

Stochastic filtering for diffusion processes with level crossings

Agostino Capponi, Ibrahim Fatkullin, and Ling Shi

Abstract—We provide a general framework for computing the state density of a noisy system given the sequence of hitting times of predefined thresholds. Our method relies on eigenfunction expansion corresponding to the Fokker-Planck operator of the diffusion process. For illustration, we present a particular example in which the state and the noise are one-dimensional Gaussian processes and observations are generated when the magnitude of the observed signal is a multiple of some threshold value. We present numerical simulations confirming the convergence and the accuracy of the recovered density estimator. Applications of the filtering methodology will be illustrated.

Index Terms—Nonlinear Filtering, Fokker-Planck equation, Diffusion processes

I. INTRODUCTION

Many applications in science and engineering require estimation of the state of a system on the basis of noisy observations. When both the state and the observation processes are Gaussian and observations are generated continuously the optimal filter is the well-known Kalman-Bucy filter [1]. However, in many practical cases the above assumptions fail and a different approach, e.g., the nonlinear filtering, must be employed.

The modern theory of nonlinear filtering was developed between late 1960s and early 1970s in works of Kallianpur and Striebel [2], [3], Kushner [4], and in the case of diffusion processes by Zakai [5]. The martingale approach to nonlinear filtering, based on the innovations process and on the systematic use of martingale representation theorems, was developed by Kailath and Frost in [6].

We consider diffusion processes whose drift and volatility coefficients are sufficiently smooth functions and the corresponding forward Kolmogorov (Fokker-Planck) operator is uniformly elliptic and has a complete set of eigenfunctions in appropriate Hilbert spaces. This allows us to use eigenfunction expansion of the joint state-noise density solving the problem explicitly. Although in the particular example that we consider here, the governing equations are linear, the method is equally applicable to a more general class of time-homogeneous diffusion processes. Note also that since observations are comprised by the hitting times of some predefined thresholds, the filtering

problem is nontrivial even when the processes themselves are Gaussian.

Assume that the observation (measurement) process is represented as

$$Y_t = X_t^{(1)} + X_t^{(2)}, \quad (1)$$

where $X_t^{(1)}$ is the actual state of the system and $X_t^{(2)}$ is the noise component. Now, suppose that we do not know the complete sample paths of Y_t and only receive signals when Y_t hits thresholds γk , $k = 0, \pm 1, \pm 2, \dots$ (for some fixed $\gamma > 0$). More precisely, once a signal corresponding to some k is received, the next signal is generated when Y_t hits the value of $\gamma(k \pm 1)$. Accordingly, we introduce the overlapping domains

$$\Omega_k = \{\mathbf{x} = (x_1, x_2) : |x_1 + x_2 - k\gamma| < \gamma\}. \quad (2)$$

The boundary $\partial\Omega_k$ consists of two parallel lines that belong to Ω_{k-1} and Ω_{k+1} , denoted by Γ_{k-1} and Γ_{k+1} respectively, see Figure 1 for illustration.

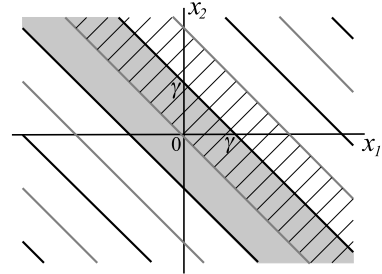


Fig. 1. Partition of the plane into domains Ω_k given by equation (2). Ω_0 is shaded grey, Ω_1 is marked with diagonal strokes.

Filtering Problem: We study the following problem. Suppose the initial (at $t = t_0 = 0$) probability density of $\mathbf{X}_t = (X_t^{(1)}, X_t^{(2)})$ is $p_0(\mathbf{x})$ with support in Ω_0 . Given an increasing sequence of positive reals (signal times)

$$0 = t_0 < t_1 < t_2 < \dots \quad (3)$$

and a sequence of integers (threshold indices)

$$0 = k_0, k_1, k_2, \dots, \quad \text{such that} \quad k_{j+1} = k_j \pm 1, \quad (4)$$

reconstruct the conditional probability density function $p_t(\mathbf{x})$, where the conditioning set is defined as

$$\mathcal{F}_t = \{\tau_j = t_j, i_j = k_j \text{ for all } j \text{ such that } t_j < t\}. \quad (5)$$

Here τ_j are the hitting times of the boundaries Γ_{i_j} , i.e., i_j and τ_j are random variables defined recursively as (setting $\tau_0 = 0$)

$$\begin{aligned} i_j &= (X_{\tau_j}^{(1)} + X_{\tau_j}^{(2)})/\gamma, \\ \tau_{j+1} &= \inf\{s > \tau_j : \mathbf{X}_s \notin \Omega_{i_j}\}. \end{aligned} \quad (6)$$

Agostino Capponi is with the Department of Industrial Engineering, Purdue University, 47906, West Lafayette, Email: capponi@purdue.edu.

Ibrahim Fatkullin is with the Department of Mathematics, University of Arizona, 85721-0089, Tucson, Email: ibrahim@math.arizona.edu.

Ling Shi is with the Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology Clear Water Bay, Kowloon, Hong Kong, Email: eesling@ust.hk.

Ibrahim Fatkullin is supported by the NSF grant DMS-0807332.

To the best of our knowledge this type of filtering problem has not received sufficient attention in the literature, even though it occurs naturally in various contexts. For example, a framework similar to ours has been suggested for structural credit risk modeling [7], where the state equation models the dynamics of the asset value of the firm, typically assumed one-dimensional. However, the authors do not provide analytic expressions for the density of states. The work which is most closely related to ours is Cvitanic et al. [8], who develop a non-linear filtering method to estimate the volatility from high frequency security prices which are observed discretely, and possibly at random times. Aihara and Bagchi [9] derive the exact volatility filter for a discrete version of the Heston model with the aid of the particle filter algorithm. Most recently, Ding et al. [10] derive a mathematical model of a non-uniformly sampled system which allows reconstructing uniquely the continuous time model from its non-uniformly sampled discrete-time counterpart. Their sampling scheme may be used in our context to construct the continuous time filtering model consistent with the available data or sensor observations, after which the filtering methodology can be applied. The problem of estimating the parameters of filtering models where the output error system has colored measurement noises is considered in [11], where the authors resort to multi-innovation identification theory to construct a stochastic gradient algorithm for parameter identification.

General Result: We show the following: suppose that $X_t^{(1)}$ and $X_t^{(2)}$ are time-homogeneous diffusion processes whose joint probability density admits expansion in terms of the eigenfunctions of the Fokker-Planck operator, \mathcal{L}^* . Let t_j be the last hitting time before time t . Then, we provide a recursive analytical expression for the joint probability density of the system at time $t > t_j$, given all the information about collisions up to time t , in terms of the quantities computed at the previous hitting time t_{j-1} .

The method is specifically designed and efficient for sufficiently large values of the threshold γ , as in this case the optimal filter density can be efficiently and accurately reconstructed using a small number of terms.

In the rest of the paper we first derive general formulas for the filtering framework presented above, and then specialize our analysis to the case when $X_t^{(1)}$ is a Wiener process and $X_t^{(2)}$ is an Ornstein-Uhlenbeck process.

II. SOLUTION OF THE FILTERING PROBLEM

We provide some background and notation in Section II-A. We give the main result in Section II-B. We provide a short discussion of the error analysis in Section II-C. We provide the steps for the derivation of the filter equations in Section II-D.

A. Notation and Background

Let \mathbf{X}_t be a stochastic process satisfying the stochastic differential equation

$$d\mathbf{X}_t = \mathbf{b}(\mathbf{X}_t)dt + \boldsymbol{\sigma}(\mathbf{X}_t)d\mathbf{W}_t, \quad (7)$$

where $\mathbf{W}_t = (W_t^{(1)}, W_t^{(2)}, \dots, W_t^{(m)})$ is an m -dimensional Wiener process. The generator of the Ito diffusion process in Eq. (7), is denoted by \mathcal{L} , and defined by

$$\mathcal{L}f(\mathbf{x}) = \sum_{i=1}^n b_i(\mathbf{x}) \frac{\partial f(\mathbf{x})}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^n \sigma_{ij}(\mathbf{x}) \frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j}. \quad (8)$$

The adjoint of the generator \mathcal{L} , is denoted by \mathcal{L}^* , and defined by

$$\mathcal{L}^*f(\mathbf{x}) = - \sum_{i=1}^n \frac{\partial [b_i(\mathbf{x})f(\mathbf{x})]}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^n \frac{\partial^2 [\sigma_{ij}(\mathbf{x})f(\mathbf{x})]}{\partial x_i \partial x_j} \quad (9)$$

The transition probability density $p_{t,s}(\mathbf{x}|\mathbf{y})$ for $\mathbf{x}_t = \mathbf{x}$ given $\mathbf{x}_s = \mathbf{y}$, where $t > s$, satisfies the Fokker-Planck partial differential equation (with appropriate boundary conditions):

$$\begin{aligned} \frac{\partial}{\partial t} p_{t,s}(\mathbf{x}|\mathbf{y}) &= \mathcal{L}^* p_{t,s}(\mathbf{x}|\mathbf{y}) \\ p_{s,s}(\mathbf{x}|\mathbf{y}) &= \delta(\mathbf{x} - \mathbf{y}). \end{aligned} \quad (10)$$

B. Main Result

The joint probability density of the system at time $t > t_j$, given all the information about collisions up to time t , may be reconstructed as

$$\begin{aligned} p_t(\mathbf{x}) &= \frac{\tilde{p}_t(\mathbf{x})}{\int_{\Omega_{k_j}} \tilde{p}_t(\mathbf{x}) d\mathbf{x}}, \\ \tilde{p}_t(\mathbf{x}) &= \sum_{n=0}^{\infty} e^{\lambda_n^{(k_j)}(t-t_j)} \phi_n^{(k_j)}(\mathbf{x}) \tilde{p}_n^{(j)}. \end{aligned} \quad (11)$$

The coefficients $\tilde{p}_n^{(j)}$ are computed recursively according to

$$\begin{aligned} \tilde{p}_n^{(j)} &= \sum_{m=0}^{\infty} e^{\lambda_m^{(k_{j-1})}(t_j-t_{j-1})} T_{nm}^{(k_{j-1}, k_j)} \tilde{p}_m^{(j-1)}, \\ \tilde{p}_n^{(0)} &= \int_{\Omega_0} p_0(\mathbf{y}) \varphi_n^{(0)}(\mathbf{y}) d\mathbf{y}. \end{aligned} \quad (12)$$

Here $\{\phi_n^{(k)}\}$, $n = 0, 1, \dots$ are the eigenfunctions (vanishing on the boundary) of \mathcal{L}^* in domains Ω_k ; $\lambda_n^{(k)}$ are the respective eigenvalues (in decreasing order). The functions $\{\varphi_n^{(k)}\}$ are such that jointly with $\{\phi_n^{(k)}\}$ they form bi-orthogonal sets in respective domains, i.e. $\int_{\Omega_k} \phi_m^{(k)}(\mathbf{x}) \varphi_n^{(k)}(\mathbf{x}) d\mathbf{x} = \delta_{mn}$. The transfer coefficients $T_{nm}^{(k_{j-1}, k_j)}$ may be computed as follows:

$$T_{nm}^{(k_{j-1}, k_j)} = \int_{\Gamma_{k_j}} \varphi_n^{(k_j)}(\mathbf{y}) \mathcal{F} \phi_m^{(k_{j-1})}(\mathbf{y}) \cdot \mathbf{n} d\mathbf{y}, \quad (13)$$

where \mathcal{F} is the flux operator corresponding to \mathcal{L} , i.e., $\mathcal{L}^* = \nabla \cdot \mathcal{F}$; \mathbf{n} is the unit vector normal to Γ_{k_j} outward with respect to $\Omega_{k_{j-1}}$.

Notice that for a practical implementation we need to truncate the infinite series. The computational complexity of the method is driven by the number of hitting times and the number of approximating eigenfunctions used. If the number of hitting times before time t is T , then we need to iterate the computation of \tilde{p} 's via formula (12) T times. Since each iteration accounts for multiplying the vector $\tilde{p}^{(j-1)}$ by the transfer matrix T_{mn} , assuming that we use N eigenfunctions in

each domain, we get $O(N^2)$ operations per iteration. Thus the overall complexity of the filtering methodology is $O(TN^2)$.

As already pointed out in the introduction, the method becomes efficient when the threshold γ is large. This is evident from the expression of $\tilde{p}_n^{(j)}$ in Eq. (12), which shows the exponentially-faster decay of contributions corresponding to the smaller eigenvalues. Thus, as larger values of γ result in larger inter-arrival times between consecutive boundary hits, fewer eigenfunctions are needed to approximate the coefficients $\tilde{p}_n^{(j)}$ with high degree of accuracy.

C. Error Analysis

For practical purposes, summation in equations (11) and (12) must be truncated which introduces an approximation error. For example, provided that the error in coefficients $p_n^{(j)}$ has l^2 -norm r_j , a rough estimate on the L^2 -norm of the error in the approximate probability density, $\hat{p}_t(\mathbf{x})$, is given by

$$\|p_t(\mathbf{x}) - \hat{p}_t(\mathbf{x})\| \leq (C_0 + C_1 r_j + C_2 r_j^2) e^{(\lambda_1^{(j)} - \lambda_0^{(j)})(t - t_j)} \quad (14)$$

where C_i are positive constants which depend on the order of truncation and specifics of the problem, i.e., domains Ω_k and the generator \mathcal{L} . Truncating the sum in equation (12) introduces similar errors:

$$r_j \leq (C_3 + C_4 r_{j-1} + C_5 r_{j-1}^2) e^{(\lambda_1^{(j-1)} - \lambda_0^{(j-1)})(t_j - t_{j-1})}. \quad (15)$$

The constants C_i diminish as the order of truncation increases; the exponential factor depends on the difference between the first two eigenvalues, since it controls the rate of convergence to the stationary distribution of the process conditioned on staying within the domains Ω_k .

Derivation of the error estimates (14) and (15) is quite straightforward, but rather cumbersome; therefore, for the purposes of this technical note, we only provide the principal idea behind it. Between the hitting times the process remains within the respective domains Ω_k , thus the (unnormalized) probability density of the system satisfies the Fokker-Planck equation (18) with zero boundary conditions. The normalized probability density function is then exponentially attracted to the stationary distribution,

$$\phi_0^{(k_j)}(\mathbf{x}) = \lim_{t \rightarrow \infty} p_t(\mathbf{x}) = \lim_{t \rightarrow \infty} \frac{\tilde{p}_t(\mathbf{x})}{\int_{\Omega_{k_j}} \tilde{p}_t(\mathbf{y}) d\mathbf{y}}. \quad (16)$$

The rate of attraction is exactly the difference between the first and second eigenvalues of \mathcal{L}^* in Ω_{k_j} ; this yields the exponential factor in equations (14) and (15). After some technical calculations ensuring control over the growth of the expansion coefficients of $\tilde{p}_t(\mathbf{x})$ in terms of the eigenfunctions $\phi_n^{(k_j)}(\mathbf{x})$, using the Gronwall's inequality, we arrive at formula (14). Formula (15) may be obtained similarly by estimating the error in distribution of exit locations $\pi_{t_j}(\mathbf{y})$.

Note also that the eigenfunction expansion method provides convergence of the density in L^2 which does not guarantee its positivity. However, this is not essential if the probability density function is employed to compute expected values of various quantities, e.g., the moments. Moreover, in our experiments, we found that including a sufficient number of terms in the approximation achieves positivity whenever $t - t_j$ is not too small.

D. Derivation of the Result

Since time $t = t_j$ is the hitting time of the last boundary, Γ_{k_j} , by the Markov property we have that for all $t > t_j$,

$$p_t(\mathbf{x}) = \int_{\mathbf{y} \in \Gamma_{k_j}} p_t^{(k_j)}(\mathbf{x}|\mathbf{y}) \pi_{t_j}(\mathbf{y}) d\mathbf{y}, \quad (17)$$

where $\pi_{t_j}(\mathbf{y})$ is the distribution of exit locations from $\Omega_{k_{j-1}}$ and $p_t^{(k_j)}(\mathbf{x}|\mathbf{y})$ is the transition probability density for \mathbf{X}_t given that $\mathbf{X}_{t_j} = \mathbf{y}$ and $\mathbf{X}_t \in \Omega_{k_j}$ for all $t > t_j$. (Before the first hitting time, i.e., for $j = 0$, $p_t(\mathbf{x}) = \int_{\mathbf{y} \in \Omega_0} p_t^{(0)}(\mathbf{x}|\mathbf{y}) p_0(\mathbf{y}) d\mathbf{y}$.) The transition density is the normalized solution of the Fokker-Planck equation,

$$\partial_t \tilde{p}_t^{(k_j)}(\mathbf{x}|\mathbf{y}) = \mathcal{L}^* \tilde{p}_t^{(k_j)}(\mathbf{x}|\mathbf{y}), \quad (18)$$

with the boundary condition $\tilde{p}_t^{(k_j)}(\mathbf{x}|\mathbf{y}) = 0$ for $\mathbf{x} \in \partial\Omega_{k_j}$ and initial data $\tilde{p}_{t_j}^{(k_j)}(\mathbf{x}|\mathbf{y}) = \delta(\mathbf{x} - \mathbf{y})$ (tilde is used to emphasize that the solution must be normalized to produce the actual transition probability density). It may be represented using the standard methods, see e.g., [12], as

$$\tilde{p}_t^{(k_j)}(\mathbf{x}|\mathbf{y}) = \sum_{n=0}^{\infty} e^{\lambda_n^{(k_j)}(t-t_j)} \phi_n^{(k_j)}(\mathbf{x}) \varphi_n^{(k_j)}(\mathbf{y}). \quad (19)$$

Substituting this expression into equation (17) and introducing

$$\tilde{p}_n^{(j)} = \int_{\mathbf{y} \in \Gamma_{k_j}} \varphi_n^{(k_j)}(\mathbf{y}) \pi_{t_j}(\mathbf{y}) d\mathbf{y}, \quad (20)$$

we recover the asserted formula (11). Further on, by the same argument for $t \in (t_{j-1}, t_j)$, $j \geq 1$,

$$\tilde{p}_t(\mathbf{x}) = \sum_{n=0}^{\infty} e^{\lambda_n^{(k_{j-1})}(t-t_{j-1})} \phi_n^{(k_{j-1})}(\mathbf{x}) \tilde{p}_n^{(j-1)}. \quad (21)$$

Since the (unnormalized) distribution of exit locations from $\Omega_{k_{j-1}}$ is given by $\pi_{t_j}(\mathbf{y}) = \mathcal{F} \lim_{t \uparrow t_j} \tilde{p}_t(\mathbf{y})$, we obtain

$$\pi_{t_j}(\mathbf{y}) = \sum_{n=0}^{\infty} e^{\lambda_n^{(k_{j-1})}(t_j-t_{j-1})} \mathcal{F} \phi_n^{(k_{j-1})}(\mathbf{y}) \cdot \mathbf{n} \tilde{p}_n^{(j-1)}. \quad (22)$$

Using this expression in formula (20), we recover relations (12) and (13).

III. EXPLICIT EXAMPLE AND NUMERICAL SIMULATIONS

Consider a system described by the following stochastic differential equations:

$$\begin{cases} dX_t^{(1)} = dW_t^{(1)} & \text{(state)} \\ dX_t^{(2)} = -\kappa X_t^{(2)} dt + \sigma dW_t^{(2)} & \text{(measurement noise),} \end{cases} \quad (23)$$

where $W_t^{(1)}$ and $W_t^{(2)}$ are independent one-dimensional Wiener processes, $\kappa > 0$, and $\sigma > 0$ are parameters. Using Eq. (8), we obtain that generator of \mathbf{X}_t is given by

$$\mathcal{L} = -\kappa x_2 \partial_{x_2} + \frac{1}{2} (\partial_{x_1 x_1}^2 + \sigma^2 \partial_{x_2 x_2}^2). \quad (24)$$

The Fokker-Planck operator \mathcal{L}^* is the (formal) adjoint of the generator \mathcal{L} , and can be obtained using Eq. (9) as

$$\mathcal{L}^* = \kappa \partial_{x_2} x_2 + \frac{1}{2} (\partial_{x_1 x_1}^2 + \sigma^2 \partial_{x_2 x_2}^2). \quad (25)$$

Note that since \mathcal{L}^* is translation-invariant with respect to the x_1 variable, the eigenvalue problems in all domains Ω_k are equivalent, and the eigenfunctions may be produced by simple translation; we therefore use Ω_0 for explicit calculations. The eigenvalues and eigenfunctions of \mathcal{L}^* together with their orthogonal counterparts, $\varphi_{vw}(\mathbf{x})$, and the transfer coefficients are computed in Appendix A and are given by

$$\begin{aligned} \lambda_{vw} &\approx -\frac{\pi^2 v^2}{8\gamma^2} - \kappa w \\ \phi_{vw}(\mathbf{x}) &\approx \frac{\sqrt{\kappa\pi}}{4\gamma} e^{-\kappa y_2^2} H_w(\sqrt{\kappa}y_2) \sin \frac{\pi v(x_1 + x_2 + \gamma)}{2\gamma} \\ \varphi_{vw}(\mathbf{x}) &\approx \frac{4}{\pi 2^w w!} H_w(\sqrt{\kappa}y_2) \sin \frac{\pi v(x_1 + x_2 + \gamma)}{2\gamma} \\ T_{(v,w);(v',w')}^{(k_{j-1},k_j)} &= 0 \text{ if } v = 2u, \text{ or } w < w' \\ T_{(v,w);(v',w')}^{(k_{j-1},k_j)} &\approx -\frac{(-1)^{u+v'} \pi v' ((k_{j-1} - k_j) \sigma \gamma \sqrt{\kappa})^{w-w'}}{4\gamma^2 (w-w')!} \\ &\quad \text{if } v = 2u + 1, w \geq w', = 1 \end{aligned} \quad (26)$$

Here $y_2 = \frac{\sigma x_1 - x_2 / \sigma}{\sqrt{1 + \sigma^2}}$ and the double index notation ($v = 1, 2, \dots$, $w = 0, 1, \dots$) is convenient since the eigenfunctions are computed using the separation of variables. The symbol \approx reflects the fact that in order to obtain analytical expressions we used perturbation theory considering σ a small parameter. In general, this is not necessary and all quantities in (26) may be precomputed numerically, which is a necessity for more complicated diffusion processes.

Now we evaluate the accuracy, convergence and efficiency of our density estimator. As time increases, so does the number of hitting times. Since at each hitting time t_j , when a transition from the domain Ω_{t_j} to the domain $\Omega_{t_{j+1}}$ occurs, the truncation approximation has to be reapplied, we need to ensure that the truncation error is under control. We next provide evidence via simulations that the proposed methodology can be applied over a large time horizon to approximate the actual conditional state probability density function with a small error. The diffusion parameters used in our numerical tests are $\kappa = 0.5$, $\sigma = 0.1$, $\gamma = 2$.

We first compare the probability density conditional on the observation process Y_t staying within the first domain Ω_0 computed using our methodology with the probability density estimated from Monte-Carlo runs. The first density is computed using equations (11) and (12) with eigenfunctions and eigenvalues given by the analytical expressions in Eq. (26). We choose the initial condition corresponding to the initial distribution to be $p_0(\mathbf{x}) = \delta(\mathbf{x})$. This choice is representative for the purpose of error estimation, since in this particular case the eigenfunction expansion produces large coefficients for the higher harmonics and thus the truncation results in greater errors. The Monte-Carlo density is obtained simulating the Brownian paths via the forward Euler scheme,

$$X_{t+\Delta t}^{(1)} = X_t^{(1)} + \sqrt{\Delta t} Z, \quad (27)$$

with Z being a standard Gaussian. We simulate the Ornstein-Uhlenbeck process $X_t^{(2)}$ as [13],

$$X_{t+\Delta t}^{(2)} = X_t^{(2)} \mu + \tilde{\sigma} \xi, \quad (28)$$

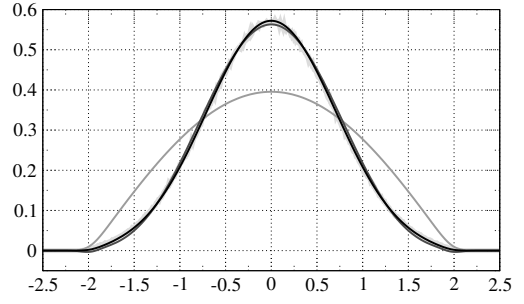


Fig. 2. Probability density function of the state variable $X_t^{(1)}$ at $t = 0.5$ with the initial condition $\mathbf{X}_t = 0$ obtained using 1, 10, and 45 eigenfunctions (from grey to black lines respectively), and the Monte-Carlo simulation (light grey line).

where ξ is a standard Gaussian independent of Z and

$$\mu = e^{-\kappa \Delta t}, \quad \tilde{\sigma}^2 = \frac{\sigma^2}{2\kappa} (1 - \mu^2). \quad (29)$$

We set the time discretization $\Delta t = 0.01$. The results presented in Figure 2 show that a number of 10 eigenfunctions already provide a very accurate approximation of the actual density, with a perfect match reached when using 45 eigenfunctions, thereby indicating that the approach is accurate and efficient in a single domain.

We next analyze the number of eigenfunctions needed to approximate the actual filter density exactly at the hitting time, and at a time t shortly after the hitting time t_j , where $t - t_j = 0.1$. We set the coefficient $\tilde{p}_n^{(0)} = 1$ if $n = 0$ and 0 otherwise, which corresponds to hitting the boundary from the stationary distribution, $p_\infty(\mathbf{x}) = \phi_{10}(\mathbf{x})$. The density right at the hitting time is a singular distribution on the boundary and thus requires the greatest number of eigenfunctions to be approximated. We sort the eigenvalues in the decreasing order and define n_k to be the rank of the eigenvalue λ_{k0} in the sorting. In our particular case, the corresponding eigenfunctions happen to contribute the most into the expansion, so it is sensible to approximate the true density using all eigenfunction up to order n_k . We have that $n_1 = 1$, $n_3 = 10$, $n_5 = 45$, $n_7 = 129$, $n_9 = 279$, and $n_{11} = 516$. We would like to remark that among the n_k eigenfunctions used, only few of them contribute noticeably to the reconstruction of our density estimator (the coefficients of the transfer matrix corresponding to the other eigenfunctions are either very small or zero), thus in a practical application of the method most of them could be removed without altering the quality of the density reconstruction.

The results in Figure 3 show that even though approximations converge rather slowly at the hitting time (the positivity is also an issue due to Gibbs phenomenon when approximating a singular distribution), even shortly after the hitting time, the convergence improves dramatically. This is due to the damping of the higher-order contributions by the exponential factors $e^{\lambda_n^{(k_j)}(t-t_j)}$ in formula (11). Figure 3 also evidences a very good match between the Monte-Carlo density computed shortly after the hitting time, and our eigenfunction expansion density estimator. The systematic shift on the left of our density estimator with respect to the Monte-Carlo density is due to

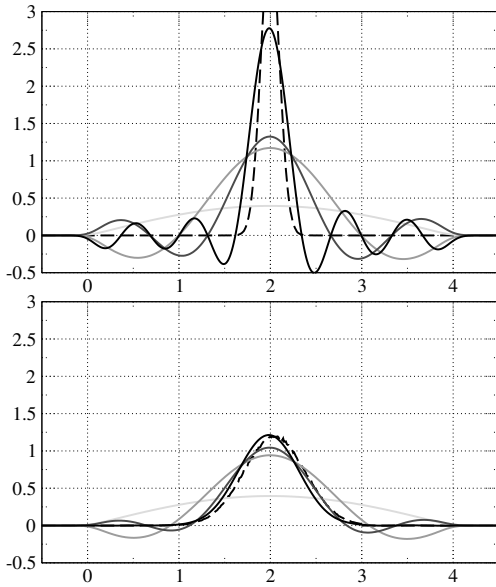


Fig. 3. The left plot shows the probability density function of the state variable $X_t^{(1)}$ computed using n_k eigenfunctions, $k = \{1, 3, 5, 11\}$ (from light to dark color) at the time when Γ_1 is hit. The right plot shows the probability density function on the state variable computed using the same eigenfunctions at the time t such that $t - t_1 = 0.1$, with t_1 being the time when Γ_1 is hit. In both plots, the dashed line represents the Monte-Carlo density.

the small σ approximation used to obtain the eigenfunctions. Clearly, the match is not perfect on the boundary, because the density becomes singular and cannot be approximated using an eigenfunction expansion in L^2 .

We finally demonstrate that the method converges fast to the actual state filter density over a large time horizon. We simulate state and noise diffusion paths and record all the conditioning information to compute the state filter density, which consists of pairs of the form (index of hit boundary, hitting time). We pick a realization where the difference between the final simulation time, $t = 40$, and the last hitting time is small, so that the filter density has not yet converged to the stationary density and the higher order approximations are needed. The simulated conditioning path is given by

$$P = \{(1, 4.78), (2, 7.43), (1, 14.28), (2, 19.14), (1, 20.93), (0, 21.9), (-1, 23.46), (0, 26.01), (-1, 28.1), (-2, 30.82), (-1, 35.93), (-2, 39.79)\} \quad (30)$$

We compute the density at time $t = 40$ on the path specified in Eq. (30) using the n_k eigenfunction approximation as above. As the eigenfunction expansion method is guaranteed to converge in L^2 , and therefore also our filter methodology converges in L^2 , we compute the L^2 distance between n_k and n_{k-2} order approximations of our density estimator. It appears from the L^2 distance plots in the right panel of Fig.4 that the difference between the n_9 and n_{11} order approximations is negligible (of the order of 10^{-4}). The grey line in the L^2 distance plot fits the L^2 distance with an exponential curve ($y = 0.8e^{-0.6x}$), thus demonstrating that the convergence is exponentially fast with respect to v .

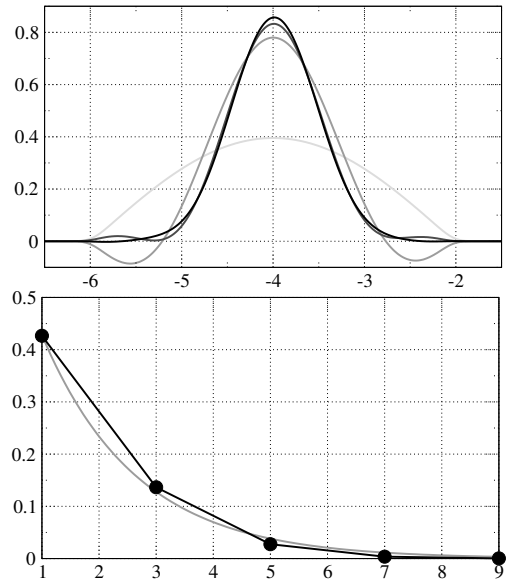


Fig. 4. The left plot is the state filter density at $t = 40$ (different shades correspond to $k = 1, 3, 5, 7$, from light to dark). The right plot is the L^2 distance between n_k and n_{k+2} order approximations of the density, with $k = 1, 5, 7, 9$. The smooth curve is an exponential fitting.

Finally notice that even at the order of $n_{11} = 516$, the matrix-vector multiplications needed for iterating our scheme require roughly 266,000 operations at each hitting time, which is computationally fast (moreover, e.g., for our particular processes, the transfer matrix is sparse and thus only a few of the first 516 eigenfunctions actually contribute into the final result). Since all this is independent of the conditioning information set, then a pre-analysis of the problem can further reduce the computational burden of the methodology.

IV. CONCLUSION

We presented a general framework to solve a non-linear filtering problem, where the non-linearity comes from the conditioning information. The observations are generated at random times when the measurement process hits pre-specified thresholds. We illustrated the method on a particular example where the state and the noise are modeled by Wiener and Ornstein-Uhlenbeck processes respectively. Our numerical simulations demonstrate that reasonable accuracy is achieved even when only a few eigenfunctions are used for expansion of the probability density function of the system.

As the threshold γ gets smaller, our framework formally reproduces the continuous time filters, e.g., [2]: the signals are generated more often, until, in the limiting case, they are generated continuously and the conditioning observation set then consists of the whole process Y_t . However, in this case the practical use of our method is limited, as in order to keep the approximation errors small it is necessary to account for a large number of eigenfunctions.

REFERENCES

- [1] R. Kalman and R. Bucy, "New results in linear filtering and prediction theory," *Trans. ASME J. Basic Eng.*, vol. 83, pp. 95–108, 1961.

- [2] G. Kallianpur and C. Striebel, "Estimation of stochastic systems: Arbitrary system process with additive white noise observation errors," *Ann. Math. Statist.*, vol. 39, pp. 785–801, 1968.
- [3] —, "Stochastic differential equations occurring in the estimation of continuous parameter stochastic processes," *Teor. Veroyatnost. i Primenen.*, vol. 14, no. 4, pp. 597–622, 1969.
- [4] H. Kushner, "Dynamical equations for optimal nonlinear filtering," *J. Diff. Eq.*, vol. 3, pp. 179–190, 1967.
- [5] M. Zakai, "On the optimal filtering of diffusion processes," *Z. Wahrsch. Verw. Geb.*, vol. 11, pp. 230–243, 1969.
- [6] P. Frost and T. Kailath, "An innovations approach to least-squares estimation-part iii: Nonlinear estimation in white gaussian noise," *IEEE Trans. Autom. Control*, vol. 16, no. 3, pp. 217–226, 1971.
- [7] R. Jarrow, P. Protter, and D. Sezer, "Information reduction via level crossings in a credit risk model," *Finance Stoch.*, vol. 11, no. 2, pp. 195–212, 2007.
- [8] J. Cvitanic, R. Liptser, and B. Rozovski, "A filtering approach to tracking volatility from prices observed at random times," *Ann. Appl. Probab.*, vol. 16, no. 3, pp. 1633–1652, 2006.
- [9] S. Aihara, A. Baghi, and S. Saha, "On parameter estimation of stochastic volatility models from stock data using particle filter-application to aex index," *Int. J. Innov. Comput.*, vol. 5, no. 1, pp. 17–28, 2009.
- [10] F. Ding, L. Qiu, and T. Chen, "Reconstruction of continuous-time systems from their non-uniformly sampled discrete-time systems," *Automatica*, vol. 45, no. 2, pp. 324–332, 2009.
- [11] F. Ding, P. Liu, and G. Liu, "Auxiliary model based multi-innovation extended stochastic gradient parameter estimation with colored measurement noises," *Signal Process.*, vol. 89, no. 10, pp. 1883–1890, 2009.
- [12] G. Gardiner, *Handbook of stochastic methods for physics, chemistry and the natural sciences*, ser. Springer Series in Synergetics. Springer-Verlag Berlin Heidelberg, 2004, vol. 13.
- [13] D. Gillespie, "Exact numerical simulation of the ornstein-uhlenbeck process and its integral," *Phys. Rev. E*, vol. 54, no. 22, pp. 2084–2091, 1996.

APPENDIX A EIGENFUNCTIONS AND EIGENVALUES OF THE FOKKER-PLANCK OPERATOR

First, let us find all eigenvalues and eigenfunctions (vanishing on the boundary) for the operator

$$\begin{aligned}\mathcal{L}^* &= \mathcal{L}_1^* + \mathcal{L}_2^*, \\ \mathcal{L}_1^* &:= \frac{1}{2}\partial_{x_1x_1}^2 + \kappa\partial_{x_1}x_1, \\ \mathcal{L}_2^* &:= \frac{1}{2}\partial_{x_2x_2}^2\end{aligned}\quad (31)$$

in the domain $\Omega = \mathbb{R} \times [0, L]$. In order to find the eigenvalues ν_n and the eigenfunctions $\xi_n(x_1)$ of \mathcal{L}_1^* , we observe that the greatest (smallest by the absolute value) eigenvalue ν_0 and the corresponding eigenfunction $\xi_n(x_1)$ are given by

$$\nu_0 = 0, \quad \xi_0(x_1) = \sqrt{\frac{\kappa}{\pi}}e^{-\kappa x_1^2}, \quad (32)$$

Notice that the eigenfunction ξ_0 is normalized to have total integral unity. Setting $\xi_n(x_1) = \xi_0(x_1)h_n(x_1)$ we obtain equations for $h_n(x_1)$:

$$\partial_{x_1x_1}^2 h_n(x_1) - 2\kappa x_1 \partial_{x_1} h_n(x_1) = 2\nu_n h_n(x_1). \quad (33)$$

This immediately implies that $h_n(x_1)$ may be expressed using Hermite polynomials as

$$h_n(x_1) = H_n(\sqrt{\kappa}x_1), \quad (34)$$

thus we obtain

$$\begin{aligned}\nu_n &= -\kappa n, \\ \xi_n(x_1) &= \sqrt{\frac{\kappa}{\pi}}e^{-\kappa x_1^2} H_n(\sqrt{\kappa}x_1), \quad n = 0, 1, 2, \dots\end{aligned}\quad (35)$$

The operator \mathcal{L}_2^* has eigenvalues $\mu_m = -\pi^2 m^2 / 2L^2$ and eigenvectors

$$\psi_m(x_2) = \frac{\pi}{2L} \sin \frac{\pi m x_2}{L}, \quad m = 1, 2, 3, \dots \quad (36)$$

Since

$$\begin{aligned}\mathcal{L}^*[\xi_n(x_1)\psi_m(x_2)] &= (\mathcal{L}_1^* + \mathcal{L}_2^*)[\xi_n(x_1)\psi_m(x_2)] \\ &= (\nu_n + \mu_m)\xi(x_1)\psi_m(x_2),\end{aligned}\quad (37)$$

\mathcal{L}^* has the following eigenvalues and eigenvectors:

$$\begin{cases} \lambda_{mn} &= -\frac{\pi^2 m^2}{2L^2} - \kappa n, \\ \phi_{mn}(x) &= \frac{\sqrt{\kappa\pi}}{2L} e^{-\kappa x_1^2} H_n(\sqrt{\kappa}x_1) \sin \frac{\pi m x_2}{L}, \end{cases}\quad (38)$$

where $m = 1, 2, 3, \dots$, $n = 0, 1, 2, \dots$. Finally, the functions $\varphi_{mn}(\mathbf{x})$ such that $\{\phi_{mn}(\mathbf{x}), \varphi_{mn}(\mathbf{x})\}$ form a bi-orthogonal set are given by

$$\varphi_{mn} = \frac{4}{\pi 2^n n!} H_n(\sqrt{\kappa}x_1) \sin \frac{\pi m x_2}{L}. \quad (39)$$

For the discussion in Section III, we need to find the eigenvalues and eigenfunctions (with vanishing boundary values) of the Fokker-Planck operator

$$\mathcal{L}^* = \kappa\partial_{x_2x_2} + \frac{1}{2}(\partial_{x_1x_1}^2 + \sigma^2\partial_{x_2x_2}^2) \quad (40)$$

in the domain $\Omega = \{(x_1, x_2) : |x_1 + x_2| < \gamma\}$. Introducing

$$y_1 = \frac{x_1 + x_2}{\sqrt{1 + \sigma^2}}, \quad y_2 = \frac{\sigma x_1 - \frac{x_2}{\sigma}}{\sqrt{1 + \sigma^2}}, \quad (41)$$

we obtain

$$\begin{aligned}\mathcal{L}^* &= \frac{1}{2}[\partial_{y_1y_1}^2 + \partial_{y_2y_2}^2] + \\ &\quad \frac{\kappa}{1 + \sigma^2}[\partial_{y_2y_2} - \sigma(y_1\partial_{y_2} + y_2\partial_{y_1}) + \sigma^2\partial_{y_2y_2}],\end{aligned}$$

while the domain transforms into

$$\Omega' = \left\{ (y_1, y_2) : |y_1| < \frac{\gamma}{\sqrt{1 + \sigma^2}} \right\}. \quad (42)$$

For sufficiently small σ we may use the regular perturbation theory and results in (38) and (39) reproducing relations (26). The transfer matrix $T_{(m,n):(m',n')}^{(k_j-1,k_j)}$ may be computed explicitly using the approximate eigenfunctions obtained from the small σ approximation and formula (13).