Asymptotically Exact Unweighted Particle Filter for Manifold-Valued Hidden States and Point Process Observations

Simone Carlo Surace, Anna Kutschireiter, and Jean-Pascal Pfister

***Abstract*—The filtering of a Markov diffusion process on a manifold from counting process observations leads to 'large' changes in the conditional distribution upon an observed event, corresponding to a multiplication of the density by the intensity function of the observation process. If that distribution is represented by unweighted samples or particles, they need to be jointly transformed such that they sample from the modified distribution. In previous work, this transformation has been approximated by a translation of all the particles by a common vector. However, such an operation is ill-defined on a manifold, and on a vector space, a constant gain can lead to a wrong estimate of the uncertainty over the hidden state. Here, taking inspiration from the feedback particle filter (FPF), we derive an asymptotically exact filter (called ppFPF) for point process observations, whose particles evolve according to intrinsic (i.e., parametrization-invariant) dynamics that are composed of the dynamics of the hidden state plus additional control terms. While not sharing the gain-times-error structure of the FPF, the optimal control terms are expressed as solutions to partial differential equations analogous to the weighted Poisson equation for the gain of the FPF. The proposed filter can therefore make use of existing approximation algorithms for solutions of weighted Poisson equations.**

***Index Terms*—Filtering, estimation, stochastic systems, mean field games, stochastic optimal control.**

I. Introduction

A LARGE number of natural and engineered systems and datasets have states that are naturally described as elements of smooth manifolds. Classical cases are the motion of a body constrained by equality constraints, motion on the surface of the earth, or the attitude of a rigid body. Increasingly, the systems are very high-dimensional, whereas data points often lie on relatively low-dimensional manifolds, whose structure can be exploited for filtering and estimation problems.

In filtering, the state of the system (called the *hidden* state) needs to be estimated from the history of observations. In practise, observations often arrive sparsely, randomly and in digital form. One example is when observations are simple event counts. Such counting or point process observations arise in a variety of applications of time series models, e.g., neuroscience, geosciences, or finance.

The exact solution of the filtering problem is intractable in most cases and requires numerical approximation. One approach has been the class of interacting particle algorithms, in which an unweighted ensemble of N particles is propagated based on the known dynamics of the hidden state and the incoming observations. The feedback particle filter (FPF) [1]–[2] is such an algorithm that is based on mean-field optimal control, with a gain$\times$error structure that is reminiscent of the Kalman filter. The gain is given by the solution of a partial differential equation (PDE), which makes the FPF exact in the limit of large N even for nonlinear problems. Although in practise the gain has to be estimated from the particles, unweighted approaches hold the promise of scaling to high-dimensional problems, in contrast to particle algorithms with importance weights [3].

In this letter, we consider the problem of finding an FPF-like algorithm for systems whose hidden states evolve continuously in time on a known smooth manifold and observations are given by a conditional Poisson process. The FPF for manifold-valued hidden states and diffusion observations has been introduced in [4]. A filter for a hidden state in $\mathbb{R}^n$ and point process observations was introduced in [5], called EKSPF. While it is reminiscent of the FPF, having a gain$\times$error structure, it uses a constant gain. As a result, the filter is exact only to first order and does not properly reflect higher-order statistics. For example, when particles are initially spread out and an incoming event confers evidence that the hidden state is in some narrow region of the state space, we should find the updated particles concentrated in that region. However, upon

Manuscript received July 23, 2019; revised September 30, 2019; accepted October 23, 2019. Date of publication November 4, 2019; date of current version December 19, 2019. This work was supported by the Swiss National Science Foundation under Grant PP00P3_179060. Recommended by Senior Editor J.-F. Zhang. *(Corresponding author: Simone Carlo Surace.)*

S. C. Surace is with the Department of Physiology, University of Bern, 3012 Bern, Switzerland (e-mail: surace@pyl.unibe.ch).

A. Kutschireiter is with the Department of Neurobiology, Harvard Medical School, Boston, MA 02115 USA, and also with the Department of Physiology, University of Bern, 3012 Bern, Switzerland.

J.-P. Pfister is with the Department of Physiology, University of Bern, 3012 Bern, Switzerland, and also with the Institute of Neuroinformatics, University of Zurich and ETH Zurich, 8057 Zürich, Switzerland.

Digital Object Identifier 10.1109/LCSYS.2019.2951093

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see http://creativecommons.org/licenses/by/4.0/

source: https://doi.org/10.7892/boris.144041 | downloaded: 25.4.2024

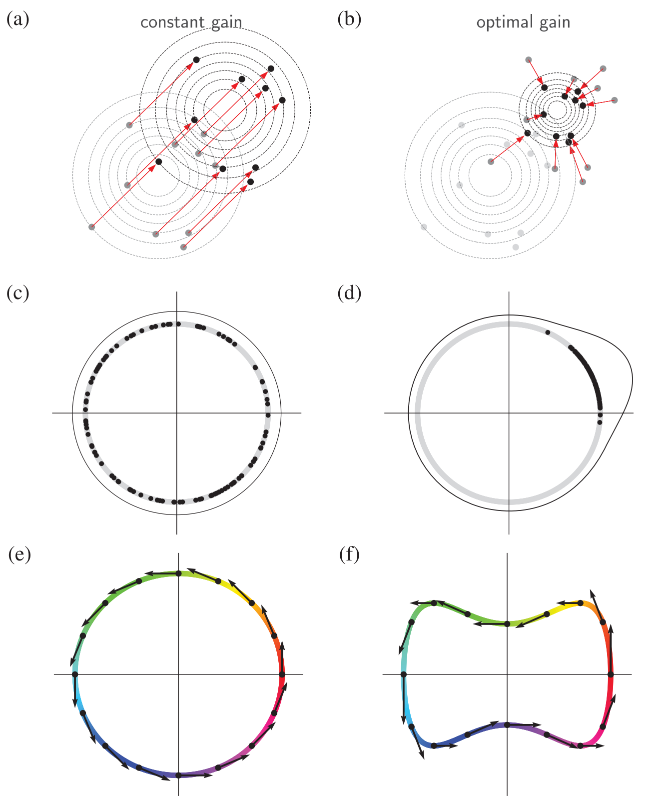

Fig. 1. (a) In Euclidean space, a constant gain for the update associated to an event leads to a translation of all the particles by a common translation vector. This leads to the correct mean but over-estimates the variance. (b) By contrast, the optimal gain also takes into account the reduction of uncertainty, and performs a scaling in addition to the translation. (c, d) Event-induced update on S^1. (c) Before the event, particles are uniformly distributed on the circle. (d) An event indicates that the hidden state is likely to be found in the upper right quadrant (black line). The optimal gain should transform the particle ensemble such that it becomes concentrated in the appropriate region after the event. This cannot be accomplished by a 'constant' update, which would simply lead to identical (i.e., independently of their position) rotation of all the particles, preserving the uniform distribution. (e, f) In addition, the notion of constancy on a manifold may depend on its embedding. (e) S^1 in its standard embedding in $\mathbb{R}^2$, with a tangent vector field of constant Euclidean length. (f) S^1 in a non-standard embedding. The same vector field from (e) is now of non-constant Euclidean length. An intrinsic notion of constancy is needed in order to avoid these ambiguities.

an event the EKSPF translates all particles by the same vector, see Figures 1a-1b.

The reliance on this uniform translation also leads to difficulties in extending the EKSPF to hidden states evolving on a manifold. In fact, when the EKSPF is applied naïvely on some arbitrary chart of the manifold, filtering performance can be poor (see Section IV for an example). This is because the meaning of a 'translation' is fundamentally ill-defined on a manifold. Since a translation in coordinate chart A does not necessarily correspond to a translation in coordinate chart B, the performance of the EKSPF depends on the choice of coordinates. However, the filtering problem on a manifold is intrinsic, i.e., independent of the choice of coordinates. It would therefore be desirable for a particle filter, and the transformation of particles in particular, to be defined in a coordinate-independent way. This would be advantageous even if the state space carries additional structure, such as the vector space structure on $\mathbb{R}^n$. A large class of estimation problems in $\mathbb{R}^n$, such as, e.g., satellite tracking, are naturally described in curvilinear coordinates.

For infinitesimal motion of particles, the notion of constancy of a vector field,[1] and thus of a constant gain approximation, depends on additional structure on the manifold, namely a connection; a mathematical structure that prescribes how to parallel transport a vector between different points. This can be visualised for the example of the unit circle S^1 that (regarded as a smooth manifold) can be embedded in different ways in, say, $\mathbb{R}^2$ (see Figures 1e-1f). If the constancy of a tangent vector field is made to depend on the embedding, then we obtain different vector fields for different embeddings. On many manifolds, there are no nontrivial parallel vector fields, which precludes the choice of a nontrivial constant gain. While this problem also affects a constant gain approximation of the FPF gain, the problem can be circumvented by seeking a non-constant gain estimate. Meanwhile, the constant gain assumption is 'baked' into the EKSPF.

In this letter, we derive an exact FPF-like filter on a manifold for point process observations, called ppFPF, from first principles, addressing the limitations of a constant gain in the EKSPF. The result is a filter whose control terms are given by solutions of PDEs analogous to the Poisson equation for the gain of the FPF. However, the gain$\times$error structure of the FPF is not strictly preserved. Instead, for the conceptual reasons stated above, the control term associated to an event is fundamentally distinct and treated separately from the term in-between events.

The remainder of this letter is structured as follows: in Section II, we introduce the mathematical notation, review the filtering problem for the Gaussian white noise observation case, and re-derive the FPF in the manifold setting, making some observations regarding the symmetry of the problem. In Section III, we present our main contribution: we derive the ppFPF, which is an adaptation of the FPF to point-process observations. In Section IV, we present numerical examples that illustrate the differences in performance and uncertainty quantification (UQ) between the ppFPF and other filters.

II. Preliminaries and Background

A. Notations and Conventions

Tangent vectors at a point $p \in M$ are written in a local chart as $a^i \partial_i|_p$, where Einstein's summation convention is used. A vector field $X \in \mathrm{Vect}(M)$ is a smooth section of the tangent bundle TM and is written locally as a first-order differential operator $X^i \partial_i$. The Lie derivative with respect to the vector field V is denoted by $\mathscr{L}_V$ and acts on sections of tensor product bundles of TM. If $\varphi \in C^\infty(M)$, then its differential $d\varphi$ is a one-form or smooth section of the cotangent bundle T^*M. More generally, a differential form of degree k is a smooth section of $\Omega^k(M) := \bigwedge^k T^*M$, where the wedge denotes the exterior product. Top degree forms are elements of $\Omega^n(M)$, where n is the dimension of M. A nowhere-vanishing element of $\Omega^n(M)$ is an orientation; if such an element exists then M

[1]As we will explain in the next section, the control terms in the FPF can be viewed as vector fields.

is called orientable, and we can then distinguish positive top degree forms, which we call volume forms. Normalized volume forms will be used to describe smooth nowhere-vanishing distributions on M. This letter d is used for exterior derivatives on differential forms $\omega \in \Omega^k(M)$ as $d\omega$, and for stochastic differentials on stochastic processes X_t as dX_t. The interior derivative on $\omega \in \Omega^k(M)$ wrt. $X \in \mathrm{Vect}(M)$ is written as $i_X\omega$. The notation $\mathscr{F}_t^Y$ is used for the filtration generated by the process $(Y_t)_{t\geq 0}$.

B. Filtering Problem and Filtering Equations

We consider a filtering problem in which the hidden state X_t evolves as a Markov diffusion process on an n-dimensional manifold[2] M, described by a Stratonovich stochastic differential equation (SDE) of the form

$$dX_t^i = V_0^i dt + V_j^i \circ dB_t^j \tag{1}$$

in local coordinates, where $B^1, \ldots, B^r$ are mutually independent standard Brownian motions.[3] We will use the index-free notation $dX_t = V_0 dt + V_j \circ dB_t^j$ for such an SDE on M. This SDE corresponds to an infinitesimal generator

$$\mathscr{A} = V_0 + \frac{1}{2}\sum_{j=1}^{r} V_j^2, \tag{2}$$

where $V_0, V_1, \ldots, V_r$ are vector fields on M. This is a second-order differential operator, which can be expressed in local coordinates as $\mathscr{A} = V_0^i\partial_i + \frac{1}{2}\sum_{k=1}^r V_k^i \partial_i V_k^j \partial_j$.

The classical observation model in nonlinear filtering is a diffusion process with additive noise, also referred to as observations in Gaussian white noise, i.e.,

$$dY_t = h(X_t)dt + dW_t, \tag{3}$$

where W_t is a Brownian motion independent of X_t. Although this letter is concerned with point process observations, in order to explain the background of this letter this section will focus exclusively on the model in Eq. (3). Later, in Section III, we shall consider point process observations, adapting an approach that has been used in the case of Gaussian white noise.

Probability distributions over the manifold M will be described by positive top-degree forms μ (volume forms) that integrate to one, i.e., $\int_M \mu = 1$. This convention avoids the superfluous appearance of a reference measure on M, and therefore emphasizes the metric-independent nature of the filtering problem. Of course, for concreteness, it is always possible to pick a reference volume form λ (for example, take the riemannian volume measure with respect to some riemannian metric on M, e.g., the Lebesgue measure for $M = \mathbb{R}^n$), and then to express μ in terms of a density p as $\mu = p\lambda$.

If the distribution of X_0 is described in terms of a volume form μ_0, the conditional distribution μ_t of X_t, given observations $\mathscr{F}_t^Y$, evolves according to the equation

$$d\mu_t = (\mathscr{A}^\dagger \mu_t)\, dt + \left(h - \hat{h}_t\right)\mu_t(dY_t - \hat{h}_t dt), \tag{4}$$

[2]To avoid further complications, we assume M to be connected and orientable.

[3]We use Einstein's summation convention.

where $\hat{h}_t = \int_M h\mu_t$ and $\mathscr{A}^\dagger$ is the adjoint of $\mathscr{A}$ with respect to the dual pairing $\langle \mu\varphi \rangle$ of volume forms and smooth functions, i.e., for all bounded $\varphi \in C^\infty(M)$ and all volume forms μ we have

$$\int_M \varphi \mathscr{A}^\dagger \mu = \int_M (\mathscr{A}\varphi)\mu. \tag{5}$$

Eq. (4) is known as the Kushner-Stratonovich equation, see, e.g., [6].

C. Unweighted Particle Filters

In unweighted particle filtering, the goal is to find a Monte-Carlo approximation of μ_t, i.e., for any $N = 1, 2, \ldots,$, the objective is to find processes $S_t^{(i)}$, $i = 1, \ldots, N$, called *particles* such that $\mu_t \approx \frac{1}{N}\sum_{i=1}^N \delta_{S_t^{(i)}}$. The processes $S_t^{(i)}$ should be adapted to $\mathscr{F}_t^{N,Z}$, where Z is a vector-valued process independent of X and N that can capture additional noise in the particle dynamics. Usually, one is interested in 'symmetric' particle representations in which all $S_t^{(i)}$ have identical distributions. The problem thus is to specify dynamics for a representative process S_t that depend on the particle ensemble.

D. Feedback Particle Filter

For Gaussian white noise observations, a recipe for building such a particle filter is known. Let us briefly review the derivation of the feedback particle filter (FPF) [2] (see [4] for the manifold setting). The FPF uses particle dynamics given by the prior dynamics plus a feedback control term dU_t that is chosen such that the Fokker-Planck equation for a single particle gives the same change in distribution as the filtering equation. An ansatz of $dU_t = K_t \circ dY_t + \Omega_t dt$ gives

$$dS_t = V_0 dt + V_j \circ dZ_t^j + K_t \circ dY_t + \Omega_t dt, \tag{6}$$

where Z_t^j is an independent copy of B_t^j. A corresponding equation for the conditional distribution of S_t given $\mathscr{F}_t^Y$, denoted by $\bar{\mu}_t$, can be derived by an integration-by-parts argument using Lie derivatives:

$$\begin{aligned} d\int_M \varphi\bar{\mu}_t &= \int_M (\mathscr{A}\varphi dt + K_t\varphi \circ dY_t + \Omega_t\varphi dt)\bar{\mu}_t \\ &= \int_M \varphi\Big(\mathscr{A}^\dagger\bar{\mu}_t dt - \mathscr{L}_{K_t}\bar{\mu}_t \circ dY_t - \mathscr{L}_{\Omega_t}\bar{\mu}_t dt\Big) \\ &\quad + \text{boundary terms}. \end{aligned} \tag{7}$$

In the first line, the Stratonovich chain rule is used. In the second line, directional derivatives are replaced by Lie derivatives,[4] and we performed integration by parts, reducing exact top-degree forms to boundary terms using Stokes' theorem. It is customary to demand that K, Ω be tangent to the boundary of M (if ∂M is nonempty), or even completely vanish on ∂M. This assumption implies $i_K\mu = 0$ on ∂M, such that the boundary terms can be discarded. After switching back to Itô calculus, one obtains

$$d\bar{\mu}_t = \Big(\mathscr{A}^\dagger\bar{\mu}_t - \mathscr{L}_{\Omega_t}\bar{\mu}_t + \tfrac{1}{2}\mathscr{L}_{K_t}^2\bar{\mu}_t\Big)dt - (\mathscr{L}_{K_t}\bar{\mu}_t)dY_t. \tag{8}$$

[4]On smooth functions, the Lie derivative agrees with the directional derivative, i.e., $\mathscr{L}_X\varphi = X\varphi = d\varphi(X)$ for all $\varphi \in C^\infty(M)$, $X \in \mathrm{Vect}(M)$.

Matching the terms of Eq. (8) with Eq. (4) (conditioned on $\bar{\mu}_t = \mu_t$) leads to the system of equations[5]

$$\mathscr{L}_{K_t}\mu_t = -(h - \hat{h}_t)\mu_t, \tag{9}$$

$$\mathscr{L}_{\Omega_t}\mu_t = \frac{1}{2}\Big(h^2 - \hat{h}^2 - K_t h\Big)\mu_t. \tag{10}$$

Given a vector field K_t solving Eq. (9), called a *gain* for the FPF, setting

$$\Omega_t = -\tfrac{1}{2}(h + \hat{h})K_t \tag{11}$$

gives an associated solution to Eq. (10).[6]

E. Uniqueness, Approximation, and Estimation of the Gain

The solutions of Eqs. (9) and (10) are not unique, as any pair (K_t, Ω_t) of solutions can be modified by adding an arbitrary divergence-free[7] vector field V, i.e., such that $\mathscr{L}_V\mu_t = 0$. Uniqueness can be obtained by fixing a riemannian metric g, and then demanding that the gain take the form $K_t = \operatorname{grad}\phi_t$. This leads to the equation $\mathscr{L}_{\operatorname{grad}\phi_t}\mu_t = -(h - \hat{h}_t)\mu_t$. Moreover, if vol_g denotes the riemannian volume form and μ_t is expressed in terms of the density p_t as $\mu_t = p_t\mathrm{vol}_g$, Eq. (9) reduces to a (weighted) Poisson equation

$$\operatorname{div}_{\mathrm{vol}_g}(p_t \operatorname{grad}\phi_t) = -(h - \hat{h}_t)p_t. \tag{12}$$

Existence and uniqueness of a solution is guaranteed under mild assumptions on p_t and h (see [7, Th. 2.2]), and $K_t = \operatorname{grad}\phi_t$ minimizes the functional $K \mapsto \int_M g(K, K)\mu$ among all solutions of Eq. (9) (see [8, Lemma 8.4.2]). In the case $M = \mathbb{R}^n$, Euclidean g, Gaussian p_t, and linear h, this gain reduces to the Kalman gain.

Sometimes it is desirable to approximate the vector field $K_t = \operatorname{grad}\phi_t$, where ϕ_t solves Eq. (12), by a constant. As mentioned in the introduction, in order to define the notion of constancy on a manifold, an additional structure ∇, called connection, has to be defined. One may choose the Levi-Civita connection corresponding to some (already given) g, but other choices are possible. A constant gain K^{CG} can then be defined as the minimum of $\|K - \operatorname{grad}\phi^2\|$ over all parallel K (i.e., $\nabla K = 0$). For example, when $M = \mathbb{R}^n$, g is the Euclidean metric, and ∇ its Levi-Civita connection,

$$K^{\mathrm{CG}} = \int_{\mathbb{R}^n} (\operatorname{grad}\phi)\mu = \int_{\mathbb{R}^n} x(h(x) - \hat{h})\mu(dx). \tag{13}$$

The right-hand representation is obtained by multiplying the Eq. (12) by x, integrating by parts, and using $\operatorname{grad} x_i = \partial_{x_i}$. Eq. (13) is convenient because the RHS can be estimated by a sample, but on some manifolds, topological obstructions make this approach infeasible. On S^1 with the standard metric and connection, a constant vector field cannot be a gradient of a smooth function. Insisting and performing the calculation on a chart leads to $K^{\mathrm{CG}} = \int_0^{2\pi}\theta(h(\theta) - \hat{h})p(\theta)d\theta + 2\pi K(0)p(0)$. It is unclear how to estimate the additional term that depends on the *exact* gain. In other cases the situation is still worse: many manifolds with connection do not have *any* nontrivial parallel vector fields (a common example is S^2 with its standard connection).

In practise, the gain $K_t = \operatorname{grad}\phi_t$ has to be estimated from a finite number of particles $S_t^{(i)} \in M$, $i = 1, \dots, N$, thought to be i.i.d. samples from μ_t. If only the gain at the particle locations is needed, we denote the mapping particles→gains by $\mathbf{K}_t = \mathfrak{G}(\mathbf{S}_t, h)$, where $\mathbf{K}_t = ((K_t)_{S_t^{(i)}})_{i=1}^N$ and $\mathbf{S}_t = (S_t^{(i)})_{i=1}^N$. This is called the gain estimation problem. For the purposes of this letter, the question of how to optimally estimate the gain shall be left aside and we refer to, e.g., [9], [10], [11] and the references therein. The aim is to show that the construction of an FPF-like algorithm for point processes can be fully reduced to the same types of equations as for the FPF gain, i.e., to equations of the following form:

Definition 1: For every positive volume form μ with $\int_M \mu = 1$ and every smooth function φ we denote by $\mathcal{E}(\mu, \varphi)$ the equation

$$\mathcal{E}(\mu, \varphi): \quad \mathscr{L}_V\mu = -\left(\varphi - \int_M \varphi\mu\right)\mu, \tag{14}$$

whose unknown quantity is the vector field V.

III. FPF for Point Process Observations

Now, we consider the case where the hidden state X_t is a diffusion on a manifold as in Section II, but the observation process is now a counting process[8] N_t, counting the number of events since time $t = 0$, with intensity function $h(X_t)$, where $h : M \to (0, \infty)$ is called the *observation function*. Here, the observations are corrupted by Poisson noise.

An equation for the optimal filter is known also in this setting. If the distribution of X_0 is described in terms of a volume form μ_0, the conditional distribution μ_t of X_t given observations $\mathscr{F}_t^N$ evolves according to the equation

$$d\mu_t = (\mathscr{A}^\dagger\mu_t)\,dt + \left(\frac{h}{\hat{h}_{t^-}} - 1\right)\mu_{t^-}(dN_t - \hat{h}_t dt), \tag{15}$$

where t^- denotes left limits. Eq. (15) will be referred to as the filtering equation for point process observations. It is sometimes called Kushner-Stratonovich-Poisson equation (see [5] for further references).

The goal of the present section is to carry out the derivation of an FPF for point process observations. We will call the resulting filter *feedback particle filter for point process observations*, or ppFPF for short.

In the following two subsections, we will separately derive the drift and the jump terms of the particle dynamics. The separation of these two aspects is necessary because the drift

[5] $K_t h = dh(K_t) = i_{K_t}dh$ denotes the directional derivative of h in the direction of the vector field K_t, whereas hK_t is the vector field K_t scaled point-wise by the function h.

[6] This can be shown by using Cartan's magic formula and the graded product rule for the interior derivative, or simply by observing that $\mathscr{L}_{\varphi X}\mu = \varphi\mathscr{L}_X\mu + (X\varphi)\mu$ for all $\varphi \in C^\infty(M)$, $X \in \mathrm{Vect}(M)$, and $\mu \in \Omega^n(M)$.

[7] The divergence of a vector field V with respect to a volume form μ is the function $\operatorname{div}_\mu V$ defined implicitly by $\mathscr{L}_V\mu = (\operatorname{div}_\mu V)\mu$. Using Cartan's magic formula and the fact that $d\mu = 0$, the divergence can also be written as $\operatorname{div}_\mu V = \frac{di_V\mu}{\mu}$. It follows that for $f > 0$ we have $f\operatorname{div}_{f\mu} V = \operatorname{div}_\mu(fV) = df(V) + f\operatorname{div}_\mu V$.

[8] By convention, N_t is right-continuous with left limits (càdlàg).

term is infinitesimal, i.e., a vector field, whereas the event term is an instantaneous transformation of the particles from the prior to the posterior. Since a vector field (infinitesimal) and a finite transformation cannot be easily mixed, the ppFPF lacks the gain×error structure of the FPF, with a common prefactor. This will be shown below.

A. Derivation of the Drift Term

We first consider the terms proportional to dt in Eq. (15), describing the evolution of the conditional distribution *in-between events*, and make the following ansatz for the particle dynamics:

$$dS_t = V_0 dt + V_j^i \circ dZ_t^j + \Omega_t dt. \tag{16}$$

Since the modification is deterministic, the corresponding equation for the conditional distribution of S_t given $\mathscr{F}_t^Y$ simply reads

$$d\bar{\mu}_t = \left(\mathscr{A}^\dagger \bar{\mu}_t - \mathscr{L}_{\Omega_t}\bar{\mu}_t\right)dt. \tag{17}$$

Matching this to Eq. (15) (again, setting $\bar{\mu}_t = \mu_t$) yields the relation

$$\mathscr{L}_{\Omega_t}\mu_t = (h - \hat{h}_t)\mu_t, \tag{18}$$

which is $\mathcal{E}(\mu_t, -h)$, up to a sign the same as Eq. (9) for the gain of the FPF. Thus, up to divergence-free terms, the drift of the ppFPF is identical to the negative gain of the corresponding FPF (i.e., with the same h).

B. Derivation of the Jump Term

Upon an event, Eq. (15) prescribes a change of the conditional distribution as follows:

$$\mu_{t^-} \mapsto \mu_t = \frac{h}{\hat{h}_{t^-}}\mu_{t^-}, \tag{19}$$

i.e., the distribution is multiplied by the observation function and subsequently renormalized. This requires a corresponding instantaneous change of the particle positions, i.e., $S_{t^-} \mapsto S_t = T_{t^-}(S_{t^-})$, where $T_t : M \to M$ satisfies the constraint

$$(T_{t^-})_*\mu_{t^-} = \frac{h}{\hat{h}_{t^-}}\mu_{t^-}, \tag{20}$$

where $_*$ denotes the pushforward. In rare cases, such as for Gaussian p and exponential h, this functional equation has exact closed-form solutions. In the absence of an exact solution, a solution T_{t^-} to Eq. (20) can be approximated by an iterative procedure, also used in [12], [13], by an adaptation of Moser's classical result [14]. The idea is to define an interpolation[9] $\tilde{\mu}_{t,s}$ of $\mu_t = \mu_{t^-}$ and $\frac{h}{\hat{h}_{t^-}}\mu_{t^-}$:

$$\tilde{\mu}_{t,s} = \frac{h^s\mu_{t^-}}{\int_M h^s\mu_{t^-}}, \quad 0 \le s \le 1. \tag{21}$$

[9]The chosen interpolation is sometimes called log-homotopy and has the virtue of producing a PDE analogous to the one for the drift term. Other smooth interpolations can be used as needed.

Algorithm 1 Log Homotopy Particle Flow (Deterministic)

Input: $\mathbf{S}_0, n, \log h$
$\mathfrak{G}$ = gain estimation method (see Section II-E)
Set $ds = 1/n$
for $i = 1$ to n **do**
  Estimate vector field: $\mathbf{V}_{i \cdot ds} := \mathfrak{G}(\mathbf{S}_{(i-1)ds}, \log h)$
  for $j = 1$ to N **do**
    $S^j_{i\cdot ds} \leftarrow S^j_{(i-1)ds} + V^j_{i\cdot ds} ds$
  end for
end for
return $\mathbf{S}_1$

Algorithm 2 point-process feedback particle filter

Input: $dt, T = ndt, \mathscr{A}, \mu_0, h, (N_t)_{t=0}^T, N$
$\mathfrak{G}$ = gain estimation method (see Section II-E)
EM = Euler-Maruyama method
Sample S_0^j from μ_0 for $j = 1$ to N
for $i = 1$ to n **do**
  $t = i \cdot dt$
  Sample dZ_t^k from $\mathcal{N}(0, dt)$ for $k = 1$ to r
  Estimate gain: $\Omega_t := \mathfrak{G}(\mathbf{S}_{(i-1)dt}, -h)$
  for $j = 1$ to N **do**
    Predict: $S_t^j \leftarrow \mathrm{EM}(S^j_{(i-1)dt}, \mathscr{A}, dt, d\mathbf{Z}_t)$
    Correct: $S_t^j \leftarrow S_t^j + \Omega_t^j dt$
    $k := N_t - N_{t-dt}$
    while $k > 0$ **do**
      Transform: $S_t^j \leftarrow T_t(S_t^j)$ (e.g. by Algorithm 1)
      $k \leftarrow k - 1$
    end while
  end for
end for
return $(\mathbf{S}_{i\cdot dt})_{i=0}^n$

We then match this flow of probability distributions with a flow of particles, i.e., the flow of an s-dependent vector field $V_{t,s}$ satisfying

$$\mathscr{L}_{V_{t,s}}\tilde{\mu}_{t,s} = -\frac{d}{ds}\tilde{\mu}_{t,s} = -\left(\log h - \int_M (\log h)\tilde{\mu}_{t,s}\right)\tilde{\mu}_{t,s}, \tag{22}$$

which is equation $\mathcal{E}(\mu_{t,s}, \log h)$ in Definition 1. This procedure results in Algorithm 1.

C. Exactness of the Particle Filter

Thus, the ppFPF is defined in terms of the following dynamics, yielding a càdlàg process:

$$\text{in-between events: } dS_t = V_0 dt + V_j^i \circ dZ_t^j + \Omega_t dt, \tag{23}$$

$$\text{event at time } t\text{: } S_t = T_{t^-}(S_{t^-}), \tag{24}$$

where Ω_t is a vector field that solves Eq. (18) and T_{t^-} is the diffeomorphism constructed in Section III-B. The PDEs to be solved for both steps are of the forms $\mathcal{E}(\mu, -h)$ and $\mathcal{E}(\mu, \log h)$, and are therefore analogous to the PDE for the gain of the FPF. As a result, all considerations in Section II-E apply to the ppFPF. By construction, the ppFPF has the following property of being exact:

Theorem 1: Let μ_t denote the conditional distribution of X_t given $\mathscr{F}_t^N$. Under assumption A, if the distribution of S_0 coincides with μ_0, and if the process $(S_t)_{t\ge 0}$ is defined according to Eqs. (23)-(24), then the conditional distribution of S_t given $\mathscr{F}_t^N$ coincides with μ_t for all $t \ge 0$.

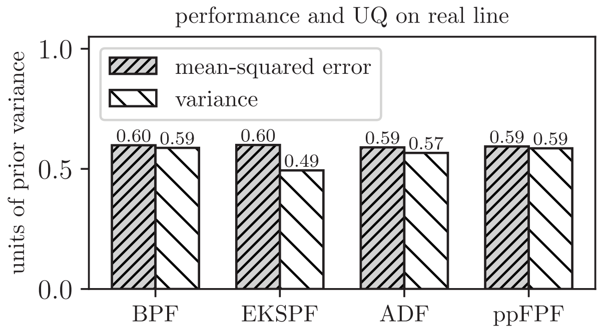

Fig. 2. Comparison of filter performance and UQ for the model $dX_t = -X_t dt + \sqrt{2} dW_t$ and $h(x) = 2e^x$ on $M = \mathbb{R}$ (BPF: bootstrap particle filter, EKSPF: filter from [5], ADF: Gaussian assumed-density filter [15], ppFPF: this letter). Simulations used $dt = 0.01$ and were run for 10^8 time-steps. BPF, EKSP, and ppFPF used $N = 200$ particles. The gain estimation method for the ppFPF used parameters $\epsilon = 10$ and $\lambda = 10^{-7}$. While all filters have comparable performance (first-order statistics), the uncertainty is more strongly under-estimated for the EKSPF and ADF compared to the asymptotically exact filters (BPF and ppFPF).

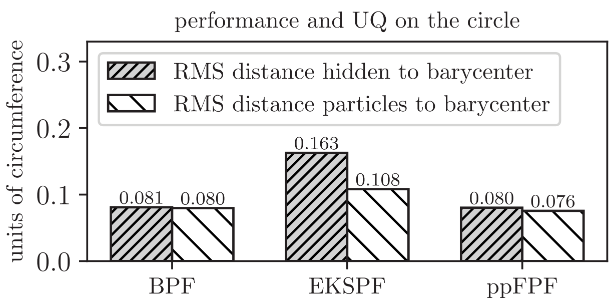

Fig. 3. Comparison of filters on $M = S^1$ (see Fig. 2). The model is Brownian motion on S^1 and observations are independent Poisson processes with intensity functions $h_i(\theta) = 20 \exp(10(\cos(\theta - i\pi/2) - 1))$, $i = 1, \ldots, 4$. The Riemannian distance $d(\theta_1, \theta_2) = \pi - ||\theta_1 - \theta_2| - \pi|$ is used to compute riemannian barycenters and errors. The gain estimation method for the ppFPF used $\lambda = 10^{-2}$ and a von Mises kernel with $\kappa = 0.1$. Simulations used $dt = 0.01$, $N = 200$, and were run for 10^5 time-steps. In this example, both performance and UQ is compromised for the EKSPF.

Algorithms 1 & 2 additionally require the choice of a specific gain estimation algorithm.

IV. Numerical Results

Simulations were conducted in order to study the performance (in terms of mean-squared error) and UQ (in terms of posterior variance) of the ppFPF in comparison to other well-known approximate filters for a filtering problem on $M = \mathbb{R}$ (see Fig. 2) as well as $M = S^1$ (Fig. 3). The ppFPF was implemented with the differential loss reproducing kernel Hilbert space method from [16] (see figure captions for parameters). The bootstrap particle filter (BPF) was resampled when N_{eff}/N dropped below 1/2, where $N_{\text{eff}} = 1/\sum_{i=1}^{N}(w^{(i)})^2$. For $M = S^1$, the EKSPF was naïvely[10] applied to the chart on the interval $[0, 2\pi)$.

[10] We emphasize that the EKSPF was not intented/designed to be used in this way. This example only serves to illustrate that a naïve application can lead to poor performance, which is to be expected due to the conceptual reasons outlined in the introduction.

V. Conclusion

In this letter, we reviewed the problem of designing unweighted particle filters for a manifold-valued hidden process observed in Poisson noise. We provided conceptual arguments as well as numerical illustrations that the existing approach from [5] (EKSPF) is limited by an intrinsic constant gain approximation, which compromises higher-order statistics as well as the ability to be extended to manifolds. We then derived an asymptotically exact unweighted particle filter, called ppFPF, by matching the particle forward equation with the equation for the optimal filter. This approach starts from first principles and is analogous to the derivation of the FPF. The resulting filter does not have the gain×error structure of the FPF, but can otherwise be reduced to partial differential equations that are completely analogous to the ones in the FPF. This makes it possible to leverage existing and future approaches to gain estimation in the FPF. As an unweighted filter, the ppFPF is expected to scale to high-dimensional problems [3].

References

[1] T. Yang, P. G. Mehta, and S. P. Meyn, “A mean-field control-oriented approach to particle filtering,” in *Proc. Amer. Control Conf.*, San Francisco, CA, USA, 2011, pp. 2037–2043.
[2] T. Yang, P. G. Mehta, and S. P. Meyn, “Feedback particle filter,” *IEEE Trans. Autom. Control*, vol. 58, no. 10, pp. 2465–2480, Oct. 2013.
[3] S. C. Surace, A. Kutschireiter, and J.-P. Pfister, “How to avoid the curse of dimensionality: Scalability of particle filters with and without importance weights,” *SIAM Rev.*, vol. 61, no. 1, pp. 79–91, 2019.
[4] C. Zhang, A. Taghvaei, and P. G. Mehta, “Feedback particle filter on Riemannian manifolds and matrix lie groups,” *IEEE Trans. Autom. Control*, vol. 63, no. 8, pp. 2465–2480, Aug. 2018.
[5] M. Venugopal, R. M. Vasu, and D. Roy, “An ensemble Kushner–Stratonovich–Poisson filter for recursive estimation in nonlinear dynamical systems,” *IEEE Trans. Autom. Control*, vol. 61, no. 3, pp. 823–828, Mar. 2016.
[6] A. Bain and D. Crisan, *Fundamentals of Stochastic Filtering* (Stochastic Modelling and Applied Probability), vol. 60. New York, NY, USA: Springer, 2009.
[7] R. S. Laugesen, P. G. Mehta, S. P. Meyn, and M. Raginsky, “Poisson’s equation in nonlinear filtering,” *SIAM J. Control Optim.*, vol. 53, no. 1, pp. 501–525, 2015.
[8] L. Ambrosio, N. Gigli, and G. Savaré, *Gradient Flows*, 2nd ed. Basel, Switzerland: Birkhäuser, 2008.
[9] A. Taghvaei and P. G. Mehta, “Gain function approximation in the feedback particle filter,” in *Proc. IEEE 55th Conf. Decis. Control (CDC)*, Las Vegas, NV, USA, 2016, pp. 5446–5452.
[10] A. Radhakrishnan, A. Devraj, and S. Meyn, “Learning techniques for feedback particle filter design,” in *Proc. IEEE 55th Conf. Decis. Control (CDC)*, 2016, pp. 5453–5459.
[11] K. Berntorp and P. Grover, “Feedback particle filter with data-driven gain-function approximation,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. 54, no. 5, pp. 2118–2130, Oct. 2018.
[12] F. Daum, J. Huang, and A. Noushin, “Exact particle flow for nonlinear filters,” in *Proc. SPIE Signal Process. Sensor Fusion Target Recognit. XIX*, vol. 7697. Orlando, FL, USA, 2010, Art. no. 769704. [Online]. Available: https://www.spiedigitallibrary.org/conference-proceedings-of-spie/7697/769704/Exact-particle-flow-for-nonlinear-filters/10.1117/12.839590.full?SSO=1
[13] S. Reich, “A dynamical systems framework for intermittent data assimilation,” *BIT Numer. Math.*, vol. 51, no. 1, pp. 235–249, 2011.
[14] J. Moser, “On the volume elements on a manifold,” *Trans. Amer. Math. Soc.*, vol. 120, no. 2, pp. 286–294, 1965.
[15] J.-P. Pfister, P. Dayan, and M. Lengyel, “Know thy neighbour: A normative theory of synaptic depression,” in *Proc. Adv. Neural Inf. Process. Syst.*, Vancouver, BC, Canada, 2009, pp. 1464–1472.
[16] A. Radhakrishnan and S. Meyn, “Feedback particle filter design using a differential-loss reproducing kernel Hilbert space,” in *Proc. Annu. Amer. Control Conf. (ACC)*, Milwaukee, WI, USA, 2018. pp. 329–336.