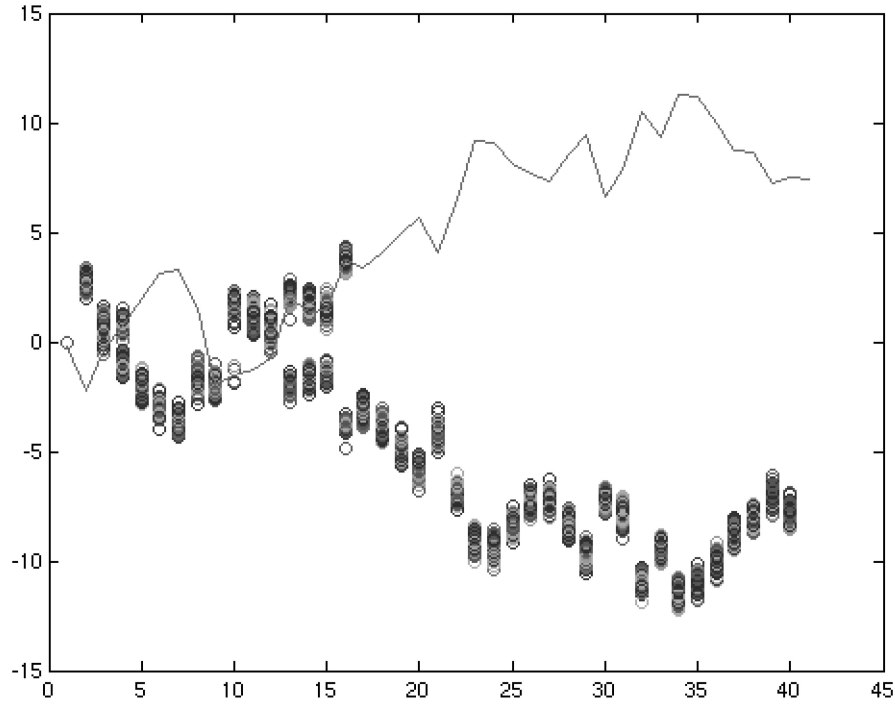
Fig. 7. Particle cloud for sample of signal and measurement models described by (126) using SIR particle filter with different initial seed. Particle filter picks wrong branch in latter part.

differential equation (ODE) (127) is satisfied (or nearly so). This implies that, to a good approximation, the particle moves along the deterministic trajectory specified by (127), as one would intuitively expect. The path integral formula gives a precise sense in which the deviation from the deterministic path is probabilistically possible as the variance of the noise $\hbar_\nu$ increases. This is analogous to the situation in quantum physics, where the role played by $\hbar_\nu$ is played by the Planck's constant $\hbar$.

B. Path Integral Monte Carlo

Recall that the prediction-correction step in the filtering process, i.e.,

$$p(t_k, x(t_k) \mid t_{k-1}, x(t_{k-1}) \mid Y(t_k)) \propto p(y(t_k) \mid x(t_k))$$

$$\times \int p(t_k, x(t_k) \mid t_{k-1}, x(t_{k-1})) p(t_{k-1}, x(t_{k-1}) \mid Y(t_{k-1})).$$

(128)

The following points needs to be addressed in assessing the accuracy of a filtering algorithm. Firstly, it is important to assess the accuracy of the DF approximation for $p(t_k, x(t_k) \mid t_{k-1}, x(t_{k-1}))$ for noninfinitesimal time steps. Secondly, the accuracy of the filtered output using the DF approximation needs to be assessed. It is noted that in the examples studied the results are promising. For lower dimensional problems, the filtered output can be computed efficiently using grid methods using sparse tensor techniques [33]. Note that this includes the case where

the diffusion matrix is not diagonal, a case not easy to solve using PDE methods (which are usually based on operator-splitting methods).

For higher dimensional problems, it is clear that a Monte-Carlo approach is essential. It is also essential if the time step is too large. The Monte-Carlo approach to computing path integrals is based on the use of a Metropolis type of algorithm to compute the finite-dimensional integral approximation of the path integral formulas. For an introductory discussion see [34].

However, the key point in filtering is that in addition to the transition probability density, the measurements are also known. This can be incorporated in the Monte Carlo algorithm. Specifically, for the model considered

$$d\mathbf{x}(t) = f(\mathbf{x}(t))dt + d\mathbf{v}(t), \qquad \mathbf{x}(t), f(\mathbf{x}(t)), \mathbf{v}(t) \in \mathbb{R}^n,$$

$$Q_{ij} = \hbar_\nu \delta_{ij},$$

(129)

$$\mathbf{y}(t) = h(\mathbf{x}(t)) + \mathbf{w}(t), \qquad \mathbf{y}(t), h(\mathbf{x}(t)), \mathbf{w}(t) \in \mathbb{R}^m,$$

$$R = I.$$

Then, from (128), it is apparent that one needs to sample $p(y(t_k) \mid x)p(t)k, x(t_k) \mid t_{k-1}, x(t_{k-1}))$, and then use it to compute the filtered conditional probability density.

In fact, something similar has been proposed in [35], [36]. They considered the quantity

$$p(x(t_0), x(t_1), \ldots, x(t) \mid y(t_1), y(t_2), \ldots, y(t)) \propto \exp(-U(t, t_0)).$$

Using the DF approximation, $U(t_0, t) = \sum_{k=1}^{k} U(t_i, t_{i-1}))$ is

$$U(t_k, t_{k-1})) = \frac{1}{2\hbar_\nu} \sum_{i=1}^{n} \left(\frac{x_i(t_k) - x_i(t_{k-1})}{t_{k-1} - t_k} - f_i(x^{(r)}(t_k))) \right)^2$$

$$+ r \sum_{i=1}^{n} \frac{\partial f_i}{\partial x_i}(x^{(r)}(t_k)) + \frac{1}{2} \sum_{l=1}^{m} (y_l(t_k) - h_l(x(t_k)))^2.$$

(130)

Strictly speaking, they do not do filtering which requires the additional steps of computing the transition probability density and integration; they merely sample from $U(t, t_0)$. 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. Some other recent related work has been done by Friston [37]–[39].

It is crucial to note the inclusion of the measurements in generating the samples. In contrast, in bootstrap particle filtering, only the state model is used to generate the samples, which are then weighted according to the measurements. Since the measurements are also used for generating the samples, it is likely to be a more robust solution than the bootstrap particle filter. In addition, one can do "path integral particle filtering" where one generates samples from the state model using the DF action only.

C. Comments on the Integral of S-T. Yau and S. S-T. Yau

A solution in terms of ordinary integrals was presented in [17]. The formal solution derived by S-T. Yau and S. S-T. Yau is of the form

a more general equation than the FPKfe. Secondly, the derivations here assume that the signal model noise is additive, whereas in [17] a weaker condition is assumed, namely that the diffusion matrix is orthogonal. However, unlike [17] the FPKfe is solved for the explicit time-dependent case.

X. CONCLUSION

In this paper the path integral formula has been derived for the FPKfe that arises from the explicitly time-dependent Langevin dynamical equation with additive noise with a rectangular diffusion vielbein. Note that it has also been used to tackle the continuous-continuous nonlinear filtering problem [40, 41]. This can be applied to solve the continuous-discrete filtering problems.

The path integral formulation has several advantages.

1) *Exact Solution*: It is possible to obtain exact expressions for the fundamental solution using path integral (and also less simply via operator methods). For example, for the case of Beněs drift an exact expression for the fundamental solution is obtained by noting that it is related to the quantum simple harmonic oscillator [42].

2) *Improved Actions*: The DF approximation is the simplest approximation for the one-step action. There is a large body of work on "improved actions" that are more accurate for larger time steps.

3) *Path Integral Monte Carlo and Variational Filtering*: Although lower dimensional problems can be tackled reliably using the DF algorithm using sparse tensor methods [33], it is clear that for higher dimensional problems Monte Carlo methods are more

$$u(t, x) = \int_{-\infty}^{\infty} \{d^n \xi\} (2\pi t)^{n/2} \exp \left(-\frac{1}{2t} \sum_{j=1}^{n} (x_j - y_j)^2 + \int_0^1 \sum_{i=1}^{n} (x_i - y_i) f_i(y + t(x, y)) dt \right) \left[1 + \sum_{i=1}^{\infty} \tilde{a}_i(x, y) t^i \right] \sigma_0(\xi)$$

(131)

for arbitrary initial condition $u(0, x) = \sigma_0(x)$ (for details see [17]). The important thing to note is that this is an infinite series in t. In addition, they present an estimate of the time interval on which their solution converges. In contrast, the path integral solution is an infinite-dimensional integral but with a much simpler structure.[6]

Also, note that the result in [17] differs from that obtained here in a few aspects. Firstly, they do not assume current conservation, and so solve

appropriate. As noted earlier, particle filters are one such solution. In fact, there is recent work on the use of path integral Monte Carlo (PIMC) methods for indirectly observable state estimation problems [36]. It is based on the application of Metropolis type algorithms to a "Hamiltonian" that is formed from the DF action ($r = 0$) and the measurement likelihood function. Further improvements are possible using improved actions, and in some cases exact expressions, for the fundamental solution. Other related work includes variational filtering [39].

The Feynman path integral method has been successfully used to solve challenging problems in quantum field theory not solvable using other

---

[6]It is also important to note that they have represented an infinite-dimensional path integral as a finite-dimensional ordinary integral. This may be useful in other areas, such as quantum mechanics and field theory.

approaches. Furthermore, from a conceptual point of view, the path integral formulation has proven to be extremely valuable. The insights from path integrals to problems in several other fields, especially quantum mechanics and field theory, has been tremendous. Hence it is a very promising framework for studying the nonlinear filtering problem.

APPENDIX I. GAUSSIAN INTEGRALS

It is remarkable that the foundation of the sophisticated methods of path integrals is based on Gaussian integrals. A Gaussian integral is defined as one where the integrand is an exponential function with an exponent quadratic in the variables. The simplest Gaussian integral

$$I = \int_{-\infty}^{\infty} e^{-x^2} dx \tag{132}$$

can by evaluated by noting that

$$I^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(x^2+y^2)} dx\,dy. \tag{133}$$

Transforming to polar coordinates with the substitutions

$$\begin{aligned} x &= r\cos\theta, & 0 < r \leq \infty \\ y &= r\sin\theta, & 0 < \theta \leq 2\pi \end{aligned} \tag{134}$$

so that the measure becomes $dx\,dy = r\,dr\,d\theta$, it follows that

$$\begin{aligned} I^2 &= \int_0^{2\pi} d\theta \int_0^{\infty} e^{-r^2} r\,dr \\ &= 2\pi \int_0^{\infty} e^{-z}\frac{dz}{2}, \qquad (z = r^2) \\ &= \pi. \end{aligned} \tag{135}$$

Hence,

$$\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi} \tag{136}$$

or more generally (assuming $a > 0$ so that the integral is convergent),

$$\begin{aligned} \int_{-\infty}^{\infty} e^{-ax^2} dx &= \int_{-\infty}^{\infty} e^{-z^2}\frac{dz}{\sqrt{a}}, \qquad z = \sqrt{a}x \\ &= \sqrt{\frac{\pi}{a}}. \end{aligned} \tag{137}$$

The general one-dimensional Gaussian integral can be evaluated by completing squares:

$$\begin{aligned} &\int_{-\infty}^{\infty} e^{-(ax^2+bx+c)} dx \\ &= e^{-c} \int_{-\infty}^{\infty} \exp\left[-a\left[\left(x+\frac{b}{2a}\right)^2 - \frac{b^2}{4a^2}\right]\right] dx \\ &= e^{((b^2/4a)-c)} \int_{-\infty}^{\infty} e^{-z'^2}\frac{dz'}{\sqrt{a}}, \qquad z' = \sqrt{a}\left(x+\frac{b}{2a}\right) \\ &= \sqrt{\frac{\pi}{a}} e^{((b^2/4a)-c)}. \end{aligned} \tag{138}$$

Define the n-dimensional measure

$$\{d^n x\} \equiv dx_1 dx_2 \cdots dx_n. \tag{139}$$

Then, it follows that

$$\begin{aligned} \int \{d^n x\} e^{-\sum_{i=1}^{n} a_i x_i^2} dx &= \prod_{i=1}^{n} \int_{-\infty}^{\infty} dx_i e^{-a_i x_i^2} \\ &= \sqrt{\frac{\pi^n}{\prod_{j=1}^{n} a_j}}, \qquad a_i > 0. \end{aligned} \tag{140}$$

Next, consider the quadratic form

$$\alpha \sum_{i,j=1}^{n} x_i A_{ij} x_j \tag{141}$$

where α is a positive real number and A is a real $n \times n$ symmetric matrix.[7] A real, symmetric matrix can always be diagonalized by an orthogonal transformation O:

$$A = O^T \Lambda(A) O. \tag{144}$$

Here, $\Lambda(A)$ is a diagonal matrix with its diagonal entries being the eigenvalues $\lambda_i(A)$ of A. Thus,

$$\begin{aligned} \alpha \sum_{i,j=1}^{n} x_i A_{ij} x_j &= \alpha \sum_{i,j=1}^{n} y_i \Lambda_{ij}(A) y_j \\ &= \alpha \sum_{i,j=1}^{n} \lambda_i(A) y_i^2, \qquad y_i = \sum_{j=1}^{n} O_{ij} x_j. \end{aligned} \tag{145}$$

Since the Jacobian is unity ($\det O = 1$),

$$\begin{aligned} \int \{d^n x\} e^{-\alpha \sum_{i,j=1}^{n} x_i A_{ij} x_j} &= \int \{d^n y\} e^{-\alpha \sum_{i=1}^{n} \lambda_i(A) y_i^2} \\ &= \left(\frac{\pi}{\alpha}\right)^{n/2} \frac{1}{\sqrt{\prod_{i=1}^{n} \lambda_i(A)}} \\ &= \left(\frac{\pi}{\alpha}\right)^{n/2} \frac{1}{\sqrt{\det A}}, \qquad \alpha > 0. \end{aligned} \tag{146}$$

---

[7]Note that any matrix can be written as a sum of a symmetric matrix (A^S) and an antisymmetric matrix because of the identity

$$A = \frac{A+A^T}{2} + \frac{A-A^T}{2} \equiv A^S + A^A, \qquad A_{ij}^S = A_{ji}^S, \quad A_{ij}^A = -A_{ji}^A. \tag{142}$$

However, since x_i are commuting variables, i.e., $x_i x_j = x_j x_i$,

$$\begin{aligned} \sum_{i,j=1}^{n} x_i A_{ij}^A x_j &= \sum_{i,j=1}^{n} x_j A_{ji}^A x_i, \qquad (i \leftrightarrow j) \\ &= -\sum_{i,j=1}^{n} x_j A_{ij}^A x_i = -\sum_{i,j=1}^{n} x_i A_{ij}^A x_j = 0. \end{aligned} \tag{143}$$

Hence, without loss of generality, A may be assumed to be a symmetric matrix.

The most general quadratic form can be written as follows:

$$\alpha\beta \sum_{i,j=1}^{n} x_i A_{ij} x_j + \beta \sum_{i=1}^{n} b_i x_i + c$$

$$= \alpha\beta \sum_{i,j=1}^{n} \left(x_i + \frac{1}{2\alpha} \sum_{k=1}^{n} (A^{-1})_{ik} b_k \right)$$

$$\times A_{ij} \left(x_j + \frac{1}{2\alpha} \sum_{k=1}^{n} (A^{-1})_{jk} b_k \right)$$

$$- \frac{\beta}{4\alpha} \sum_{i,j=1}^{n} b_i (A^{-1})_{ij} b_j + c. \tag{147}$$

Under the substitution

$$y_i = x_i + \frac{1}{2\alpha} \sum_{k=1}^{n} (A^{-1})_{ik} b_k \tag{148}$$

$dy_i = dx_i$ and the limits of integration are unchanged. Therefore, the general n-dimensional Gaussian integral is

$$Z(A, b, c) \equiv \int \{d^n x\} \exp\left[-\beta \left[\alpha \sum_{i,j=1}^{n} x_i A_{ij} x_j + \sum_{i=1}^{n} b_i x_i + c/\beta \right] \right]$$

$$= \int \{d^n y\} \exp\left[-\sum_{i,j=1}^{n} y_i (\alpha\beta A)_{ij} y_j \right.$$

$$\left. + (\beta/4\alpha) \sum_{i,j=1}^{n} b_i (A^{-1})_{ij} b_j - c \right]$$

$$= \left[\left(\frac{\pi}{\alpha\beta} \right)^{n/2} \frac{e^{-c}}{\sqrt{\det A}} \right] e^{(\beta/4\alpha)\Sigma_{i,j=1}^{n} b_i A_{ij}^{-1} b_j}. \tag{149}$$

The normalized measure

$$[d^n x] \equiv \left[\sqrt{\det A} \left(\frac{\alpha\beta}{\pi} \right)^{n/2} \right] \tag{150}$$

is defined so that

$$\int [d^n x] e^{-\alpha \Sigma_{i,j=1}^{n} x_i A_{ij} x_j}$$

$$= \int \left\{ d^n x \left(\frac{\alpha\beta}{\pi} \right)^{n/2} \sqrt{\det A} \right\} e^{-\alpha \Sigma_{i,j=1}^{n} x_i A_{ij} x_j}$$

$$= 1. \tag{151}$$

In the calculation of the expectation values of polynomials in random vectors with a Gaussian distribution, integrals of the form arise:

$$\int [d^n x] x_{k_1} x_{k_2} \cdots x_{k_n} e^{-\alpha \Sigma_{i,j=1}^{n} x_i A_{ij} x_j}$$

$$= \frac{\int \{d^n x\} x_{k_1} x_{k_2} \cdots x_{k_l} e^{-\alpha\beta \Sigma_{i,j=1}^{n} x_i A_{ij} x_j}}{\int \{d^n x\} e^{-\alpha\beta \Sigma_{i,j=1}^{n} x_i A_{ij} x_j}}. \tag{152}$$

From (149), it follows from repeated differentiation with respect to βb, that

$$\langle \mathbf{x}_{k_1} \mathbf{x}_{k_2} \cdots \mathbf{x}_{k_l} \rangle$$

$$= \left(\left(\frac{\alpha\beta}{\pi} \right)^{n/2} \sqrt{\det A} \right)$$

$$\times \left[\frac{(-1)^l}{\beta^l} \frac{\partial}{\partial b_{k_1}} \frac{\partial}{\partial b_{k_2}} \cdots \frac{\partial}{\partial b_{k_l}} Z(A, b, 0) \right] \Bigg|_{b=0}$$

$$= \left(\frac{(-1)^l}{\beta^l} \frac{\partial}{\partial b_{k_1}} \frac{\partial}{\partial b_{k_2}} \cdots \frac{\partial}{\partial b_{k_l}} e^{(\beta/4\alpha)\Sigma_{i=1}^{n} b_i (A^{-1})_{ij} b_j} \right) \Bigg|_{b=0}. \tag{153}$$

In fact, if $F(x_1, x_2, \ldots, x_n)$ is a power series in x_i:

$$\langle F(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n) \rangle = \left(F \left(-\frac{1}{\beta} \frac{\partial}{\partial b_1}, -\frac{1}{\beta} \frac{\partial}{\partial b_2}, \ldots, -\frac{1}{\beta} \frac{\partial}{\partial b_n} \right) \right.$$

$$\left. \times e^{(\beta/4\alpha)\Sigma_{i,j=1}^{n} b_i A_{ij}^{-1} b_j} \right) \Bigg|_{b=0}. \tag{154}$$

An example is the two-point function:

$$\langle \mathbf{x}_i \mathbf{x}_j \rangle = \frac{(-1)^2}{\beta^2} \frac{\beta}{2\alpha} (A^{-1})_{ij}$$

$$= \frac{1}{2\alpha\beta} (A^{-1})_{ij}. \tag{155}$$

The expectation can be simplified further. Note that if l is odd, the expectation vanishes. When l is even, it is easy to show (using the method of induction) that:

$$\langle x_{k_1} x_{k_2} \cdots x_{k_l} \rangle$$

$$= \sum_{\text{Permutations of } \{k_1, k_2, \ldots, k_l\}} \left(\frac{\beta}{2\alpha} \right)^l A_{k_{P_1} k_{P_2}}^{-1} \cdots A_{k_{P_{l-1}} k_{P_l}}^{-1}$$

$$= \sum_{\text{Permutations of } \{k_1, k_2, \ldots, k_l\}} \left(\frac{\beta}{2\alpha} \right)^l \langle x_{P_1} x_{k_{P_2}} \rangle \cdots \langle x_{k_{P_{l-1}}} x_{k_{P_l}} \rangle. \tag{156}$$

This can be formally generalized to an infinite number of variables, i.e., fields, and is then referred to as the Wick's theorem for bosonic fields in quantum field theory.

APPENDIX II.   DELTA-FUNCTION AND CONSTRAINTS

The delta function[8] has several representations. The most useful representation for our purposes is the following representation of the n-dimensional delta

---

[8] Strictly speaking, the delta function should be called the delta distribution defined rigorously as a limit of sequence of certain functions; it is not mathematically meaningful to view the Dirac delta function as a function. However, for most of our applications, the manipulations done here are valid if the delta function is always considered to be a part of the integrand.

function

$$\delta^n(x - x') \equiv \frac{1}{(2\pi)^n} \int \{d^n k\} e^{i\sum_{i=1}^n k_i(x_i - x_i')}$$

$$\equiv \int [d^n k] e^{i\sum_{i=1}^n k_i(x_i - x_i')}. \tag{157}$$

The Fourier transform and its inverse are defined as follows:

$$\tilde{f}(k) = \int \{d^n x\} e^{-i\sum_{i=1}^n k_i x_i} f(x)$$

$$f(x) = \frac{1}{(2\pi)^n} \int \{d^n k\} e^{i\sum_{i=1}^n k_i x_i} \tilde{f}(k). \tag{158}$$

Consider the problem of writing down the expression of a function $\sigma(x)$ at the point where $f(x) = 0$ without actually solving the constraint equation. Assume that the solution to the equation $f(x) = 0$ is unique. Then,

$$\sigma(x)|_{f(x)=0} = \sigma(x_c), \qquad f(x_c) = 0$$

$$= \int \{d^n x\} \delta(x - x_c)\sigma(x). \tag{159}$$

From the sequence of identities

$$1 = \int \{d^n y\} \delta^n(y)$$

$$= \int \{d^n f(x)\} \delta^n(f(x))$$

$$= \int \{d^n x\} J(x) \delta^n(f(x)), \qquad J = \det\left(\frac{\partial f_i}{\partial x_j}(x)\right)$$

$$= \int \{d^n x\} \delta^n(x - x_c) \tag{160}$$

it follows that

$$\delta^n(f(x)) = \frac{\delta^n(x - x_c)}{J(x)} \tag{161}$$

$$\delta^n(x - x_c) = \delta^n(f(x))J(x).$$

Hence,

$$\sigma(x) = \int \{d^n x\} \delta^n(f(x))J(x)\sigma(x). \tag{162}$$

Using the Fourier integral representation of the delta function in (157), it is clear that

$$\sigma(x)|_{f(x)=0} = \frac{1}{(2\pi)^n} \int \{d^n x\}\{d^n \lambda\} e^{i\lambda f(x)} J(x)\sigma(x). \tag{163}$$

APPENDIX III.   FUNCTIONAL CALCULUS: A BRIEF NOTE

Recall that a function $f(x)$ gives a number for each point x. The infinite-dimensional generalization of a function is called a functional $f[x(t)]$, which gives a number for each function $x(t)$. Loosely speaking, the

functional derivative may be defined analogous to the ordinary derivative as follows:

$$\frac{\delta f[x(t)]}{\delta x(t')} = \lim_{\epsilon \to 0} \frac{f[x(t) + \epsilon\delta(t - t')] - f[x(t)]}{\epsilon}. \tag{164}$$

Alternatively, the functional derivative $(\delta f/\delta x(t'))[x(t)]$ of the functional $f[x(t)]$ with respect to variation of the function $x(t)$ is defined by the following equation

$$f[x(t) + \eta(t)] = f(x(t)) + \int \frac{\delta f}{\delta x(s')}[x(s)]\eta(s')ds' + \cdots. \tag{165}$$

Note that the functional derivative of a functional is also a functional. Functional differentiation obeys the standard algebraic rules (linearity and Leibnitz's rule):

$$\frac{\delta}{\delta x(t')}\left[\sum_{i=1}^n f_i[x(t)]\right] = \sum_{i=1}^n \frac{\delta f_i}{\delta x(t')}[x(t)] \tag{166}$$

$$\frac{\delta}{\delta x(t')}[f_1[x(t)]f_2[x(t)]] = f_1[x(t)]\frac{\delta}{\delta x(t')}f_2[x(t)]$$

$$+ f_2[x(t)]\frac{\delta}{\delta x(t')}f_1[x(t)]. \tag{167}$$

Just as the derivative of x with respect to x is 1, the functional derivative of $x(t)$ with respect to $x(t')$ is the unit matrix in infinite dimensions, namely the delta function:

$$\frac{\delta x(t)}{\delta x(t')} = \delta(t - t'). \tag{168}$$

The functional derivatives of a formal power series in function $x(t)$ can be seen to be analogous to the ordinary derivative of a power series in the variable t. Specifically, it is straightforward to verify that if

$$f[x(t)] = \sum_{n=0}^\infty \frac{1}{n!} \int dt_1 \cdots dt_n f^{(n)}(t_1, t_2, \ldots, t_n)x(t_1)\cdots x(t_n) \tag{169}$$

then

$$f^{(r)}(t_1, t_2, \ldots, t_r) = \left\{\left(\prod_{i=1}^r \frac{\delta}{\delta x(t_i)}\right)f\right\}\bigg|_{x=0}. \tag{170}$$

In particular, a functional representable as a power series in functions, like the exponential functional, can be functionally differentiated using this result. Thus

$$\frac{\delta}{\delta J(x)} \exp\left(\int dx' J(x')A(x')\right)$$

$$= A(x)\exp\left(\int dx' J(x')A(x')\right). \tag{171}$$

The functional delta function may be written as

$$\delta(f[x(t)]) = \int [d\lambda(t)] \exp\left(i\int dt\lambda(t)f(x(t))\right). \tag{172}$$

Note that (160) becomes

$$1 = \int [\mathcal{D}f[x(t')]] \det\left[\frac{\delta f}{\delta x(t')}[x(t)]\right] \delta(f[x(t)]).$$
(173)

In the application at hand, determinants of operators (matrices) arise in Gaussian integration in the infinite (finite)-dimensional case. Of particular interest is the operator $\det(\delta(x-y) + K(x,y))$, for some operator $K(x,y)$, where it is assumed that traces of all powers of K exist. From the identity

$$\ln\det M = \operatorname{tr}\ln M$$
(174)

it follows that

$$\ln\det[1+K] = \int dx K(x,x) - \frac{1}{2}\int dx_1 dx_2 K(x_1,x_2)K(x_2,x_3)$$
$$+ \cdots + \frac{(-1)^{n+1}}{n}\int dx_1 \cdots dx_n K(x_1,x_2)K(x_2,x_3)$$
$$\cdots K(x_n,x_1) + \cdots .$$
(175)

Note that if $K(x,y) = \theta(x-y)\tilde{K}(x,y)$, only the first term is non-zero.

REFERENCES

[1] Jazwinski, A. H.
     Stochastic Processes and Filtering Theory.
     New York: Dover, 2007.

[2] Kalman, R. E.
     A new approach to linear filtering and prediction problems.
     Transactions of ASME, Journal of Basic Engineering, **82D** (Mar. 1960), 35–45.

[3] Kalman, R. E. and Bucy, R. S.
     New results in linear filtering and prediction problems.
     Transaction of ASME, Journal of Basic Engineering, **83** (1961), 95–108.

[4] Kailath, T., Sayed, A. H., and Hassibi, B.
     Linear Estimation.
     Upper Saddle River, NJ: Prentice-Hall, 2000.

[5] Sarkka, S.
     On unscented Kalman filtering for state estimation of continuous-time nonlinear systems.
     IEEE Transactions on Automatic Control, **52**, 9 (Sept. 2007), 1631–1641.

[6] Elliott, R. and Krishnamurthy, V.
     New finite-dimensional filters for parameter estimation of discrete-time linear Gaussian models.
     IEEE Transactions on Automatic Control, **44**, 5 (May 1999), 938–951.

[7] Farina, A., Benvenuti, D., and Ristic, B.
     A comparative study of the Benes filtering problem.
     Signal Processing, **82**, 2 (2002), 133–147.

[8] Thomée, V.
     Finite difference methods for linear parabolic equations.
     In *Handbook of Numerical Analysis*, vol. 1, Amsterdam, Holland: North-Holland Publishing Co., 1990, pp. 5–196.

[9] Marchuk, G. I.
     Splitting and alternating direction methods.
     In *Handbook of Numerical Analysis*, vol. 1, Amsterdam, Holland: North-Holland Publishing Co., 1990, pp. 197–462.

[10] Canuto, C., et al.
     Spectral Methods: Fundamentals in Single Domains.
     New York: Springer, 2006.

[11] Thomée, V.
     Galerkin Finite Element Methods for Parabolic Problems (Springer Series in Computational Mathematics) vol. 25.
     New York: Springer, 1997.

[12] Challa, S. and Bar-Shalom, Y.
     Nonlinear filter design using Fokker-Planck-Kolmogorov probability density evolutions.
     IEEE Transactions on Aerospace and Electronic Systems, **36**, 1 (Jan. 2000), 309–315.

[13] Lototsky, S. and Rozovskii, B.
     Recursive nonlinear filter for a continuous-discrete time model: Separation of parameters and observations.
     1998. [Online], available:
     citeseer.ist.psu.edu/lototsky98recursive.html.

[14] Gordon, N., Salmond, D., and Smith, A.
     Novel approach to nonlinear/non-Gaussian Bayesian state estimation.
     IEE Proceedings F—Radar and Signal Processing, **140**, 2 (Apr. 1993), 107–113.

[15] Maskell, S., et al.
     Tracking using a radar and a problem specific proposal distribution in a particle filter.
     IEE Proceedings—Radar, Sonar and Navigation, **152**, 5 (Oct. 2005), 315–322.

[16] Arulampalam, M., et al.
     A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking.
     IEEE Transactions on Signal Processing, **50**, 2 (Feb. 2002), 174–188.

[17] Yau, S-T. and Yau, S. S-T.
     Explicit solution of a Kolmogorov equation.
     Applied Mathematics and Optimization, **34**, 3 (1996), 231–266.

[18] Dirac, P. A. M.
     The Lagrangian in quantum mechanics.
     Physikalische Zeitschrift der Sowjetunion, **3** (1933), 64–72.

[19] Dirac, P. A. M.
     The Principles of Quantum Mechanics (4th ed.).
     New York: Oxford, 1982.

[20] Feynman, R. P. and Hibbs, A. R.
     Quantum Mechanics and Path Integrals.
     New York: McGraw-Hill, 1965.

[21] Feynman, R. P.
     Space-time approach to non-relativistic quantum mechanics.
     Reviews of Modern Physics, **20** (1948), 367–387.

[22] Siegel, W.
     Fields.
     http://www.arxiv.org/abs/hep-th/9912205, 1999. [Online], available: http://www.citebase.org/abstract?id=oai: arXiv:hep-th/9912205.

[23] Glimm, J. and Jaffe, A.
     Quantum Physics: A Functional Integral Point of View (2nd ed.).
     New York: Springer, May 1987.

[24] Rothe, H. J.
     Lattice Gauge Theories: An Introduction (3rd ed.).
     Singapore: World Scientific Publishing Co., Ltd., May 2005.

[25] Risken, H.
     The Fokker-Planck Equation: Methods of Solution and Applications (2nd ed.).
     New York: Springer, 1999.

[26] Zinn-Justin, J.
Quantum Field Theory and Critical Phenomena
(International Series in Monographs on Physics).
New York: Oxford, 2002.

[27] Liang, Z. and Yau, S. S-T.
Wavelet-Galerkin method for the Kolmogorov equation.
Mathematical and Computer Modelling, **40**, 9–10 (Nov. 2004), 1093–1121.

[28] Balaji, B.
Continuous-discrete path integral filtering.
Entropy, **11**, 3 (2009), 402–430. [Online], available:
http://www.mdpi.com/1099-4300/11/3/402.

[29] Bharucha-Reid, A. T.
Elements of the Theory of Markov Processes and Their Applications.
New York: McGraw-Hill, 1960.

[30] Moral, P. D.
Feynman-Kǎc Formulae.
New York: Springer, Mar. 2004.

[31] Budhiraja, A., Chen, L., and Lee, C.
A survey of numerical methods for nonlinear filtering problems.
Physica D: Nonlinear Phenomena, **230** (2007), 27–36.

[32] Chen, L., et al.
PFLib—An object oriented MATLAB toolbox for particle filtering.
In *Proceedings to SPIE Signal Processing, Sensor Fusion, and Target Recognition* XVI, vol. 6567, Apr. 2007.

[33] Balaji, B.
Continuous-discrete filtering using the Dirac-Feynman algorithm.
In *IEEE Radar Conference* (RADAR'08), Rome, Italy, May 2008.

[34] Lepage, P. G.
Lattice QCD for novices.
June 2005. [Online], available: http://arxiv.org/abs/hep-lat/0506036.

[35] Alexander, F. J., Eynik, G. L., and Restrepo, J. M.
Accelerated Monte Carlo for optimal estimation of time series.
Journal of Statistical Physics, **119** (2005), 1331–1345.

[36] Restrepo, J. M.
A path integral method for data assimilation.
Physica D Nonlinear Phenomena, **237** (Jan. 2008), 14–27.

[37] Friston, K., et al.
Variational free energy and the laplace approximation.
NeuroImage, **34**, 1 (2007), 220–234. [Online], available:
http://www.sciencedirect.com/science/article/B6WNP-4M57H65-1/2/e0c2e1ec4be9a4bb669a5a791685cc6e.

[38] Friston, K., Trujillo-Barreto, N., and Daunizeau, J.
DEM: A variational treatment of dynamic systems.
NeuroImage, **41**, 3 (2008), 849–885. [Online], available:
http://www.sciencedirect.com/science/article/B6WNP-4S19RH8-1/2/48c1aa6f77adaeeaba06cd57573a986d.

[39] Friston, K.
Variational filtering.
NeuroImage, **41**, 3 (2008), 747–766. [Online], available:
http://www.sciencedirect.com/science/article/B6WNP-4S3G3X5-D/2/a3e8b19e56c4993a5b76ad6640a6e5ca.

[40] Balaji, B.
Estimation of indirectly observable Langevin states: Path integral solution using statistical physics methods.
Journal of Statistical Mechanics: Theory and Experiment, **2008**, 01 (2008), P01014.

[41] Balaji, B.
Universal nonlinear filtering using path integrals II: The continuous-continuous model with additive noise.
PMC Physics A, **3**, 2 (Feb. 10, 2009).

[42] Balaji, B.
The exact fundamental solution for the Benes tracking problem.
In I. Kadar (Ed.), *SPIE*, **7336**, 1 (2009), p. 73360V, doi:10.1117/12.817736. Available: http://link.aip.org/link/?PSI/7336/73360V/1.

Bhashyam Balaji was born in Bangalore, India in 1970 and graduated from Kendriya Vidyalaya, Sector 8, R. K. Puram, New Delhi, India in 1987. He received his B.Sc. (Hons) degree in physics from St. Stephen's College, University of Delhi, India and Ph.D. in theoretical particle physics from Boston University, Boston, MA, in 1997.

Since 1998, he has been a scientist at the Defence Research and Development Canada, Ottawa, Canada. His radar research interests include synthetic aperture radar, interferometry, space-time adaptive processing, multi-target tracking, and application of stochastic context-free grammars to syntactic tracking. His recent research also includes the application of Feynman path integral and quantum field theory methods to the problems of nonlinear filtering and stochastic control.