Gauge Freedom within the Class of Linear Feedback Particle Filters

Ehsan Abedi, Simone Carlo Surace

Abstract—Feedback particle filters (FPFs) are Monte-Carlo approximations of the solution of the filtering problem in continuous time. The samples or particles evolve according to a feedback control law in order to track the posterior distribution. However, it is known that by itself, the requirement to track the posterior does not lead to a unique algorithm. Given a particle filter, another one can be constructed by applying a time-dependent transformation of the particles that keeps the posterior distribution invariant. Here, we characterize this gauge freedom within the class of FPFs for the linear-Gaussian filtering problem, and thereby extend previously known parametrized families of linear FPFs.

I. INTRODUCTION

The filtering problem is the problem of estimating a quantity evolving in time that is accessible only through partial and noisy observations. This is commonly formalized as the problem of finding the conditional distribution of the hidden state at time \( t \) given all observations up to that time. Despite its long history and rich developments around its theory (see [1], Section 1.3), the principal challenge of implementing filters in practical applications centers around the lack of closed-form solutions and the resulting necessity to find efficient and robust numerical approximations.

The issue has become even more severe in the era of big data where the dimensionality of the processes is very large. The main approach to approximate the conditional distribution, sequential Monte-Carlo or particle filtering (see [2] for a survey and pointers to the literature), is known to exhibit a curse of dimensionality as the number of dimensions of the observations grows [3]-[6]. The problem can be traced down to the use of importance weights and their increasing degeneracy as time progresses (see [6] and the references therein).

More recently, particle filters without importance weights [7]-[11] have been gaining attention. While lacking the principal vulnerability of weighted particle filters, many theoretical questions remain open [12]-[14]. One question that has recently received some interest is the non-uniqueness of the law of the process approximating the filtering distribution. It has been shown [15]-[16] that at least in the linear-Gaussian case there are many ways to construct particle dynamics – both deterministic and stochastic – that track the exact posterior distribution (given by the Kalman-Bucy filter).

E. Abedi performed this research during a semester project at University of Bern, Switzerland, while enrolled in the Master Program in Computational Science and Engineering at EPFL, Lausanne, Switzerland, and was supported by the Vahabzadeh Foundation. ehsan.abedi@epfl.ch

S.C. Surace is with the Department of Physiology, University of Bern, Switzerland, and is supported by the Swiss National Science Foundation, grant. surace@pyl.unibe.ch

Here, we systematically investigate the degrees of freedom in choosing dynamics of the particles (within some constraints) for the linear-Gaussian filtering problem while keeping their distribution aligned with the exact conditional distribution (assuming that the initial distribution of the filter is Gaussian). We characterize a group of transformations (which we call gauge transformations, taking inspiration by similar transformations appearing in theoretical physics) that preserve the conditional distribution and describe how it acts on certain classes of feedback particle filters. Moreover, we propose a cost function on the family of all such filters. We identify a known and a new feedback particle filter as a (constrained) optimum of this cost function.

The remainder of the paper is structured as follows: in Section II, we discuss the filtering problem in general and the linear-Gaussian case that is the focus of this paper in particular. We define the notion of particle filter that is the focus of this work and provide an overview of previous work regarding such filters and the history of the non-uniqueness problem. In Section III we discuss the symmetries (gauge transformations) of the problem, define certain classes of filters on which the symmetries act, and describe this action in detail. In Section IV, we introduce an optimality criterion and optimize it with and without constraints, obtaining two specific types of feedback particle filters. Lastly, in Section V we discuss the implications of our results and comment on future directions.

II. PRELIMINARIES AND BACKGROUND

The classical filtering problem is to find the conditional distribution of the hidden state \( X_t \) given observations \( \mathcal{F}_t^Y \) for the stochastic system given by

\[
\begin{align*}
    dX_t &= f(X_t)dt + g(X_t)dW_t, \quad (1) \\
    dY_t &= h(X_t)dt + dV_t, \quad (2)
\end{align*}
\]

where \( X_t \) and \( Y_t \) are valued in \( \mathbb{R}^n \) and \( \mathbb{R}^m \) respectively, \( f, g, h \) are (known) vector- or matrix-valued functions satisfying suitable conditions for the well-posedness of the stochastic differential equations, and \( W_t \) and \( V_t \) are independent \( n' \) and \( m \)-dimensional Brownian motions respectively. The distribution of \( X_0 \) is assumed to be known and independent of the Brownian motions.

A. Linear-Gaussian filtering problem

In this paper, we will focus on the special linear-Gaussian case of the above problem, where \( f, g, \) and \( h \) are chosen...
such that
\[ dX_t = AX_t dt + BdW_t, \]
\[ dY_t = CX_t dt + dV_t, \]
for some \( A \in \mathbb{R}^{n \times n} \), \( B \in \mathbb{R}^{n \times n'} \), and \( C \in \mathbb{R}^{m \times n} \), and \( X_0 \) has Gaussian distribution. This filtering problem has an exact solution, called the Kalman-Bucy filter [17]. The conditional distribution of \( X_t \) given \( \mathcal{F}_t^Y \) is multivariate Gaussian with mean \( \mu_t \) and covariance matrix \( P_t \) which jointly evolve as
\[ d\mu_t = A\mu_t dt + P_t C^T (dY_t - C\mu_t dt), \]
\[ dP_t = (BB^T + AP_t + P_t A^T - P_t C^T CP_t) dt, \]
where \( \mu_0 \) and \( P_0 \) are chosen according to the distribution of \( X_0 \). For convenience, throughout this article it will be assumed that \( P_0 \) is strictly positive definite such that \( P_t \) remains strictly positive definite for all \( t \geq 0 \) (see Proposition 1.1 in [18]).

B. Particle filters

In the context of this paper, a particle filter is any approximation of the conditional distribution of \( X_t \) given \( \mathcal{F}_t^Y \) by a set of (unweighted) samples or particles \( S_i^{(t)} \) for \( i = 1, \ldots, N \), such that \( S_i^{(t)} \) are \( \mathcal{F}_t^{Y,Z} \)-adapted processes. Here, \( \mathcal{F}_t^{Y,Z} \) is the filtration generated by the process \( (Y_t, Z_t) \), where \( Z_t \) is some process independent from \( \mathcal{F}_t^{X,Y} \) (for example, some additional noise in the particle dynamics). In the following, we will only consider ‘symmetric’ particle filters for which all particles have the same conditional distribution given \( \mathcal{F}_t^Y \). We will therefore talk about the dynamics of a single representative particle \( S_t \), omitting the particle index \( i \). The distribution of \( S_t \) will still depend on the number of particles \( N \). Such a particle filter is called asymptotically exact if the distribution of \( S_t \) given \( \mathcal{F}_t^Y \) converges to the conditional distribution of \( X_t \) as \( N \to \infty \).

C. Feedback particle filter (FPF)

An asymptotically exact filter has been found in [9] and [10] based on mean-field optimal control with an algorithm called stochastic feedback particle filter (sFPF). Specifically, the objective of the sFPF is to find control laws \( u \) and \( K \) for the particle dynamics
\[ dS_t = f(S_t) dt + g(S_t) dW_t + u(S_t, t) dt + K(S_t, t) dY_t, \]
such that the particle filter whose particles \( S_i^{(t)} \) evolve according to the SDE (7) is exact. Here, \( W_t \) is an \( n' \)-dimensional Brownian motion that is independent of \( \mathcal{F}_t^{X,Y} \) and plays the role of \( Z_t \) in the previous paragraph. The sFPF can be derived by ‘aligning’ the Frobenius-Planck equation of the particle filter with the Kushner-Stratonovich filtering equation in an appropriate sense (see [10], Section 2), which leads to a McKean-Vlasov or mean-field equation where \( u(\cdot, t) \) and \( K(\cdot, t) \) depend on the distribution of \( S_t \).

As a result, the FPF is exact in the mean-field sense if initialized with the correct initial distribution. However, the computation of the gain term \( K \) requires the solution of a linear boundary value problem (BVP) at each instant in time, which incurs a large computational cost.

In the linear-Gaussian case, however, the solution of the BVP is given by the Kalman filter (see [10], Section 4.2), and the resulting FPF – which will henceforth be referred to as the stochastic linear FPF (sFPF) – is equivalent to the square-root form of the Ensemble Kalman-Bucy filter. It takes the form
\[ dS_t = AS_t dt + P_t C^T (dY_t - \frac{1}{2} C(S_t + \mu_t) dt) + BdW_t. \]

If \( S_0 \) is normally distributed with mean \( \mu_0 \) and covariance matrix \( P_0 \), the conditional distribution of \( S_t \) given \( \mathcal{F}_t^Y \) has mean and covariance \( \mu_t \) and \( P_t \) respectively for all \( t \geq 0 \).

D. Non-uniqueness of the control law

The sFPF is not the only choice of dynamics for \( S_t \) for which the distribution agrees with the Kalman-Bucy filter. If the only requirement is to track the posterior distribution, one may construct filters that do not satisfy the condition that gives rise to the sFPF. A number of examples are known in the linear-Gaussian case. In [15] a deterministic linear FPF (deterministic refers to the fact that there is no independent noise term) was derived from an optimal transport perspective. We will refer to it as optimal transport deterministic linear feedback particle filter (OTdetFPF). Its dynamics are given by
\[ dS_t = AS_t dt + P_t C^T (dY_t - \frac{1}{2} C(S_t + \mu_t) dt) + \hat{\Omega}_t P_t^{-1} (S_t - \mu_t) dt, \]
where \( \hat{\Omega}_t \) is a skew-symmetric matrix determined uniquely from the equation
\[ \hat{\Omega}_t P_t^{-1} + P_t^{-1} \hat{\Omega}_t = A^T - A + \frac{1}{2} \left( P_t^{-1} BB^T - BB^T P_t^{-1} + P_t C^T C - C^T CP_t \right). \]
The OTdetFPF replaces the Brownian motion term of the sFPF by a repulsive term that drives the particles away from their mean. In addition, it has a skew-symmetric term that arises from the minimization of the transportation cost. Since the skew-symmetric term does not affect the distribution, a parametrized family of deterministic FPFs (detFPF) may be constructed by choosing it arbitrarily, in particular by setting it to zero.

More recently, [20] introduced another approach based on optimal control theory and duality formalisms to extend the FPF in the linear case. They constructed a two-parameter family of FPFs with parameters \( \gamma_1, \gamma_2 \in [0, 1] \). For \( \gamma_2 = 0 \), the parameter \( \gamma_1 \) interpolates between the sFPF (\( \gamma_1 = 1 \)) and a detFPF (\( \gamma_1 = 0 \)) that corresponds to Eq. (9) with \( \hat{\Omega}_t = 0 \). On the other hand, the parameter \( \gamma_2 \) trades off the ‘hedging’ in the feedback term (how to mix or hedge the mean of all
the particle positions with the individual particle position) with additional observation noise.

E. Notations and definitions

We denote by $S$ the process $(S_t)_{t \geq 0}$ as a random variable with values in a suitable subspace of functions $\mathbb{R} \to \mathbb{R}^n$ (for most cases, the space of continuous functions will suffice), with its law given by a probability measure on the corresponding Borel $\sigma$-algebra. If the conditional distribution of $S_t$ given $\mathcal{F}^Y_t$ agrees with the conditional distribution of $X_t$ given $\mathcal{F}^Y_t$ for all $t \geq 0$, such a process is called particle filter $\mathcal{F}^Y_t$. If $S$ is adapted to $\mathcal{F}^Y_t$, where $Z$ consists of $r \geq 0$ Brownian motions independent of $Y$ and if $S$ is the solution to a stochastic differential equation, it will be called feedback particle filter. If $r = 0$ it is called deterministic feedback particle filter. We denote by $\text{Skew}(n)$ the real vector-space of skew-symmetric $n \times n$-matrices, i.e. of matrices $X$ such that $X^\top = -X$, where $\cdot^\top$ denotes transposition. The notation $\text{GL}(n)$ is used for the group of invertible $n \times n$-matrices.

III. The class of linear particle filters and their symmetries

In the linear-Gaussian case, a particle filter is exact if the conditional distribution of $S_t$ given $\mathcal{F}^Y_t$ is multivariate Gaussian with mean equal to $\mu_t$ and covariance matrix equal to $P_t$ (the mean and covariance matrix of the Kalman-Bucy filter). If we split off the mean as $S_t = \mu_t + E_t$, the process $E_t$ can be any Gaussian process with mean equal to zero and covariance function $K(t, t')$ such that $K(t, t) = P_t$ for all $t \geq 0$. In this section, we first look at a group of linear time-dependent transformations that preserve this structure, called gauge transformations. Then we consider the subclass $\mathcal{F}$ of filters for which $E_t$ is the solution to a stochastic differential equation, which we will call the group of linear feedback particle filters, and the subclass $\mathcal{F}_{\text{det}} \subset \mathcal{F}$ for which $E_t$ does not have an independent source of noise, called deterministic linear feedback particle filters. We then describe the action of the group on $\mathcal{F}$, proving that the group of gauge transformations acts transitively on $\mathcal{F}_{\text{det}}$.

A. Gauge freedom of linear particle filters

If a particle filter $S_t = \mu_t + E_t$ is given, another particle filter may be constructed by applying an affine transformation $g_t$, i.e. $\tilde{S}_t = g_t(S_t - \mu_t) + \mu_t$, or simply $\tilde{S}_t = \mu_t + \tilde{E}_t$, where $\tilde{E}_t = g_t E_t$ and $g_t \in \text{GL}(n)$ is adapted to $\mathcal{F}^Y_t$ and satisfies

$$g_t P_t g_t^\top = P_t$$

for all $t \geq 0$. For fixed $t$, the subset consisting of those $g \in \text{GL}(n)$ that satisfy Eq. (11) forms a (random and time-dependent) Lie subgroup of $\text{GL}(n)$, which we denote by $G_t$. Its Lie algebra $\mathfrak{g}_t$ consists of all matrices of the form $\Omega P_t^{-1}$, where $\Omega \in \text{Skew}(n)$ is arbitrary.

\[2\] We only consider exact particle filters here. The idea is that for finite $N$ the mean-field quantities appearing in the control law are replaced by empirical estimates, yielding an asymptotically exact filter (although the details are nontrivial, see e.g. [21]).

In principle, the choice of the function $g : [0, \infty) \to \mathcal{G}_t(n)$, $t \mapsto g_t$ does not have to be regular. For example, after simulating a sample of particles up to time $t$, a sample for $s \leq t$ can be modified by a transformation $g \in \mathcal{G}_s$ without concern for samples at other times as long as only information up to time $s$ is used. However, in the following we will restrict to choices of $g$ with more regularity.

For example, $g$ may be chosen to be a continuously differentiable function. We may then write $\tilde{g}_t = M_t P_t^{-1} g_t$. Eq. (11) then implies that $M_t = \frac{1}{2}(P_t - g_t P_t g_t^\top) + \Upsilon_t^{(0)}$, where $\Upsilon_t^{(0)}$ is a continuous but otherwise arbitrary function with values in $\text{Skew}(n)$. The skew-symmetric component of $M_t$ accounts for motion of $g_t$ along the group $\mathcal{G}_t$, whereas the symmetric component is due to the change in $\mathcal{G}_t$ itself.

More generally, $g_t$ may be given by the solution of a stochastic differential equation involving the observations, i.e.

$$dg_t = \left( M_t dt + \sum_{i=1}^m \Upsilon_t^{(i)} dW_t^{(i)} \right) P_t^{-1} g_t,$$

where $W_t^{(i)}$ denotes the component number $i$ of $W_t$. By differentiating Eq. (11) and matching terms, we obtain the constraint that

$$M_t = \frac{1}{2} \left( \dot{P}_t - g_t \dot{P}_t g_t^\top - \sum_{i=1}^m \Upsilon_t^{(i)} P_t^{-1} (\Upsilon_t^{(i)})^\top \right) + \Upsilon_t^{(0)},$$

where $\Upsilon_t^{(i)} \in \text{Skew}(n), i = 0, \ldots, m$.

We call any such function $g$, where $g_t$ evolves according to an SDE of the form (12) with (13) and arbitrary choices of $\Upsilon_t^{(i)}, i = 0, \ldots, m$, a deterministic gauge transformation. The orbit of a filter $S$ under all such transformations is called the deterministic extension of $S$.

B. Classes of linear feedback particle filters

Suppose that $S_t = \mu_t + E_t$, where the dynamics of $E_t$ are given by

$$dE_t = \left( G_t dt + \sum_{i=1}^m \Omega_t^{(i)} dW_t^{(i)} \right) P_t^{-1} E_t + \sum_{j=1}^r H_t^{(j)} dW_t^{(j)},$$

where $r \in \{0, 1, 2, \ldots\}$ (for $r = 0$ the sum over $j$ disappears by convention), $W_t^{(j)}$ are scalar Brownian motions independent of each other and of $\mathcal{F}_t^{X,Y}$, and $E_0$ has zero mean and covariance matrix equal to $P_0$. In addition, all coefficients are assumed to be adapted to $\mathcal{F}_t^{X,Y}$ and bounded. By construction, $S_t$ has conditional mean equal to $\mu_t$. Moreover, $S_t$ has covariance equal to $P_t$ for all $t \geq 0$ if and only if $\Omega_t^{(i)}, i = 0, 1, \ldots, m$ are in Skew$(n)$ and the following relation holds:

$$G_t = \frac{1}{2} \left( \dot{P}_t + \sum_{i=1}^m \Omega_t^{(i)} P_t^{-1} \Omega_t^{(i)} \right) - \sum_{j=1}^r H_t^{(j)} (H_t^{(j)})^\top + \Omega_t^{(0)}.$$

(15)
Definition 3.1: The class of all processes \( S_t = \mu_t + E_t \) with \( E_t \) having dynamics according to Eqs. (14) and (15) where \( \Omega_t^{(i)} \), \( i = 0, 1, \ldots, m \) are \( \mathcal{F}_t^Y \)-adapted Skew\((n)\)-valued processes, is denoted by \( \mathcal{S} \) and called the class of linear feedback particle filters. The subclass \( \mathcal{S}_{\text{det}} \subset \mathcal{S} \) for which \( r = 0 \) is called deterministic linear feedback particle filters.

C. Action of gauge transformations on the class of linear feedback particle filters

The following two results summarize the details of how gauge transformations act upon the classes \( \mathcal{S} \) and its subclasses defined above.

Proposition 3.2: Let \( S \in \mathcal{S} \) and \( \tilde{S}_t = \mu_t + \tilde{E}_t \), where \( \tilde{E}_t = g_t E_t \) with \( g \) a deterministic gauge transformation according to Eqs. (13) and (14). Then \( \tilde{S} \) also belongs to \( \mathcal{S} \) with \( \tilde{\Omega}_t^{(i)} \), \( i = 0, 1, \ldots, m \) and \( \tilde{H}_t^{(j)} \), given by

\[
\tilde{\Omega}_t^{(0)} = g_t \Omega_t^{(0)} g_t^\top + \Upsilon_t^{(0)},
\]
\[
\tilde{\Omega}_t^{(i)} = g_t \Omega_t^{(i)} g_t^\top + \Upsilon_t^{(i)}, \quad i = 1, \ldots, m,
\]
\[
\tilde{H}_t^{(j)} = g_t H_t^{(j)}, \quad j = 1, \ldots, r.
\]

Proof: We have

\[
d\tilde{E}_t = d g_t E_t + g_t dE_t + dg_t \, dE_t.
\]

By using Eqs. (12) and (13), we obtain

\[
dg_t E_t = \left( M_t dt + \sum_{i=1}^{m} \mathcal{Y}_t^{(i)} d\mathcal{Y}_t^{(i)} \right) P_t^{-1} \tilde{E}_t,
\]
\[
g_t dE_t = g_t \left( G_t dt + \sum_{i=1}^{m} \Omega_t^{(i)} d\mathcal{Y}_t^{(i)} \right) P_t^{-1} E_t
\]
\[
+ \sum_{j=1}^{r} g_t H_t^{(j)} d\mathcal{W}_t^{(j)},
\]
\[
dg_t \, dE_t = \sum_{i=1}^{m} \mathcal{Y}_t^{(i)} P_t^{-1} g_t \Omega_t^{(i)} P_t^{-1} E_t dt.
\]

By rewriting \( P_t^{-1} E_t = g_t^\top P_t^{-1} g_t^{-1} \tilde{E}_t \), noting that \( g_t^\top P_t^{-1} g_t^{-1} = P_t^{-1} \), and then collecting all the terms, we obtain Eqs. (17) and (18) as well as \( G_t = g_t G_t g_t^\top + M_t \). By using Eqs. (13) and (15), we obtain

\[
g_t G_t g_t^\top + M_t = \frac{1}{2} g_t P_t g_t^\top + \frac{1}{2} \sum_{i=1}^{m} g_t \Omega_t^{(i)} P_t^{-1} \Omega_t^{(i)} g_t^\top
\]
\[
- \frac{1}{2} \sum_{j=1}^{r} g_t H_t^{(j)} (g_t H_t^{(j)})^\top + g_t \Upsilon_t^{(0)} g_t^\top
\]
\[
+ \frac{1}{2} \tilde{P}_t - \frac{1}{2} g_t \tilde{P}_t g_t^\top - \frac{1}{2} \sum_{i=1}^{m} \mathcal{Y}_t^{(i)} P_t^{-1} (\mathcal{Y}_t^{(i)})^\top + \Upsilon_t^{(0)}.
\]

By rewriting \( P_t^{-1} = g_t^\top P_t^{-1} g_t \) and then using Eq. (17) to substitute \( g_t \Omega_t^{(i)} g_t^\top \), after cancelling all terms we obtain

\[
g_t G_t g_t^\top + M_t = \frac{1}{2} \left( \tilde{P}_t + \sum_{i=1}^{m} \tilde{\Omega}_t^{(i)} P_t^{-1} \tilde{\Omega}_t^{(i)}
\]
\[
- \sum_{j=1}^{r} \tilde{H}_t^{(j)} (\tilde{H}_t^{(j)})^\top + g_t \Upsilon_t^{(0)} g_t^\top + \Upsilon_t^{(0)},
\]

from which Eq. (16) follows. This concludes the proof.

From the transformation rules we immediately obtain the following:

Corollary 3.3: The deterministic gauge transformations act transitively on \( \mathcal{S}_{\text{det}} \), i.e. every deterministic gauge transformation of an \( S \in \mathcal{S}_{\text{det}} \) in \( \mathcal{S}_{\text{det}} \) and for any pair of \( S, \tilde{S} \in \mathcal{S}_{\text{det}} \) there is a deterministic gauge transformation that maps \( S \) to \( \tilde{S} \).

D. The one-dimensional case

In the case \( n = 1 \) (\( m \) arbitrary) any exact particle filter is of the form

\[
S_t = \mu_t + \sqrt{P_t} \theta_t,
\]

where \( \theta_t \) is any stationary Gaussian process with conditional mean zero and unit variance. In the case where \( \theta_t \) is a diffusion process, i.e. \( d\theta_t = \Lambda^{(0)}_t \theta_t dt + \sum_{j=1}^{r} \Lambda^{(j)}_t dW_t^{(j)} \), where \( \Lambda^{(j)}_t \) is adapted to \( \mathcal{F}_t^Y \) for all \( j = 0, 1, \ldots, r \), we are able to split the dynamics of \( E_t = \sqrt{P_t} \theta_t \) and bring it into the form of Eq. (14) as

\[
dE_t = \frac{\tilde{P}_t}{2 \sqrt{P_t}} dt + \sqrt{P_t} d\theta_t
\]
\[
= \frac{1}{2} \left( B B^\top + 2 A P_t dt - C^\top C P_t^2 + 2 P_t \Lambda^{(0)}_t \right) P_t^{-1} E_t dt
\]
\[
+ \sum_{j=1}^{r} P_t^{1/2} \Lambda^{(j)}_t dW_t^{(j)}.
\]

Here, it is necessary that \( \Lambda^{(0)}_t = -\frac{1}{2} \sum_{j=1}^{r} (\Lambda^{(j)}_t)^2 \) in order to have the matching variance.

Since \( \text{Skew}(1) = \{0\} \), the only nontrivial deterministic gauge transformation is the constant \( g_t = -1 \). It acts to simultaneously flip the signs of all \( \Lambda^{(j)}_t, j = 1, \ldots, r \). Correspodingly, the only deterministic linear feedback particle filter is the one for which \( \Lambda^{(j)}_t = 0, j = 1, \ldots, r \). Reassembling all the terms of \( S_t = \mu_t + E_t \), it takes the form

\[
dS_t = AS_t dt + P_t C^\top (dY_t - \frac{1}{2} C(S_t + \mu_t) dt)
\]
\[
+ \frac{1}{2} B B^\top P_t^{-1} (S_t - \mu_t) dt.
\]

IV. Optimality criteria

We will now return to the multidimensional case. In order to select a particle filter among all linear feedback particle filters, additional criteria are required. Since every filter \( S \in \mathcal{S} \) is specified by a choice of \( \Omega_t^{(i)} \in \text{Skew}(n) \) for \( i = 0, 1, \ldots, m \) and a vector \( H_t^{(j)} \in \mathbb{R}^n \) for \( j = 1, \ldots, r \).
and for all $t$, an optimality criterion can be formulated as a function of these quantities. Define
\[ \mathcal{L}_{P,r}(\Omega^{(0)}, ..., \Omega^{(m)}, H^{(1)}, ..., H^{(r)}) = \text{tr} G P^{-1} G^T, \]  
where $\text{tr}$ denotes the trace operator and $G$ is given by Eq. (14). This term appears to quadratic order in $dt$ when expanding $\text{tr} [dE_t dE_t^T | \mathcal{F}^Y]$ and can be associated with the transport cost as in [15]. It is therefore unsurprising that its minimization (under the constraint that the observation and Brownian motion terms are absent) yields the OTdetFPF from [15].

**Proposition 4.1:** The filter given by $\Omega_{t}^{(i)} = 0$, $i = 1, ..., m$, $H^{(j)} = 0$, $j = 1, ..., r$, and
\[ \Omega_{t}^{(0)} = \Omega_{t}^{*} = \arg\min_{\Omega \in \text{Skew}(n)} \mathcal{L}_{P,r}(\Omega, 0, ..., 0) \]  
is identical to the optimal transport deterministic linear feedback particle filter (OTdetFPF).

**Proof:** The OTdetFPF has dynamics
\[ dE_t = \left( A + \frac{1}{2} BB^T - \frac{1}{2} P_t C^T C P_t + \dot{\Omega}_t \right) P_t^{-1} dE_t dt, \]  
where $\dot{\Omega}_t$ is the unique solution of Eq. (10). This corresponds to Eq. (14) with $\Omega_{t}^{(i)} = 0$, $i = 1, ..., m$, $H^{(j)} = 0$, $j = 1, ..., r$, and
\[ \Omega_{t}^{(0)} = \frac{1}{2} \left( AP_t - P_t A^T + \dot{\Omega}_t \right). \]  
Any critical point of the function $\mathcal{L}_{P,r}(\Omega, 0, ..., 0)$ satisfies
\[ \text{tr} \left[ X \left( P^{-1} G^T - G P^{-1} \right) \right] = 0, \quad \forall X \in \text{Skew}(n). \]  
Since $(X, Y) \mapsto \text{tr} XY$ defines an inner product on $\text{Skew}(n)$, this implies $P^{-1} G^T - G P^{-1} = 0$. Substituting $G = \frac{1}{2} \dot{P} + \Omega$, we obtain the equation
\[ \Omega P^{-1} + P^{-1} \Omega = \frac{1}{2} \left( P^{-1} \dot{P} - \dot{P} P^{-1} \right) = \frac{1}{2} \left( P^{-1} BB^T - BB^T P^{-1} + P^{-1} AP + A^T \right. \]  
\[ - A - PA^T P^{-1} + PC^T C - C^T CP \right). \]  
It can be checked that the unique solution is given by $\Omega = \frac{1}{2} (AP - PA^T + \dot{\Omega})$. \hfill \blacksquare

Whereas the optimal transport formulation in [15] always yields deterministic filters, the cost function (28) may be minimized without the constraints of Proposition 4.1. This yields a new class of filters that has not yet explicitly appeared in the literature.

**Proposition 4.2:** Let $S$ be a particle filter specified by data $\mathcal{P}_t = (\Omega_{t}^{(0)}, \Omega_{t}^{(1)}, ..., \Omega_{t}^{(m)}, H_{t}^{(1)}, ..., H_{t}^{(r)})$ consisting of $\Omega_{t}^{(i)} \in \text{Skew}(n)$ for $i = 0, 1, ..., m$ and a vector $H_{t}^{(j)} \in \mathbb{R}^n$ for $j = 1, ..., r$ and for some $r \geq 0$. The following properties are equivalent:

i) $\mathcal{L}_{P,r}$ has a critical point at $\mathcal{P}_t$.
ii) $\mathcal{L}_{P,r}$ has a global minimum at $\mathcal{P}_t$.
iii) $\mathcal{P}_t$ is given by $\Omega_{t}^{(0)} = 0$ and
\[ \sum_{j=1}^{r} H_t^{(j)} (H_t^{(j)})^T - m \Omega_{t}^{(i)} P_t^{-1} \Omega_{t}^{(i)} = \dot{P}_t. \]  

**Proof:** By differentiating $\mathcal{L}_{P,r}$, we have i) if and only if
\[ 0 = \text{tr} \left[ K^{(0)} \left( P^{-1} G^T - G P^{-1} \right) \right] \]  
\[ + \sum_{i=1}^{m} \text{tr} \left[ K^{(i)} P^{-1} \Omega_{t}^{i} \left( P^{-1}(G)^T + G P^{-1} \right) \right] \]  
\[ - \sum_{j=1}^{r} \text{tr} \left[ l^{(j)} (H^{(j)})^T (P^{-1}(G)^T + G P^{-1}) \right] \]  
for all $K^{(i)} \in \text{Skew}(n)$ and all $l^{(j)} \in \mathbb{R}^n$, $i = 0, ..., m$ and $j = 1, ..., r$. Since $\text{tr}$ is an inner product, this is equivalent to
\[ P^{-1}(G)^T - G P^{-1} = 0, \]  
\[ P^{-1}(G)^T + G P^{-1} = 0, \]  
which in turn is equivalent to $G = 0$ and hence to iii). This shows i)⇒ii)(iii). Moreover, $G = 0$ when plugged into $\mathcal{L}_{P,r}$ gives a value of 0, which is the global minimum for this function. This shows that iii)⇒ii), which concludes the proof. \hfill \blacksquare

**Remark 4.3:** For $n = 1$, since all skew-symmetric $1 \times 1$-matrices are zero, condition iii) in Proposition 4.2 reads $\sum_{j=1}^{r} (h_{t}^{(j)})^2 = b^2 + 2aP_t - c^{2}P_t^2$. This has a solution if and only if $P_0 \leq P_t \leq \frac{a + \sqrt{a^2 + 4b^2}}{2c^2}$. Otherwise there are no global minimizers of $\mathcal{L}_{P,r}$.

**V. Discussion**

In this paper, we studied the intrinsic freedom in choosing dynamics of a particle filter if the only requirement for it is that the distribution of $S_t$ match the exact posterior distribution. We studied this for the example of the linear-Gaussian filtering problem where all calculations can be done explicitly, and the transformations can be assumed to be linear. However, as we will outline in this section, the implications of this study can be generalized to the nonlinear case as well.

The phenomenon that marginal distributions at each time $t$ do not give unique dynamics is well-known. For example, within the class of stationary Gaussian processes with zero mean and covariance of the stationary distribution equal to $K_0$, any covariance function $K(t - t')$ is admissible as long as $K(0) = K_0$. Likewise, for diffusion processes, fixing a distribution $p$, one may find an infinite family of stochastic differential equations whose solution has stationary distribution $p$. More generally, under a dynamical systems perspective, given a (sufficiently well-behaved) trajectory in the space of distributions over $\mathbb{R}^n$, we can find a multitude of diffusion processes (Fokker-Planck equations) that each specify a vector field in that space such that the given trajectory is a flow line.

This picture carries over to the filtering problem. In the (nonlinear) filtering problem with fixed initial distribution $p_0$
and fixed observation history, under reasonable assumptions the filtering equation specifies a unique trajectory \((p_t)_{t \geq 0}\) in the space of distributions. Using the idea of robust filtering (i.e. approximating the observation process by a smooth process, see [22]), this trajectory can be approximated by a smooth curve in the space of distributions. Constructing a particle filter corresponds to finding a diffusion process such that this smooth trajectory is a solution of the corresponding Fokker-Planck equation, i.e. a flow line of the corresponding vector field.

In contrast, the approach in [9] and [10] matches the vector field of the particle dynamics to the vector field of the filtering equation at each point (in the space of distributions), which leads to a unique control law. Modifying the control law in ways that do not change the vector field in a certain family of distributions may change it outside of the family. For example, in the linear-Gaussian case we modified the control law such that it did not affect the vector field on certain Gaussian distributions (those for which \(E_t\) has zero mean).

Even though the choice of a suitable family of distributions on which to fix the vector field might be difficult, the same reasoning can be applied in principle to the nonlinear case. After fixing an initial distribution \(p_0\), the filtering equation yields a unique trajectory \((p_t)_{t \geq 0}\) in the space of distributions. One may use the stochastic feedback particle filter \(S_t\) from [9] and [10] to approximate this trajectory. For each \(t \geq 0\), one can then apply a transformation \(\Phi_t\), e.g. from the family of diffeomorphisms that preserve the distribution \(p_t\). One can choose the function \(\Phi_t\) to be regular. The transformed particle filter \(S_t = \Phi_t(S_t)\) will still approximate \((p_t)_{t \geq 0}\).

The question remains what consequences arise as a result of the unconstrained changes of the vector field outside of the specified family. It is reasonable to suspect that certain filters constructed in this manner lack the intrinsic stability of the filtering equation; there might be initial distributions for which the distribution of the modified particle filter deviates significantly and irrecoverably from the exact conditional distribution. It remains for future research to better understand this mechanism and to devise criteria that can distinguish between stable and unstable filters.

VI. ACKNOWLEDGEMENTS

We thank Jean-Pascal Pfister, Anna Kutschireiter, Amirhossein Taghvai, and Prashant Mehta for helpful discussions.

REFERENCES

[1] A. Bain and D. Crisan, Fundamentals of stochastic filtering, ser. Stochastic Modelling and Applied Probability. New York, NY: Springer, New York, 2009, vol. 60.

[2] A. Doucet and A. M. Johansen, “A tutorial on particle filtering and smoothing: Fifteen years later,” in The Oxford Handbook of Nonlinear Filtering. Oxford University Press, Feb. 2011, pp. 656–704.

[3] F. E. Daum and J. Huang, Curve of dimensionality and particle filters. Aerospace Conference, 2003.

[4] 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, Dec. 2008.

[5] P. Rebeschini and R. van Handel, “Can local particle filters beat the curse of dimensionality?” Annals of Applied Probability, vol. 25, no. 5, pp. 2809–2866, 2015.

[6] 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 Review, vol. 61, no. 1, pp. 79–91, Feb. 2019.

[7] G. Evensen, “The Ensemble Kalman Filter: theoretical formulation and practical implementation,” Ocean Dynamics, vol. 53, no. 4, pp. 343–367, 2003.

[8] K. Bergemann and S. Reich, “An ensemble Kalman-Bucy filter for continuous data assimilation,” Meteorologische Zeitschrift, 2012.

[9] T. Yang, P. G. Mehta, and S. P. Meyn, “Feedback particle filter,” IEEE Transactions on Automatic Control, vol. 58, no. 10, pp. 2465–2480, 2013.

[10] T. Yang, R. S. Laugesen, P. G. Mehta, and S. P. Meyn, “Multivariable feedback particle filter,” Automatica, 2016.

[11] F. E. Daum and J. Huang, “Generalized particle flow for nonlinear filters,” in SPIE Defense, Security, and Sensing, O. E. Drummond, Ed. SPIE, Apr. 2010, pp. 769801–769808–12.

[12] R. S. Laugesen, P. G. Mehta, S. P. Meyn, and M. Raginsky, “Poisson’s equation in nonlinear filtering,” arXiv.org, Dec. 2014.

[13] F. E. Daum and J. Huang, “A plethora of open problems in particle flow research for nonlinear filters, Bayesian decisions, Bayesian learning, and transport,” in SPIE Defense + Security, I. Kadar, Ed. SPIE, May 2016, pp. 984201–3.

[14] P. Del Moral and J. Tugaut, “On the stability and the uniform propagation of chaos properties of Ensemble Kalman–Bucy filters,” Annals of Applied Probability, vol. 28, no. 2, pp. 790–850, Apr. 2018.

[15] A. Täghvai and P. G. Mehta, “An optimal transport formulation of the linear feedback particle filter,” 2016 American Control Conference (ACC), pp. 3614–3619, Jul. 2016.

[16] J.-W. Kim, A. Täghvai, and P. G. Mehta, “Derivation and Extensions of the Linear Feedback Particle Filter Based on Duality Formalisms,” 2018 IEEE Conference on Decision and Control (CDC), pp. 7188–7193, Dec. 2018.

[17] R. E. Kalman and R. S. Bucy, “New results in linear filtering and prediction theory,” Journal of Basic Engineering, vol. 83, no. 1, pp. 95–108, 1961.

[18] L. Dieci and T. Eirola, “Positive definiteness in the numerical solution of Riccati differential equations,” Numerische Mathematik, vol. 67, no. 3, pp. 303–313, Apr. 1994.

[19] D. Crisan, P. Del Moral, and T. J. Lyons, “Interacting particle systems approximations of the Kushner-Stratonovich equation,” Advances in Applied Probability, pp. 819–838, 1999.

[20] J. W. Kim, A. Täghvai, and P. G. Mehta, “Derivation and Extensions of the Linear Feedback Particle Filter based on Duality Formalisms,” arXiv.org, Apr. 2018.

[21] J. Mandel, L. Cobb, and J. D. Beezley, “On the convergence of the ensemble Kalman filter,” Applications of Mathematics, vol. 56, no. 6, pp. 533–541, Jan. 2012.

[22] J. M. C. Clark and D. Crisan, “On a robust version of the integral representation formula of nonlinear filtering,” Probability Theory and Related Fields, vol. 133, no. 1, pp. 43–56, Feb. 2005.