Nonlinear Filtering with Transfer Operator

Parikshit Dutta, Abhishek Halder, and Raktim Bhattacharya

***Abstract*— This paper presents a new nonlinear filtering algorithm that is shown to outperform state-of-the-art particle filters with resampling. Starting from the Itô stochastic differential equation, the proposed algorithm harnesses Karhunen-Loéve expansion to derive an approximate non-autonomous dynamical system, for which transfer operator based density computation can be performed in exact arithmetic. It is proved that the algorithm is asymptotically consistent in mean-square sense. Numerical results demonstrate that explicitly accounting prior dynamics entail significant performance improvement for nonlinear non-Gaussian estimation problems with infrequent measurement updates, as compared to the performance of particle filters.**

I. Introduction

On a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ with filtration $\{\mathscr{F}_t\}_{t\geqslant 0}$, consider the nonlinear estimation problem associated with the Itô stochastic differential equations (SDEs)

$$dx\,(t) \;=\; f\,(x\,(t)\,,t,\delta)\;dt + d\mathcal{W}\,(\omega,t)\,, \tag{1}$$

$$dy\,(t) \;=\; h\,(x\,(t)\,,t,\delta)\;dt + d\mathcal{V}\,(\omega,t)\,, \tag{2}$$

where at time instance t, the state vector $x\,(t) \in \mathbb{R}^n$, and the measurement vector $y\,(t) \in \mathbb{R}^m$. $\delta \in \mathbb{R}^p$ is the parameter vector, and $\mathcal{W}\,(\omega,t) : \Omega \times \mathbb{R}^+ \mapsto \mathbb{R}^n$, $\mathcal{V}\,(\omega,t) : \Omega \times \mathbb{R}^+ \mapsto \mathbb{R}^m$ are mutually independent Wiener processes denoting process and measurement noise, respectively. Further, $\omega \in \Omega$, and the functions $f\,(.)$ and $h\,(.)$ represent the dynamics and measurement models, respectively.

State and parameter estimation for nonlinear systems such as above, are commonly done using *sequential Monte Carlo* (SMC) methods, *particle filter* being the most popular amongst them [1]. These algorithms follow traditional prediction-update framework where the *prior* is predicted using state dynamics, followed by a Bayesian update using measurement model, resulting the *posterior*. It is well known [2] that these methods require large number of samples for convergence, leading to higher computational cost. This problem is usually tackled by combining particle filters with *resampling* [3], [4], commonly known as *bootstrap filters* [5]. However, resampling may introduce loss of diversity amongst particles [6]. Several other methods like *regularized particle filter* [7], and *filters with Markov Chain Monte Carlo (MCMC) move step* [8], have been proposed to enhance sample diversity. At the same time, even with resampling, due to the simulation based nature of these filters, the sample size scales exponentially with state dimension [9]. To circumvent this problem, particle filters based on *Rao-Blackwellization* [10] have been proposed to partially solve the estimation problem analytically. However, its application remains limited to systems where the required partition of the state space is possible.

The main idea of this paper is to recognize the fact that much of the computational burden of particle filter, stems from the Monte Carlo approximation of the prior. Lack of statistically consistent methods for high dimensional uncertainty propagation, has stymied the accurate computation of prior density. In the previous work, the authors developed [11] *Perron-Frobenius* (PF) operator [12] based methods for numerically efficient uncertainty propagation schemes for nonlinear systems with *parametric and initial condition uncertainties*. This was achieved by solving the characteristic ordinary differential equation (ODE) corresponding to the Liouville partial differential equation (PDE), along the trajectories in the extended state space $z := [x \quad \delta]^\top$. In the estimation setting, it was observed [13] that prior probability density function (PDF) computed via PF operator, followed by Bayesian update, outperformed particle filter *in the absence of process noise*.

In the presence of process noise, the transport PDE associated with *forward Kolmogorov operator* is the *Fokker-Planck-Kolmogorov* (FPK) equation [14], which being a second order PDE, does not enjoy method-of-characteristics (MOC) based ODE formulation. Function approximation techniques for solving FPK eqn. usually suffer from the "curse of dimensionality" [15]. This severely limits the accuracy of prior computation, and hence that of the nonlinear filters. An alternative approach was proposed recently by the authors [16], where instead of directly approximating the prior, the process noise was approximated by a finite-term Karhunen-Loéve (KL) expansion resulting an approximate state dynamics. Next, the MOC based PF operator computation was performed on this approximate non-autonomous dynamical system in exact arithmetic. [16] provided strong numerical evidence that such two step "*first KL, then PF*" (henceforth *KLPF*) algorithm is *asymptotically consistent in distribution*. However, two issues remained unsettled.

1) In [16], the distributional consistency was algorithmically verified. A rigorous proof for convergence was lacking. Also, it was not clear whether the distributional convergence is only *sufficient*, i.e. whether a stronger notion of convergence holds true.
2) No detailed numerical investigation was performed to assess the filtering performance improvement resulting from KLPF algorithm, vis-a-vis with particle filter.

Parikshit Dutta is with INRIA Rhône Alpes & Laboratoire Jean Kuntzmann, 655 avenue de l'Europe, 38330 Montbonnot, France, parikshit.dutta@inria.fr; Abhishek Halder and Raktim Bhattacharya are with the Department of Aerospace Engineering, Texas A&M University, College Station, TX 77843, USA, {ahalder,raktim}@tamu.edu

This paper has two key contributions. ***First***, we prove that solution of the KL approximated dynamics, converges to that of the true Itô SDE in *mean-square (m.s.) sense*. This is indeed stronger than the distributional convergence of [16]. Further, the m.s. convergence is shown to be necessary and sufficient. ***Second***, we provide strong numerical results showing that the proposed algorithm achieves superior estimation accuracy than the particle filter with resampling. The proposed nonlinear estimation algorithm, henceforth referred as *KLPF filter*, is shown to achieve better performance with much lesser samples than particle filter.

The rest of this paper is structured as follows. Section II describes the KLPF formulation for computing prior PDF and provides m.s. convergence guarantees. The nonlinear filtering algorithm is introduced in Section III. Section IV contains numerical results for both linear Gaussian and nonlinear non-Gaussian estimation problems. Section V concludes the paper.

Notation

Most notations are standard. $\mathcal{I}_n$ denotes the n-by-n identity matrix, and $\text{diag}\,(\cdot)$ denotes the diagonal matrix. The symbol $\mathcal{N}\,(\mu, \Sigma)$ denotes joint Gaussian PDF with mean μ and covariance Σ. $\mathbb{N}$ stands for the set of natural numbers, $\text{tr}\,(\cdot)$ denotes the trace of a matrix, and $\text{div}(\cdot)$ denotes the divergence operator. The symbol δ_{ij} represents Kronecker delta.

II. Approximating Prior Dynamics

A. KLPF Formulation

Given the Itô SDE (1), we write an approximate dynamical system corresponding to its Langevin ODE form for the j^{th} state:

$$\dot{x}_N^{(j)} = f^{(j)}\left(x_N(t), t, \delta\right) + \sum_{i=1}^{N} \sqrt{\Lambda_i}\, \zeta_i^{(j)}\left(\omega\right) e_i\left(t\right), \qquad (3)$$

where $j = 1, 2, \ldots, n$. Further, $\{\Lambda_i, e_i\,(t)\}_{i=1}^{\infty}$ is the sequence of eigenvalue-eigenfunction pairs of the covariance function $C\,(t_1, t_2)$ associated with the additive stationary process noise, and $\zeta_i\,(\omega)$ are i.i.d. random variables drawn from the distribution of the noise stochastic process. For example, if $\mathcal{W}\,(\omega, t)$ is *Wiener process* with $C\,(t_1, t_2) = \sigma^2\,(t_1 \wedge t_2)$, $t_1, t_2 \in [0, T]$, then $\{\Lambda_i, e_i\,(t)\}_{i=1}^{\infty}$ is the eigenpair sequence for *Gaussian white noise* $\eta\,(\omega, t)$, and $\zeta_i\,(\omega) \sim \mathcal{N}\left(0, \sigma^2\right)$. In effect, the second term in the RHS of (3), is the N-term KL expansion for $\eta\,(\omega, t)$. We affix subscript N to the states ($x_N(t)$) of the approximate dynamics (3), to distinguish them from the sample paths ($x(t)$) of the original SDE (1).

Next, we augment (3) with the characteristic ODE

$$\dot{\varphi^-} = -\text{div}\left(\widetilde{f}\right)\varphi^-, \qquad (4)$$

where $\widetilde{f}$ denotes the RHS nonlinearity of (3), and $\varphi^-\left(x_N(t), t, \delta\right)$ denotes the prior at time t, supported over the *extended state space* $\begin{bmatrix} x_N & \delta \end{bmatrix}^\top \in \mathbb{R}^{n+p}$. Consequently, (4) computes the evolution of joint prior PDF along the characteristic curves $x_N\,(t)$. Notice that we do not assume the process noise to be Gaussian. As long as the additive noise has finite second moment, we can write down the approximate dynamical system (3) via the noise KL expansion (Table I). The overall formulation is summarized in Fig. 1.

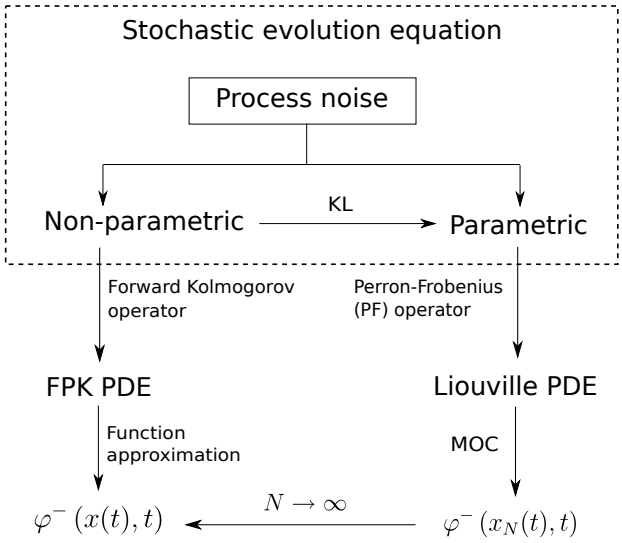


Fig. 1. Summary of the KLPF formulation.

It is well-known [19] that as $N \to \infty$, the finite-term noise KL expansion $\sum_{i=1}^{N} \sqrt{\Lambda_i}\, \zeta_i^{(j)}\left(\omega\right) e_i\left(t\right)$ converges uniformly to the unstructured noise $\eta\,(\omega, t)$ in *m.s. sense*. However, to justify our formulation, it remains to answer whether $x_N\,(t)$ converges to that of $x\,(t)$, and in what sense. The following sub-section answers this issue. For notational ease, we disregard uncertainty on parameter δ, without loss of generality. It is straightforward to verify that the following results generalize to extended state space.

B. Quality of Approximation

1) Asymptotic convergence: In [16], it was verified through simulation that as $N \to \infty$, $x_N(t) \to x(t)$ in distribution, i.e. $\varphi^-\left(x_N(t), t\right) \to \varphi^-\left(x(t), t\right)$, $\forall t \geqslant 0$. Here we prove the stronger result that $x_N(t) \to x(t)$ in m.s. sense.

Theorem 1: Let $x\,(\omega, t)$ be the solution of the nonlinear Itô SDE

$$dx(t) = f(x(t), t)dt + d\mathcal{W}(\omega, t), \qquad (5a)$$

$$\Rightarrow \frac{d}{dt}x\,(t) = f\,(x, t) + \eta(\omega, t), \qquad (5b)$$

where $dW(\omega, t) = \eta\,(\omega, t)\ dt$, and $f : \mathbb{R}^n \times [0, T] \to \mathbb{R}^n$ satisfies the following:

1) non-explosion condition: $\exists\ D \geq 0$, s.t. $|f\,(x, t)| < D(1 + |x|)$ where $x \in \mathbb{R}^n$, $t \in [0, T]$;
2) Lipschitz condition: $\exists\ C \geq 0$, s.t. $|f\,(x, t) - f\,(\breve{x}, t)| < C\,|x - \breve{x}|$, where $x, \breve{x} \in \mathbb{R}^n$, $t \in [0, T]$.

TABLE I

NOISE KL EXPANSION: EXAMPLES

Noise $\mathcal{W}(\omega,t)$ in SDE	$C\left(t_1,t_2\right)$ **for** $\mathcal{W}(\omega,t)$	**White noise $\eta(\omega,t)$ in Langevin ODE**	**KL expansion of $\eta\left(\omega,t\right)$, $0<t\leqslant T$**
Wiener process	$\sigma^2\left(t_1\wedge t_2\right)$	Gaussian white noise	$\sqrt{\frac{2}{T}}\sum_{i=1}^{\infty}\zeta_i\left(\omega\right)\cos\left(\left(i-\frac{1}{2}\right)\frac{\pi t}{T}\right)$
Compound Poisson process	$\lambda\sigma^2\left(t_1\wedge t_2\right)+(\lambda\mu)^2 t_1 t_2$	Poisson white noise (Appendix A)	$\sum_{i=1}^{\infty}\overline{\zeta}_i\left(\omega\right)\frac{\frac{2}{\beta_i}\sqrt{\Lambda_i}}{\sqrt{2T-\beta_i\sin\frac{2T}{\beta_i}}}\cos\left(\frac{t}{\beta_i}\right)$

Let $x_N\left(t\right)$ be solution of the ODE

$$\frac{d}{dt}x_N\left(t\right)=f\left(x_N(t),t\right)+\eta_N(\omega,t), \tag{6}$$

where $\eta_N(\omega,t)$ is the N-term truncated orthonormal expansion of $\eta(\omega,t)$, and $\mathbb{E}\left[\int_0^T\eta_N(\omega,t)dt\right]<\infty$. Then,

$$\lim_{N\to\infty}\mathbb{E}|x\left(t\right)-x_N\left(t\right)|^2=0, \tag{7}$$

iff $x_N\left(t\right)$ is the KL expansion of $x\left(t\right)$.

Theorem 1 states conditions upon the solutions of approximated and true systems for m.s. convergence to hold, under certain assumptions on the nonlinearities. No condition has been imposed yet on the initial states, which we investigate next.

Theorem 2: Given the stochastic dynamical system

$$dx(t)=\quad f(x(t),t)dt+d\mathcal{W}\left(\omega,t\right), \tag{8}$$

and its corresponding N-term KL approximation given by

$$dx_N^{(j)}(t)=\quad f^{(j)}(x_N(t),t)\,dt+\sum_{i=1}^{N}\sqrt{\Lambda_i}\zeta_i^{(j)}(\omega)\dot{e}_i(t)\,dt, \tag{9}$$

where, $\lim_{N\to\infty}\mathbb{E}\left|\mathcal{W}^{(j)}\left(\omega,t\right)-\sum_{i=1}^{N}\sqrt{\Lambda_i}\zeta_i^{(j)}(\omega)e_i(t)\right|^2=0$, $\forall\ j=1,2,\ldots,n$. Then, $\lim_{N\to\infty}\mathbb{E}|x(t)-x_N(t)|^2=0$, if $x(0)=x_N(0)$.

Corollary 3: Suppose $x_N(0)\neq x\left(0\right)$. If $x_N(0)$ is the *generalized polynomial chaos* (gPC) expansion of $x(0)$, then $\lim_{N\to\infty}\mathbb{E}|x(t)-x_N(t)|^2=0$.

III. KLPF FILTER

Algorithm 1 Continuous-discrete KLPF filter ('time of measurement' index: $k=1,\ldots,\tau$; sample index: $i=1,\ldots,\nu$)

Require: $\{y_k\}_{k=1}^{\tau}$ and φ_0 ▷ Measurements & initial joint state PDF
1: $\{x_{0,i}\}_{i=1}^{\nu}\leftarrow$ MCMC$(\{\varphi_{0,i}\}_{i=1}^{\nu})$ ▷ Initial sampling
2: **for** $k=0$ to $\tau-1$ **do**
3: $\{\varphi_{k+1,i}^{-},x_{k+1,i}^{-}\}_{i=1}^{\nu}\leftarrow$ Propagate$\{\varphi_{k,i}^{+},x_{k,i}^{+}\}_{i=1}^{\nu}$ ▷ MOC (4)
4: $\{\varrho_{(k+1|k+1),i}\}_{i=1}^{\nu}\leftarrow(2\pi)^{-\frac{m}{2}}|R|^{-\frac{1}{2}}\exp[-\frac{1}{2}(y_{k+1}-h(x_{k+1,i}^{-},t_k))^{\top}R^{-1}(y_{k+1}-h(x_{k+1,i}^{-},t_k))]$ ▷ Likelihood function
5: $\{\varphi_{k+1,i}^{+}\}_{i=1}^{\nu}\leftarrow$ Update$\{\varphi_{k+1,i}^{-},\varrho_{(k+1|k+1),i}\}_{i=1}^{\nu}$ ▷ Bayes'
6: $\hat{x}_{k+1}\leftarrow\sum_{i=1}^{\nu}x_{k+1,i}^{-}\,\varphi_{k+1,i}^{+}$ ▷ State estimate at $k+1^{\text{th}}$ time
7: **end for** ▷ Repeat for next measurement

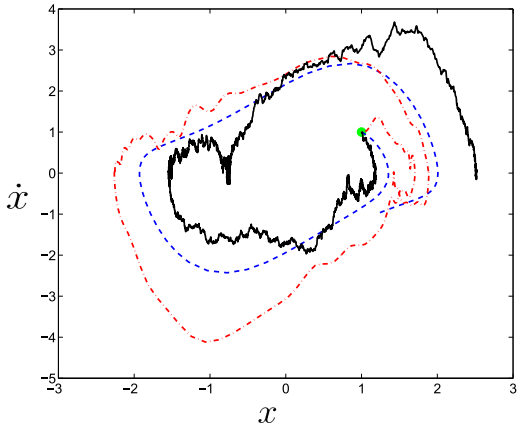


Fig. 2. This plot illustrates the asymptotic convergence results developed in Section II.B.1 for Van der Pol oscillator $\ddot{x}(t)=\left(1-x^2(t)\right)\dot{x}(t)-x(t)+\eta(\omega,t)$, with $\eta(\omega,t)$ being Gaussian white noise. Starting from the same initial condition $(1,1)$, denoted by the *filled circle*, the *dashed*, and *solid* curves show the deterministic (zero noise), and stochastic (SDE sample path with zero-mean additive Gaussian noise of variance 0.25) trajectories, respectively. The *dash-dotted* curve is the trajectory of the KL-approximated system with $N=100$ terms, starting from the same initial condition, with process noise same as that of the SDE path. As N increases, the *dash-dotted* curve converges to the *solid curve* in mean-square sense.

IV. NUMERICAL RESULTS

A. Comparison of KLPF and Particle Filter Posteriors

In this subsection, we consider two examples for which the estimation problem is exactly solvable and hence the true posterior is known. To demonstrate the performance improvement achieved by KLPF compared to particle filter, we must show that the KLPF posterior is closer to the true posterior, than particle filter. In other words, the "distance" between KLPF posterior and true posterior, must remain smaller than the "distance" between particle filter posterior and true posterior, *for all times*. The notion of distributional distance used here, is the *quadratic Wasserstein metric of order two* (denoted as ${}_2W_2$), that measures the *difference in shapes* between the two statistical distributions under comparison.

Definition 1: (**Wasserstein distance**) Consider a metric space (M,ℓ_p) and let $x,\widetilde{x}\in M$. For $q\in\mathbb{N}$, let $\mathcal{P}_q\left(M\right)$ denote the collection of all probability measures μ supported on M, which have finite q^{th} moment. Then the ℓ_p Wasserstein distance of order q, denoted as ${}_pW_q$, between two probability

measures $\varsigma_1, \varsigma_2 \in \mathcal{P}_q(M)$, is defined as

$$ {}_pW_q(\varsigma_1,\varsigma_2) := \left(\inf_{\mu\in\mathcal{M}(\varsigma_1,\varsigma_2)} \int_{M\times M} \| x-\widetilde{x} \|_{\ell_p}^q \; d\mu(x,\widetilde{x}) \right)^{\frac{1}{q}} $$

where $\mathcal{M}(\varsigma_1,\varsigma_2)$ is the set of all measures supported on the product space $M \times M$, with first marginal ς_1 and second marginal ς_2.

Remark 1: Intuitively, Wasserstein distance quantifies the minimum amount of work required to convert one distributional shape to the other, and can be interpreted as the cost for Monge-Kantorovich optimal transportation plan [20]. We set $p = q = 2$ (see [21] for details) for comparing posteriors, and for notational ease, henceforth denote ${}_2W_2$ as W. For absolutely continuous measures ς_1 and ς_2, with PDFs φ_1 and φ_2, we can write $W(\varphi_1,\varphi_2)$ in lieu of $W(\varsigma_1,\varsigma_2)$.

Remark 2: For multivariate Gaussians, W admits [22] a closed form expression, given by

$$ W(\mathcal{N}(\mu_1,\Sigma_1),\mathcal{N}(\mu_2,\Sigma_2)) = \Big(\| \mu_1-\mu_2 \|_2^2 + \mathrm{tr}(\Sigma_1+\Sigma_2) - 2\,\mathrm{tr}\left[\sqrt{\Sigma_1}\Sigma_2\sqrt{\Sigma_1}\right]^{1/2} \Big)^{1/2}. \tag{10} $$

In general, computation of W from definition 1, necessitates solving a linear program (LP). We refer the readers to [23] for details of this computation.

1) Case I. Kalman filter: Let us consider the continuous-discrete Kalman filter with continuous-time state dynamics

$$ \dot{x}(t) = -0.05\,\mathcal{I}_2\,x(t) + [1 \quad 1]^\top \eta(t), \tag{11} $$

and discrete-time measurement model

$$ y_k = [1 \quad 1]\; x_k + v_k, \qquad k \in \mathbb{N}, \tag{12} $$

where $\eta(t)$ and v_k are independent zero mean Gaussian white noise processes, with variances $Q = 1/8$ and $R = 1/4$, respectively. We assume the initial joint state PDF to be $\mathcal{N}\Big([1 \quad 1]^\top, \mathrm{diag}(1,1)\Big)$.

From this initial state PDF, we draw 100 sample sets, each with sample size 500. Then using (10), we compute two Wasserstein time histories: $W\left(\varphi^+_{\text{Kalman}}(t), \varphi^+_{\text{Particle}}(t)\right)$ and $W\left(\varphi^+_{\text{Kalman}}(t), \varphi^+_{\text{KLPF}}(t)\right)$, where $\varphi^+_{\text{Kalman}}(t)$, $\varphi^+_{\text{Particle}}(t)$ and $\varphi^+_{\text{KLPF}}(t)$ denote posteriors at time t, obtained from Kalman filter, particle filter and KLPF filter, respectively. The means and standard deviations of these time histories are shown in Fig. 3. This plot shows that the KLPF filter posterior remains indeed closer to the Kalman posterior, compared to the particle filter posterior.

2) Case II. Beneš filter: Beneš filter is one of the few [24] nonlinear filters which admit a *known finite-dimensional solution* of the nonlinear estimation problem. Here, the nonlinear drift in state dynamics, is assumed to satisfy a Riccati differential equation [25] and the measurement model is taken to be affine in states. We consider the continuous-continuous scalar Beneš filtering problem of the form:

$$ dx(t) = \frac{\kappa e^x - e^{-x}}{\kappa e^x + e^{-x}}\,dt + d\mathcal{W}(\omega,t), \tag{13} $$

$$ dy(t) = x(t)\;dt + d\mathcal{V}(\omega,t), \tag{14} $$

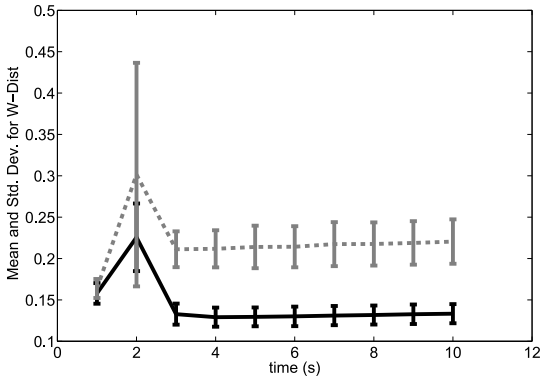


Fig. 3. Plot of means and standard deviations of the Wasserstein distances of the posteriors from KLPF filter (*solid line*) and the particle filter (*hyphenated line*) for the Kalman filter. The vertical lines about the means represent $\pm 1\sigma$ limits.

with $\kappa = 0.5$ and deterministic initial condition x_0. The process and measurement noise densities are $\mathcal{N}(0,Q)$ and $\mathcal{N}(0,R)$ respectively, with $Q = 1$, $R = 10$. It can be shown [26] that the drift nonlinearity satisfies the necessary Riccati condition and the resulting solution [27] is given by the normalized posterior density

$$ \varphi(x(t)\,|\mathcal{Y}_t) = \sqrt{\frac{\coth(t)}{2\pi}} \left(\frac{\kappa e^x + e^{-x}}{\kappa e^{I_t(y(\omega))} + e^{-I_t(y(\omega))}} \right) \exp\left(-\tfrac{1}{2}\Gamma(t)\right), \tag{15} $$

where $\mathcal{Y}_t$ is the history (filtration) till time t, and

$$ I_t(y(\omega)) := \mathrm{sech}(t)\left[x_0 + \int_0^t \sinh(s) dy_s(\omega) \right], \tag{16} $$

$$ \Gamma(t) := \tanh(t) + \coth(t)\left(x - I_t(y(\omega))\right)^2 \tag{17} $$

Notice that for this nonlinear non-Gaussian estimation problem, unlike Kalman filter case, we can not write the Wasserstein distance between the true posterior (15) and particle filter/KLPF posterior, as an analytical expression in terms of the respective sufficient statistics. Thus, in order to compute the Wasserstein time history, we resort to the LP formulation [23]. At each time, we sample (15) using the *Metropolis-Hastings MCMC* technique [28], and solve the LP between the sampled true Beneš posterior and particle filter/KLPF posterior, to result the normalized Wasserstein trajectories shown in Fig. 4. Like the Kalman filter case, as time progresses, KLPF posterior gets closer, compared to particle filter, to true Beneš posterior.

B. Application to Hypersonic Entry

The KLPF filtering technique is applied next to estimate states of a hypersonic spacecraft entering the atmosphere of Mars. The entry dynamics is given by the "noisy" version of

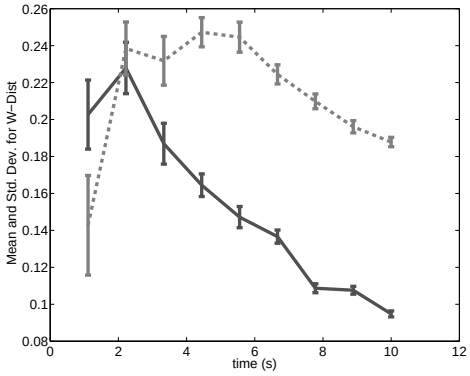


Fig. 4. Plot of means and standard deviations of the Wasserstein distances of the posteriors from KLPF filter (*solid line*) and the particle filter (*hyphenated line*) for the Beneš filter. The vertical lines about the means represent $\pm 1\sigma$ limits.

TABLE II

PARAMETERS IN EQN. (18)

Description of parameters	Values with dimensions
Mean equatorial radius of Mars	$R_m = 3397 \times 10^3$ m
Acceleration due to gravity for Mars	$g = 3.71$ m/s^2
Ballistic coefficient of the vehicle	$B_c = 72.8$ kg/m^2
Lift-to-drag ratio of the vehicle	$\frac{L}{D} = 0.3$
Nominal density at the surface of Mars	$\rho_0 = 0.0019$ kg/m^3
Scale heights for density computation	$h_1 = 9.8$ km, $h_2 = 20$ km
Escape velocity of Mars	$v_c = 5.027$ km/s
Rotational angular velocity of Mars	$\overline{\Omega} = 7.0882 \times 10^{-5}$ rad/s
Bank angle of the vehicle	$\phi = 0^o$

Vinh's equation [29]:

$$\dot{r} = v \sin\gamma + \eta_r, \tag{18a}$$

$$\dot{\theta} = \frac{v \cos\gamma \sin\xi}{r} + \eta_\theta, \tag{18b}$$

$$\dot{\lambda} = \frac{v \cos\gamma \cos\xi}{r \cos\theta} + \eta_\lambda, \tag{18c}$$

$$\dot{v} = -\frac{\rho v^2}{2B_c} - g \sin\gamma - \overline{\Omega}^2 r \cos\theta(\sin\gamma\cos\theta - \cos\gamma\sin\theta\sin\xi) + \eta_v, \tag{18d}$$

$$\dot{\gamma} = \left(\frac{v}{r} - \frac{g}{v}\right)\cos\gamma + \frac{\rho}{2B_c}\left(\frac{L}{D}\right) v \cos\phi + 2\overline{\Omega}\cos\theta\cos\xi + \frac{\overline{\Omega}^2 r}{v}\cos\theta(\cos\gamma\cos\theta + \sin\gamma\sin\theta\sin\xi) + \eta_\gamma, \tag{18e}$$

$$\dot{\xi} = \frac{\rho}{2B_c}\left(\frac{L}{D}\right) v \sin\phi - \frac{v}{r}\cos\gamma\cos\xi\tan\theta + 2\overline{\Omega}(\tan\gamma\cos\theta\sin\xi - \sin\theta) - \frac{\overline{\Omega}^2 r}{v\cos\gamma}\sin\theta\cos\theta\cos\xi + \eta_\xi. \tag{18f}$$

The 6×1 state vector comprises of the distance (in Km) of the center-of-mass of the spacecraft from Mars center (r), Mars-centric latitude (θ) and longitude (λ), total velocity (v) (in Km/s), flight path angle (γ) and azimuth angle (ξ). The variation in Mars atmospheric density (ρ in Kg/m^3) is given by the model $\rho = \rho_0 \exp\left((h_2 - h)/h_1\right)$, where $h := (r - R_m)$, denotes the altitude (in Km) measured from the mean Martian surface. The parameters in *state eqn.* (18) are described in Table II with their respective dimensions. The *measurement model* is given by

$$y = [\widetilde{q}, H, \gamma, \theta, \lambda, \xi]^\top + \vartheta, \tag{19}$$

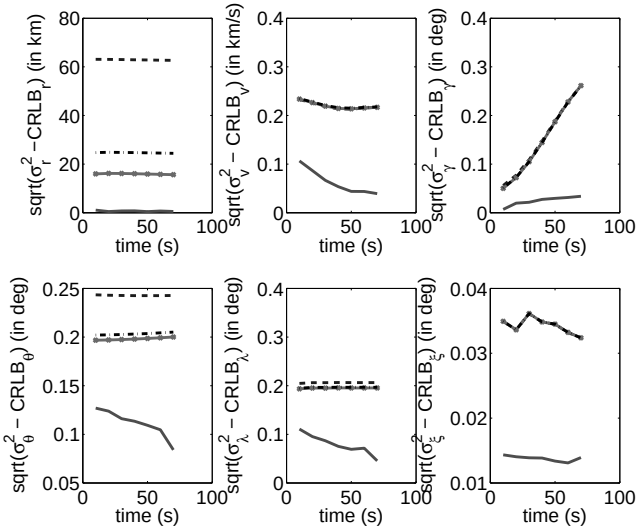


Fig. 5. Plots for $\sqrt{\sigma_{x_i}^2 - CRLB_{x_i}}$ for states $x_1, \ldots, x_6$. The *solid* line represents KLPF filter (3000 particles). The *hyphenated*, *hyphen-dotted* and *solid-asterixed* lines represent particle filters with 3000, 20,000 and 50,000 particles, respectively.

where the dynamic pressure $\widetilde{q} = \frac{1}{2}\rho v^2$, and the heating rate $H = K\rho^{\frac{1}{2}} v^{3.15}$, with $K = 4.47228 \times 10^{-9}$ being the scaled material heating coefficient [30]. Each component of the process noise vector $\eta := (\eta_r, \eta_\theta, \eta_\lambda, \eta_v, \eta_\gamma, \eta_\xi)^\top$ is a zero mean mutually uncorrelated Gaussian white noise, with appropriate units. Same holds true for the 6×1 measurement noise vector ϑ, which is also uncorrelated with η. The process and measurement noise covariances are taken as $Q = 3.6 \times 10^{-5}\mathcal{I}_6$ and $R = 3.6 \times 10^{-3}\mathcal{I}_6$, respectively.

Starting with an initial state PDF $\mathcal{N}(\mu_0, \Sigma_0)$, with $\mu_0 = [R_m + 54 \text{ Km}, -60^o, 30^o, 2.4 \text{ Km/s}, -9^o, 0.0573^o]^\top$ and $\Sigma_0 = \text{diag}\,(5.4 \text{ Km}, 3^o, 3^o, 240 \text{ m/s}, 0.9^o, 0.0057^o)$, both particle filter and KLPF filtering schemes are applied to estimate the state vector $x = [r, \theta, \lambda, v, \gamma, \xi]^\top$. Fig. 5 shows the plots for square root of the difference between the respective variance (σ_{x_i}) and *Cramer-Rao lower bound* (CRLB$_{x_i}$) for each state x_i. The performance of KLPF filter with sample size 3000, is compared with the same for particle filters with 3000, 20,000 and 50,000 particles. It can be observed that $\sqrt{\sigma_{x_i}^2 - \text{CRLB}_{x_i}}$ for KLPF filter is lower than that of the particle filters for all the states. This demonstrates that the solution obtained from proposed estimation scheme remains closer to the true minimum variance solution than that obtained from particle filters.

Next, we plot the final posterior univariate and bivariate marginals obtained from the two filtering methods, computed using the algorithm given in [11]. Fig. 6 and 7 respectively compare the univariate and bivariate marginals, obtained from KLPF estimator and particle filter, both with 3000 samples. It can be observed that the KLPF estimator is able to reduce variance and capture localization of uncertainty better than the particle filter, with same number of samples. This remains true even when the performance of KLPF estimator with 3000 samples, is compared with a particle filter with

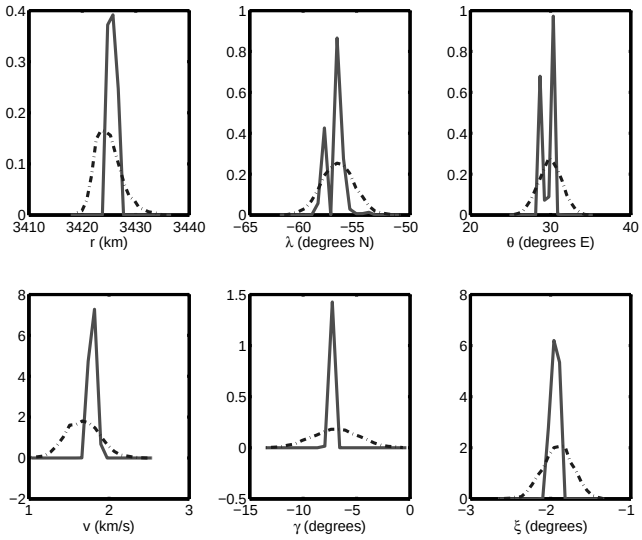


Fig. 6. Comparison of final posterior univariate marginal PDFs for all states, obtained from KLPF estimator (*solid line*) and particle filter (*hyphenated line*) with 3000 particles.

50,000 samples (Fig. 8 and 9).

V. Conclusions

A new nonlinear filtering algorithm is presented in this paper, that is shown to outperform the estimation accuracy of particle filters with even higher number of samples. This is achieved by explicitly taking the prior dynamics into account. Contrary to the traditional "top-down" approach of numerically solving the FPK PDE via function approximation, a "bottom up" approach for prior computation is developed by *first approximating the problem* via spectral parametrization of the noise, and *then solving that approximate problem in exact arithmetic* via MOC computation of the transfer operator. The resulting algorithm, dubbed as KLPF filter, is a *non-particle filter* [32], and is amenable to both Gaussian and non-Gaussian process noise. The estimation performance improvement over particle filter, is demonstrated through numerical simulations.

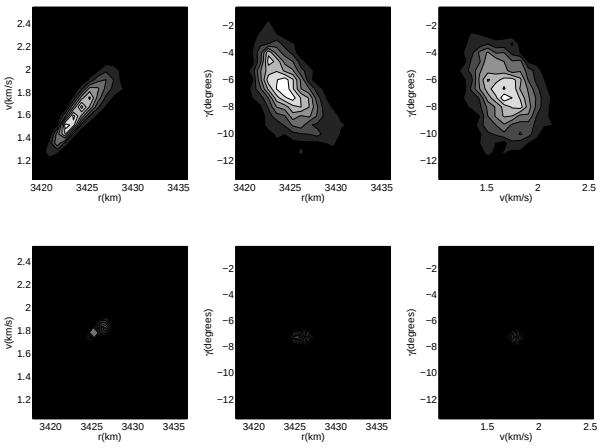
Appendix

A. KL Expansion of Poisson White Noise

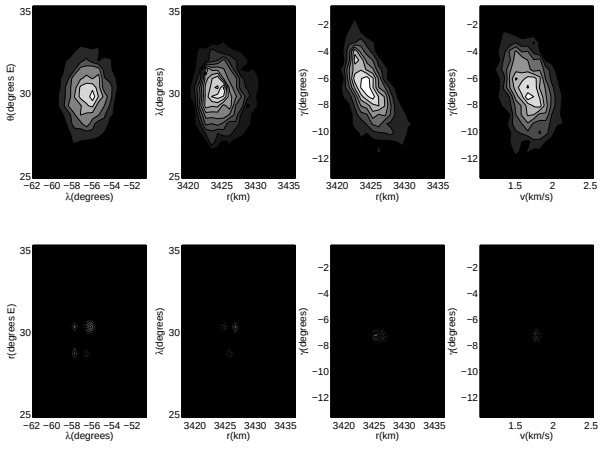
We first obtain the KL expansion of compound Poisson process [17] with covariance kernel given in Table I. This requires us to solve the associated Fredholm integral equation of second kind, detailed in the supplementary document [18]. Next, we take the formal derivative (in m.s. sense) of the KL expansion of compound Poisson process, to arrive at the KL expansion of Poisson white noise, given in the second row, right-most column in Table I. Here, $\overline{\zeta}_i(\omega)$ are i.i.d random variables from $\mathcal{N}(0,1)$, $\beta_i \triangleq \sqrt{\frac{\Lambda_i}{\lambda\sigma^2}}$, $\forall\, i \in \mathbb{N}$, and $\Lambda_i > 0$ solves

$$\tan\left(\sigma T\sqrt{\frac{\lambda}{\Lambda_i}}\right) = \left[1 + \frac{1}{\lambda T}\left(\frac{\sigma}{\mu}\right)^2\right]\left(\sigma T\sqrt{\frac{\lambda}{\Lambda_i}}\right), \qquad (20)$$

where the parameters $\lambda, \sigma, \mu, T > 0$.



(a) Top row- particle filter, bottom row- KLPF estimator.



(b) Top row- particle filter, bottom row- KLPF estimator.

Fig. 7. Plots for the final posterior bivariate marginal PDFs obtained from KLPF estimator and particle filter with 3000 particles. The *darker (lighter)* regions represent *lower (higher)* PDF values.

B. Proof for Theorem 1

($\Leftarrow$) Given (7) holds, we need to show $x_N(\omega,t)$ is the KL expansion of $x(\omega,t)$. Let $\{\psi_m(t)\}_{m=1}^{\infty}$ be any orthonormal basis. Then $x(\omega,t)$ can be written as a convergent sum in $L_2(\Omega,\mathscr{F},\mathbb{P})$, i.e. $x(\omega,t) = \sum_{m=1}^{\infty} b_m c_m(\omega)\psi_m(t)$.

Let $x_N(\omega,t)$ be an N-term m.s. convergent approximation of $x(\omega,t)$, and the resulting truncation error equals $\mathscr{E}_N(\omega,t) = \sum_{m=N+1}^{\infty} b_m c_m(\omega)\psi_m(t)$. Further, projecting $x(\omega,t)$ onto the basis $\psi_m(t)$ results $c_m(\omega) = \frac{1}{b_m}\int_0^T x(\omega,t)\psi_m(t)dt$. For convergence, the basis $\psi_m(t)$ should minimize $\int_0^T \mathbb{E}\left[\mathscr{E}_N(\omega,t)\right]dt$ subject to the orthonormality constraint $\int_0^T \phi_m(t)\phi_k(t)dt = \delta_{mk},\ \forall m,k \in \mathbb{N}$.

Introducing b_m^2 as Lagrange multipliers and using the above derived formula for $c_m(\omega)$, the first order optimality

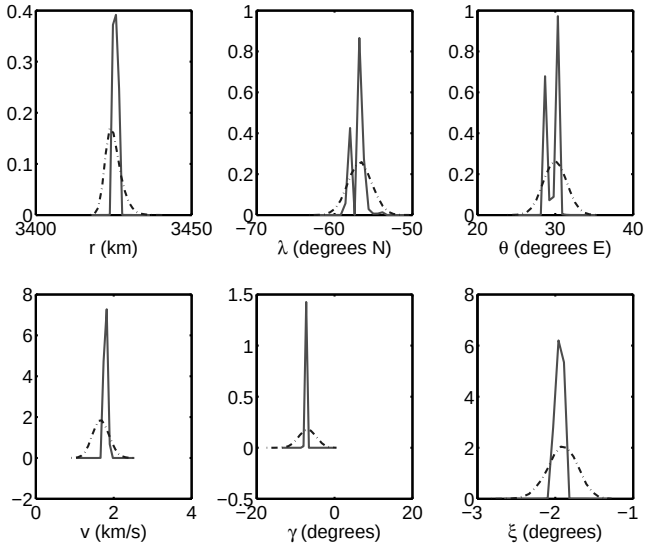


Fig. 8. Final posterior univariate marginal PDF for all states obtained from KLPF estimator (*solid line*) with 3000 particles and particle filter (*hyphenated line*) with 50,000 particles.

condition yields $\int_0^T C_{xx}(t_1,t_2)\,\psi_m(t_1)\,dt_1 = b_m^2\psi_m(t_2)$, which is the Fredholm integral equation of second kind for the covariance function of random process $x(\omega,t)$. Hence $\{b_m^2,\psi_m(t)\}_{m=1}^{\infty}$ is the eigenvalue-eigenfunction sequence for $C_{xx}(t_1,t_2)$. Thus, the original expansion is indeed a KL expansion. ■

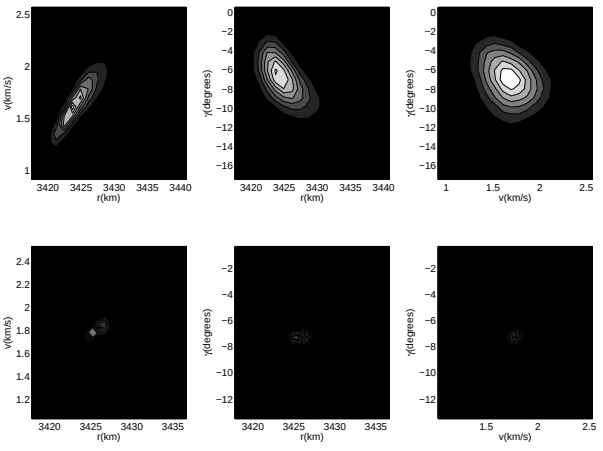
($\Rightarrow$) To proceed, we need the following uniqueness conditions on (i) solution of (5a), and (ii) KL expansion of a random process.

Proposition 1 ([33], Chap. 5): Given, the non-explosion condition and the Lipschitz condition are satisfied for $f(\cdot,\cdot)$ in (5a). Let Z be a random variable, independent of the σ-algebra generated by $\eta(\omega,t), t\geq 0$, and $\mathbb{E}\left[|Z|^2\right]<\infty$. Then the SDE (5a) where $t\in[0,T], X(\omega,0)=Z$, has a unique t-continuous solution $x(\omega,t)$ adapted to the filtration $\mathscr{F}_t^Z$ generated by Z, and $\mathbb{E}\left[\int_0^T |x(\omega,t)|^2 dt\right]<\infty$.

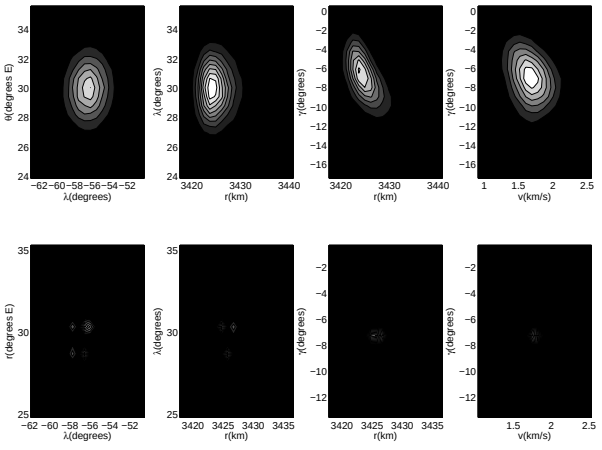
Proposition 2 ([31], Chap. 2): The Karhunen-Loève expansion of a random process $x(\omega,t)$, given by $x(\omega,t)=\sum_{i=1}^{\infty}\sqrt{\Lambda_i}\zeta_i(\omega)e_i(t)$, is unique.

Let us assume that $\breve{x}_N(\omega,t)$ is the KL expansion of $x(\omega,t)$. Furthermore, if possible, assume that $\breve{x}_N(\omega,t)\neq x_N(\omega,t)$, which is the solution of (6) and converges to the solution of (5a) in m.s. sense.

Notice that (6) has unique solution as RHS of (6) satisfies Lipschitz condition. This can be proved as follows: for RHS of (6) to satisfy Lipschitz condition, we must have $|f(x,t)+\eta_N(\omega,t)-f(\breve{x},t)-\eta_N(\omega,t)|\leq C|x-\breve{x}|$, which is true since $f(\cdot,\cdot)$ itself satisfies Lipschitz condition. Hence (5a) has unique solution that admits a unique KL expansion. Also according to our assumption, the solution of (6) converges to the solution of (5a) in m.s. sense. This contradicts our assumption that $\breve{x}_N(\omega,t)\neq x_N(\omega,t)$, which completes the proof. ■



(a) Top row- particle filter, bottom row- KLPF estimator.



(b) Top row- particle filter, bottom row- KLPF estimator.

Fig. 9. Plots for the final posterior bivariate marginal PDFs obtained from KLPF estimator with 3000 particles and particle filter with 50,000 particles. The *darker* (*lighter*) regions represent *lower* (*higher*) PDF values.

C. Proof for Theorem 2

Integrating (8) and (9) and taking the expected value of square of the difference, we obtain

$$\mathbb{E}|x(t)-x_N(t)|^2 = \mathbb{E}\left[\left|(x(0)-x_N(0))+\int_0^t (f(x,s)-f(x_N,s))\,ds+\int_0^t d(\mathcal{W}_s-\sum_{i=1}^{N}\sqrt{\Lambda_i}\zeta_i(\omega)e_i(s))\right|^2\right],$$

$$\leq \underbrace{\mathbb{E}|(x(0)-x_N(0))|^2}_{0=:B\text{ (say)}}+\mathbb{E}\left|\int_0^t (f(x,s)-f(x_N,s))ds\right|^2+$$

$$\mathbb{E}\left|\int_0^t d(W_s-\sum_{i=1}^{N}\sqrt{\Lambda_i}\zeta_i(\omega)e_i(s))\right|^2,$$

$$\leq B+t\mathbb{E}\int_0^t |f(x,s)-f(x_N,s)|^2\,ds+$$

$$\mathbb{E}\left|\int_0^t d(W_s-\sum_{i=1}^{N}\sqrt{\Lambda_i}\zeta_i(\omega)e_i(s))\right|^2, \qquad (21)$$

where in the last step, we used Chebyshev's integral inequality. Consequently, we have

$$\lim_{N\to\infty}\mathbb{E}|x(t)-x_N(t)|^2 \leq B+\lim_{N\to\infty} t\mathbb{E}\left[\int_0^t |f(x,s)-f(x_N,s)|^2 ds\right]+\lim_{N\to\infty}\mathbb{E}\left|\int_0^t d(W_s-\sum_{i=1}^{N}\sqrt{\Lambda_i}\zeta_i(\omega)e_i(s))\right|^2. \quad (22)$$

Using the Lipschitz criterion and property of KL expansion, from (22) we get

$$\underbrace{\lim_{N\to\infty}\mathbb{E}|x(t)-x_N(t)|^2}_{v(t)\text{ (say)}} \leq B+tC\int_0^t \lim_{N\to\infty}\mathbb{E}\,|x(s)-x_N(s)|^2\,ds,$$

$$\Rightarrow v(t)\leq B+A\int_0^t v(s)ds \Rightarrow v(t)\leq B\exp(At), \quad (23)$$

where the last step follows from Gronwall's inequality, with $tC\leq A, \forall t\in(0,T]$. Therefore, $\lim_{N\to\infty}\mathbb{E}|x(t)-x_N(t)|^2 = 0$, since $x(0)=x_N(0)\Rightarrow B=0$, as per our assumption. ■

D. Proof for Corollary 3

In the proof of Theorem 2, for $x(0)\neq x_N(0)$, taking the limit $N\to\infty$ yields

$$\lim_{N\to\infty}\mathbb{E}|x(t)-x_N(t)|^2 \leq \lim_{N\to\infty}\mathbb{E}|(x(0)-x_N(0))|^2+\lim_{N\to\infty} t\mathbb{E}\int_0^t |f(x,s)-f(x_N,s)|^2\,ds+\lim_{N\to\infty}\mathbb{E}\left|\int_0^t d(\mathcal{W}_s-\sum_{i=1}^{N}\sqrt{\Lambda_i}\zeta_i(\omega)e_i(s))\right|^2.$$

Going through the subsequent steps as before, we arrive at

$$\lim_{N\to\infty}\mathbb{E}|x(t)-x_N(t)|^2=0, \text{ if } \lim_{N\to\infty}\mathbb{E}|x(0)-x_N(0)|^2=0.$$

However, if $x_N(0)$ is the gPC expansion of $x(0)$, then they asymptotically converge in m.s. sense [31]. Hence $\lim_{N\to\infty}\mathbb{E}|x(0)-x_N(0)|^2=0$, which, from the Gronwall's inequality, implies that $\lim_{N\to\infty}\mathbb{E}|x(t)-x_N(t)|^2=0$. This completes our proof. ■

ACKNOWLEDGEMENT

This research was partially supported by NSF award #1016299 with D. Helen Gill as the Program Manager.

REFERENCES

[1] M.S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A Tutorial On Particle Filters for Online Nonlinear/Non-Gaussian Bayesian Tracking", *IEEE Transactions on Signal Processing*, Vol. 50, No. 2, pp. 174–188, 2002.

[2] F. Daum, and J. Huang, "Curse of Dimensionality and Particle Filters", *Proceedings of the 2003 IEEE Aerospace Conference*, Vol. 4, pp. 1979–1993, 2003.

[3] A. Doucet, S. Godsill, and C. Andrieu, "On Sequential Monte Carlo Sampling Methods for Bayesian Filtering", *Statistics and Computing*, Vol. 10, No. 3, pp. 197–208, 2000.

[4] C.S. Manohar, and D. Roy, "Monte Carlo Filters for Identification of Nonlinear Structural Dynamical Systems", *Sadhana*, Vol. 31, No. 4, pp. 399–427, 2006.

[5] N.J. Gordon, D.J. Salmond, and A.F.M. Smith, "Novel Approach to Nonlinear/Non-Gaussian Bayesian State Estimation", *IEE Proceedings of Radar and Signal Processing*, Vol. 140, No. 2, pp. 107–113, 1993.

[6] B. Ristic, S. Arulampalam, and N. Gordon, *Beyond the Kalman Filter: Particle Filters for Tracking Applications*, Artech House Publishers, 2004.

[7] N. Oudjane, and C. Musso, "Progressive Correction for Regularized Particle Filters", *Proceedings of the Third International Conference on Information Fusion*, Vol. 2, pp. THB2–10, 2000.

[8] W.R. Gilks, and C. Berzuini, "Following A Moving Target–Monte Carlo Inference for Dynamic Bayesian Models", *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, Vol. 63, No. 1, pp. 127–146, 2001.

[9] C. Snyder, T. Bengtsson, P. Bickel, and J. Anderson, "Obstacles to High-dimensional Particle filtering", *Monthly Weather Review*, Vol. 136, No. 12, pp. 4629–4640, 2008.

[10] G. Casella, and C.P. Robert, "Rao-Blackwellisation of Sampling Schemes", *Biometrika*, Vol. 83, No. 1, pp. 81–94, 1996.

[11] A. Halder, and R. Bhattacharya, "Dispersion Analysis in Hypersonic Flight During Planetary Entry Using Stochastic Liouville Equation", *Journal of Guidance Control and Dynamics*, Vol. 34, No. 2, pp. 459–474, 2011.

[12] A. Lasota, and M.C. Mackey, *Chaos, Fractals, and Noise: Stochastic Aspects of Dynamics*, Springer, Vol. 97, 1994.

[13] P. Dutta, and R. Bhattacharya, "Hypersonic State Estimation Using the Frobenius-Perron Operator", *Journal of Guidance Control and Dynamics*, Vol. 34, No. 2, pp. 325–344, 2011.

[14] H. Risken, *The Fokker-Planck Equation: Methods of Solution and Applications*, Springer, Vol. 18, 1996.

[15] R. Bellman, *Dynamic Programming*, Princeton University Press, 1957.

[16] P. Dutta, A. Halder, and R. Bhattacharya, "Uncertainty Quantification for Stochastic Nonlinear Systems using Perron-Frobenius Operator and Karhunen-Loéve Expansion", *IEEE Multi-Conference on Systems and Control*, Dubrovnik, Croatia, 2012. Available at `http://people.tamu.edu/~ahalder/MSC2012_Dobruvnik.html`

[17] M. Grigoriu, "White Noise Processes", *Journal of Engineering Mechanics*, Vol. 113, No. 5, pp. 757–765, 1987.

[18] P. Dutta, A. Halder, and R. Bhattacharya, *Supplementary Material*, 2012. Available at `http://people.tamu.edu/~ahalder/ACC2013_KLPFfilter.html`

[19] M. Grigoriu, *Stochastic Calculus: Applications in Science and Engineering*. First ed., Birkhäuser; 2002.

[20] C. Villani, *Topics in Optimal Transportation*. Graduate Studies in Mathematics, First ed., American Mathematical Society; 2003.

[21] A. Halder, and R. Bhattacharya, "Further Results on Probabilistic Model Validation in Wasserstein Metric", *51st IEEE Conference on Decision and Control*, Maui, 2012.

[22] C.R. Givens, and R.M. Shortt, "A Class of Wasserstein Metrics for Probability Distributions", *The Michigan Mathematical Journal*, Vol. 31, No. 2, pp. 231–240, 1984.

[23] A. Halder, and R. Bhattacharya, "Model Validation: A Probabilistic Formulation", *50th IEEE Conference on Decision and Control*, Orlando, 2011.

[24] F. Daum, "Exact Finite-dimensional Nonlinear Filters", *IEEE Transactions on Automatic Control*, Vol. 31, No. 7, pp. 616–622, 1986.

[25] V.E. Beneš, "Exact Finite-dimensional Filters for Certain Diffusions with Nonlinear Drift", *Stochastics: An International Journal of Probability and Stochastic Processes*, Vol. 5, No. 1-2, pp. 65–92, 1981.

[26] A. Bain, and D. Crişan, *Fundamentals of Stochastic Filtering*, Vol. 60, Springer Verlag, 2008.

[27] D. Crisan, "A Direct Computation of the Beneš Filter Conditional Density", *Stochastics: An International Journal of Probability and Stochastic Processes*, Vol. 55, No. 1-2, pp. 47–54, 1995.

[28] S. Chib, and E. Greenberg, "Understanding the Metropolis-Hastings Algorithm", *American Statistician*, Vol. 49, No. 4, pp. 327–335, 1995.

[29] J.S. Chern, and N.X. Vinh, "Optimum Reentry Trajectories of A Lifting Vehicle", *Michigan University Final Report*, Vol. 1, 1980.

[30] K. Bollino, I. Ross, and D. Doman, "Optimal Nonlinear Feedback Guidance for Reentry Vehicles", *AIAA Guidance, Navigation and Control Conference*, 2006.

[31] R.D. Ghanem, and P.D. Spanos, *Stochastic Finite Elements: A Spectral Approach*, Dover Publications, 2003.

[32] F. Daum, and M. Krichman, "Non-particle Filters", *Proceedings of SPIE*, Vol. 6236, 2006.

[33] B.K. Øksendal, *Stochastic Differential Equations: An Introduction with Applications*, Springer Verlag, 2003.