Particle Filtering with Invertible Particle Flow

Yunpeng Li, *Student Member, IEEE,* and Mark Coates, *Senior Member, IEEE*

Abstract—A key challenge when designing particle filters in high-dimensional state spaces is the construction of a proposal distribution that is close to the posterior distribution. Recent advances in particle flow filters provide a promising avenue to avoid weight degeneracy; particles drawn from the prior distribution are migrated in the state-space to the posterior distribution by solving partial differential equations. Numerous particle flow filters have been proposed based on different assumptions concerning the flow dynamics. Approximations are needed in the implementation of all of these filters; as a result the particles do not exactly match a sample drawn from the desired posterior distribution. Past efforts to correct the discrepancies involve expensive calculations of importance weights. In this paper, we present new filters which incorporate deterministic particle flows into an encompassing particle filter framework. The valuable theoretical guarantees concerning particle filter performance still apply, but we can exploit the attractive performance of the particle flow methods. The filters we describe involve a computationally efficient weight update step, arising because the embedded particle flows we design possess an invertible mapping property. We evaluate the proposed particle flow particle filters' performance through numerical simulations of a challenging multi-target multi-sensor tracking scenario and complex high-dimensional filtering examples.

Index Terms—Sequential Monte Carlo, Particle Flow, High-dimensional Filtering, Optimal Proposal Distribution.

I. INTRODUCTION

Particle filters are a family of Monte Carlo algorithms developed to solve the filtering problem of sequentially estimating a state variable. Particles (state samples) and their associated weights are advanced through time to approximate the filtering distributions of interest. The bootstrap particle filter (BPF) draws particles from the prior distribution and updates the weight of each particle using the likelihood of the latest measurement [1]. When the state dimension is high or when measurements are highly informative, the majority of particles drawn from the prior distribution will be in regions with very low likelihood, leading to negligible weights for most particles in the BPF. As the Monte Carlo approximation of the posterior distribution is dominated by a few particles, this weight degeneracy issue results in a poor representation of the posterior distribution [2], [3].

The "optimal" proposal distribution minimizes the variance of the importance weights [4] but is rarely possible to sample from. More advanced particle filters construct efficient proposal distributions by approximating the optimal proposal distribution [4], [5]. The auxiliary particle filter (APF) [6] introduces an auxiliary variable to more effectively sample particles, taking information from the new measurement into account. The Rao-Blackwellised particle filters [7] reduce the

Y. Li and M. Coates are with the Department of Electrical and Computer Engineering, McGill University, Montréal, Québec, Canada, e-mail: yunpeng.li@mail.mcgill.ca; mark.coates@mcgill.ca.

variance of Monte Carlo estimates by marginalizing some states analytically. The unscented particle filter [8] approximates the optimal proposal distribution using the unscented transformation. Although these approaches can be effective in some settings, particle filtering in high-dimensional spaces remains a challenging task and most conventional particle filters perform poorly [9]–[11]. Several directions have been explored in order to address the challenge in high-dimensional filtering and in settings where the measurements are highly informative. We provide a more detailed discussion of the research contributions most related to our work in Section I-A. One method that is promising and exhibits good performance in very high dimensions is the equivalent weights particle filter [12], [13]. In its basic form, it sacrifices the statistical consistency of the conventional particle filter but ensures that a large number of particles always have substantial weight, thus avoiding degeneracy. Another approach involves separating the state space through factorization or partitioning [14]–[17]. These techniques are promising, but rely on identifying a suitable factorization of the conditional posterior, so their applicability is restricted. A more general approach involves the incorporation of Markov Chain Monte Carlo (MCMC) methods within the particle filters [11], [18]–[26].

Although they can be effective in allowing particle filters to operate in high-dimensional state spaces, MCMC methods are almost always computationally expensive and their inclusion can render real time filtering impossible. An alternative set of methods, labelled "progressive Bayesian update", "homotopy" or "particle flow", can offer similar performance without the same computational requirements, but this comes at the cost of a more limited theoretical understanding. A framework for performing a progressive Bayesian update was introduced in [27]. In the context of particle filters, particles are "migrated" to represent the posterior distribution; the importance sampling step is eliminated. In a series of papers [28]–[37], Daum et al. link the log prior (or predictive posterior) and the log posterior distribution via a homotopy and derive partial differential equations (PDE) in order to guide particles to flow from the prior (or predictive posterior) towards the posterior distribution. Recently, Khan et al. [38] and de Melo et al. [39] have proposed alternative approaches. Although numerous such *particle flow filters* have been developed, most solutions are analytically intractable. One important exception is the "exact" particle flow [31] filter, developed under the assumption that the measurement model is linear and both the prior and posterior distributions are Gaussian.

For all particle flow filters, discretization is needed during implementation to numerically solve the derived PDEs. Particles advance through a large number of small steps. The truncation errors accumulated through the numerical integration steps, as well as approximations (Gaussianity, local linearity)

used in various other stages of implementation, can lead to particles deviating from the true posterior distribution.

An alternative perspective is to view the particle flow filters as a way to generate a proposal distribution close to the posterior for use by an encompassing particle filter. With this approach, we exploit the ability of particle flow filters to move particles into regions where the posterior is significant, but the extensive theoretical understanding of sampling-based particle filters still applies—convergence rates and large deviations bounds are available (see [40] for examples). Several filters adopting this strategy have been recently proposed [3], [39], [41]–[43]. Although elegant, most approaches are computationally very expensive.

In this paper, we present particle filtering algorithms that employ a modified deterministic particle flow approach to construct a proposal distribution that closely matches the target posterior. Our main contributions compared to the state-of-the-art work in this domain include: (i) we modify the particle flow procedures so that they constitute invertible mappings, which allow much more efficient importance weight evaluation than state-of-the-art particle filters that use particle flows; (ii) we provide the proof of the invertible mapping property of the designed particle flows; (iii) we demonstrate the performance of particle filters with invertible particle flows in very challenging filtering scenarios with either highly informative measurements or high-dimensional state space.

Preliminary results were published in abbreviated forms in conference papers [44], [45]. This paper provides a more detailed description of the proposed algorithms and related methods, a more complete performance evaluation with added high-dimensional filtering simulations, as well as the proof of the invertible mapping property.

The remainder of the paper is organized as follows. We provide a more detailed discussion of related work in Section I-A. We present the problem statement in Section II, followed by a brief review of particle flow methods in Section III. We describe the proposed particle flow particle filters and prove the invertible mapping property in Section IV. Section V details the simulation setup and presents results. The paper's contributions and results are summarized and discussed in Section VI.

A. Related Work

Over the past two decades, several directions have been explored for improving particle filtering performance in high-dimensional settings or when measurements are highly informative. One popular class of methods involve using the extended Kalman filter (EKF), the unscented Kalman filter (UKF) [46] or the ensemble Kalman filter (EnKF) [47] to generate the proposal distribution [8], [48]. These methods can be very effective for certain nonlinear models, but performance can deteriorate with non-Gaussian models. Other methods involve factorization or partitioning of the state space. In the *multiple particle filtering* approach [14], [15], particle filters are executed in parallel on low-dimensional subspaces that partition the full state space. The filters share information with one another to approximate filtering in the full state space. Rebeschini et al. proposed a similar strategy called the block particle filter [16], acknowledging that the blocking (partitioning) process introduces a bias that is difficult to quantify. The *space-time particle filters* [17] also rely on factorization of the conditional posterior, but Beskos et al. demonstrate that the filters provide consistent estimates as the number of particles increases. Although they are promising, the need to identify an effective factorization means that these algorithms are not generally applicable.

A different direction involves the incorporation of Markov Chain Monte Carlo (MCMC) methods within the particle filters. The first well-known technique in this category is the resample-move algorithm [19], which applies MCMC after the resampling step in order to diversify the particles. More recent techniques that adopt a similar approach include [23], [24]. These latter approaches modify the stochastic differential equation (SDE) which describes the state dynamics, in order to facilitate the MCMC moves and render them more efficient.

One problem associated with performing MCMC after resampling is that often very few particles are replicated after resampling in the high-dimensional space. Many MCMC iterations may then be needed to produce a set of particles that represent a reasonably independent sample from the target posterior. For this reason, Sequential Markov chain Monte Carlo (SMCMC) methods avoid the resampling step and use MCMC to generate samples directly from a proposal distribution [18], [22], [25], [26]. The identification of effective MCMC kernels often then becomes the major challenge. In [26], MCMC kernels based on Langevin diffusion or Hamiltonian dynamics have been proposed to more efficiently traverse a high-dimensional space. One of the most effective algorithms, referred to as the SmHMC algorithm, is based on the Manifold Hamiltonian Monte Carlo kernel [26]. One limitation of the SmHMC method is that it requires the target distribution to be log-concave. Solutions to this limitation are proposed in [26], but they require either analytically tractable expected values of negative Hessians or careful parameter tuning. In comparison, the particle flow particle filter we describe has a milder assumption on the smoothness of the measurement function.

Another way to mitigate weight degeneracy is to approach the true posterior density from the tractable prior density via intermediate densities, in a similar vein to simulated annealing. Such approaches were proposed in [20], [21]. Beskos et al. [11] illustrated how bridging densities and resampling could be used in conjunction with the sequential Monte Carlo sampler to improve the performance of particle filters in high dimensions. Theoretical results in [11] suggested that the approach successfully avoided the degeneracy and eliminated the need for exponential growth in the number of particles with respect to the state dimension.

Our work is most closely related to other techniques that combine particle flow or transport and particle filtering [3], [39], [41], [42]. The Gaussian particle flow importance sampling (GPFIS) algorithm developed in [3] uses approximate Gaussian flows to sample from non-Gaussian models. There is a non-trivial weight correction after every iteration of the particle flow and as a result the algorithm is computa-

tionally demanding. The particle flow particle filter (PF-PF) we describe performs an exact weight update instead of the approximate weight updates in the GPFIS [3]. The stochastic particle flow technique in [39] builds upon stationary solutions to the Fokker-Planck equation to compose Gaussian mixtures to approximate the posterior. The performance reported in [39] is impressive, but the computational requirements are significant. The optimal transport method-based approaches proposed in [41], [42] are similar in spirit to the particle flow approaches in that they involve particle transport (i.e., flow or migration) rather than importance sampling. The filter proposed in [41] incorporates weight correction after application of a transport map, so it inherits the theoretical properties of particle filters. The Gibbs flow approach in [42] requires numerical integrations of probability densities in state updates of each particle at each intermediate flow step, performed for each state dimension. So, the computational cost is very high, especially for high dimensional filtering scenarios.

The approach we propose may be interpreted as a guided sequential Monte Carlo (GSMC) method as outlined in [41]. The GSMC methods use deterministic transport maps to generate the proposal density. The importance weight evaluation step in the GSMC methods requires the calculation of the determinant of the Jacobian matrix of the transport map. This can be a major impediment for complex transport maps. As explained in [41], if we impose an extra assumption on the coupling that the transport map couples the prior and the posterior *exactly*, then the calculation of the determinant of the transport map can be avoided. Maps that achieve the exact coupling are identified in [41] for the special cases when the prior and posterior are Gaussians or mixtures of Gaussians; the maps can be implemented using the ensemble Kalman filter. In this paper, we effectively identify an alternative deterministic transport map using a modified exact Daum-Huang particle flow [31] and present complete routines to construct the proposal density using the modified invertible particle flow. The simple structure of the map allows efficient computation of the weight update. In one version of the proposed particle flow particle filter, we avoid evaluation of the determinant because the same map is applied to all particles; in the other we can evaluate it analytically for a relatively small computational overhead because of the structure of the map (the repeated application of affine transformations).

II. PROBLEM STATEMENT

The nonlinear filtering task we address involves tracking the marginal posterior distribution $p(x_k|z_{1:k})$, where x_k is the state of a system at time k and $z_{1:k} = \{z_1, \ldots, z_k\}$ is a sequence of measurements collected up to time step k. The state evolution and measurements are described by the following model:

$$x_0 \sim p_0(x) \,, \tag{1}$$
$$x_k = g_k(x_{k-1}, v_k) \quad \text{for } k \geq 1 \,, \tag{2}$$
$$z_k = h_k(x_k, w_k) \quad \text{for } k \geq 1 \,. \tag{3}$$

Here $p_0(x)$ is an initial probability density function, $g_k : \mathbb{R}^d \times \mathbb{R}^{d'} \to \mathbb{R}^d$ is the state-transition function of the unobserved

state $x_k \in \mathbb{R}^d$, $z_k \in \mathbb{R}^S$ is the measurement generated from the state x_k through a potentially nonlinear measurement model $h_k : \mathbb{R}^d \times \mathbb{R}^{S'} \to \mathbb{R}^S$. $v_k \in \mathbb{R}^{d'}$ is the process noise and $w_k \in \mathbb{R}^{S'}$ is the measurement noise. We assume that $g_k(\cdot, 0)$ is bounded, and $h_k(\cdot, 0)$ is a C^1 function [49], i.e. $h_k(\cdot, 0)$ is differentiable everywhere and its derivatives are continuous.

III. PARTICLE FLOWS

In this section we briefly review how deterministic particle flow can be used to address the nonlinear filtering problem. Suppose that we have a set of N_p particles $\{x_{k-1}^i\}_{i=1}^{N_p}$ approximating the posterior distribution at time $k-1$. After propagating particles using the dynamic model, we obtain particles $\{\tilde{x}_k^i\}_{i=1}^{N_p}$ that represent the predictive posterior distribution at time k. Particle flow is then used to migrate the particles so that they approximate the posterior distribution at time k.

We can model the particle flow as a background stochastic process η_λ in a pseudo time interval $\lambda \in [0, 1]$. To simplify notation, we do not include the time index k in the following description of the stochastic process, as the particle flow only concerns particle migration between two adjacent time steps. We denote by η_λ^i the stochastic process's i-th realization, and set $\eta_0^i = \tilde{x}_k^i$, for $i = 1, 2, \ldots, N_p$.

The *zero diffusion* particle flow filters [28]–[33] involve no random displacements of particles; the flows are deterministic. The trajectory of η_λ^i for realization i follows the ordinary differential equation (ODE):

$$\frac{d\eta_\lambda^i}{d\lambda} = \zeta(\eta_\lambda^i, \lambda) \,, \tag{4}$$

where $\zeta : \mathbb{R}^d \to \mathbb{R}^d$ is governed by the Fokker-Planck equation and additional flow constraints [32]. The Fokker-Planck equation with zero diffusion is given by

$$\frac{\partial p(\eta_\lambda^i, \lambda)}{\partial \lambda} = -\text{div}(p(\eta_\lambda^i, \lambda)\zeta(\eta_\lambda^i, \lambda))$$
$$= -p(\eta_\lambda^i, \lambda)\text{div}(\zeta(\eta_\lambda^i, \lambda)) - \frac{\partial p(\eta_\lambda^i, \lambda)}{\partial \eta_\lambda^i}\zeta(\eta_\lambda^i, \lambda) \,, \tag{5}$$

where $p(\eta_\lambda^i, \lambda)$ is the probability density of η_λ^i at time λ of the flow.

By imposing different constraints on the flow, Equation (5) can lead to a variety of particle flow filters. However, very few are analytically tractable. One exception is when the predictive posterior and the likelihood distributions are both Gaussian and the measurement model is linear, i.e., $\eta_0^i \sim N(\bar{\eta}_0, P)$, $z = h(\eta_\lambda^i, w) \sim N(H\eta_\lambda^i, R)$. The predictive covariance P and the measurement covariance R are both positive definite. H is called the measurement matrix. Since we drop the time index k in this section, we use z to denote the measurement available at time step k.

A. *The exact Daum and Huang filter*

The flow trajectory in the resultant exact Daum and Huang (EDH) filter [31] becomes:

$$\zeta(\eta_\lambda^i, \lambda) = A(\lambda)\eta_\lambda^i + b(\lambda) \,, \tag{6}$$

where

$$A(\lambda) = -\frac{1}{2}PH^T(\lambda HPH^T + R)^{-1}H, \quad (7)$$

$$b(\lambda) = (I + 2\lambda A(\lambda))[(I + \lambda A(\lambda))PH^TR^{-1}z + A(\lambda)\bar{\eta}_0]. \quad (8)$$

For nonlinear observation models, a linearization of the model is performed at the mean of the intermediate distribution, $\bar{\eta}_\lambda$, to construct $H(\lambda)$:

$$H(\lambda) = \left.\frac{\partial h(\eta,0)}{\partial \eta}\right|_{\eta=\bar{\eta}_\lambda}. \quad (9)$$

This is the $S \times d$ Jacobian matrix evaluated at $\bar{\eta}_\lambda$. A slight change must be made to the expressions of the flow, as presented in [50]:

$$A(\lambda) = -\frac{1}{2}PH(\lambda)^T(\lambda H(\lambda)PH(\lambda)^T + R)^{-1}H(\lambda), \quad (10)$$

$$b(\lambda) =(I + 2\lambda A(\lambda))[(I + \lambda A(\lambda))PH(\lambda)^TR^{-1}(z - e(\lambda)) + A(\lambda)\bar{\eta}_0], \quad (11)$$

where $e(\lambda) = h(\bar{\eta}_\lambda, 0) - H(\lambda)\bar{\eta}_\lambda$.

B. The localized exact Daum and Huang filter

The localized exact Daum and Huang filter (LEDH) [50] linearizes the system and updates the drift term for each individual particle. For the i-th particle, the drift term

$$\zeta(\eta_\lambda^i, \lambda) = A^i(\lambda)\eta_\lambda^i + b^i(\lambda), \quad (12)$$

where

$$A^i(\lambda) = -\frac{1}{2}PH^i(\lambda)^T(\lambda H^i(\lambda)PH^i(\lambda)^T + R)^{-1}H^i(\lambda), \quad (13)$$

$$b^i(\lambda) =(I + 2\lambda A^i(\lambda))[(I + \lambda A^i(\lambda))PH^i(\lambda)^TR^{-1}(z - e^i(\lambda)) + A^i(\lambda)\bar{\eta}_0]. \quad (14)$$

Here $H^i(\lambda) = \left.\frac{\partial h(\eta,0)}{\partial \eta}\right|_{\eta=\eta_\lambda^i}$ and $e^i(\lambda) = h(\eta_\lambda^i, 0) - H^i(\lambda)\eta_\lambda^i$.

C. Numerical Implementation

In the implementation of the exact particle flow algorithms, discretized pseudo-time integration is used to approximate the solution to the ODE. Suppose that a sequence of discrete steps are taken at N_λ positions $[\lambda_1, \lambda_2, \ldots, \lambda_{N_\lambda}]$, where $0 = \lambda_0 < \lambda_1 < \ldots < \lambda_{N_\lambda} = 1$. The step size $\epsilon_j = \lambda_j - \lambda_{j-1}$ for $j = 1, \ldots, N_\lambda$ can be possibly varying, and we require that $\sum_{j=1}^{N_\lambda} \epsilon_j = \lambda_{N_\lambda} - \lambda_0 = 1$.

The integral between λ_{j-1} and λ_j for $1 \le j \le N_\lambda$ is approximated and the functional mapping for the EDH flow becomes

$$\eta_{\lambda_j}^i = f_{\lambda_j}(\eta_{\lambda_{j-1}}^i) = \eta_{\lambda_{j-1}}^i + \epsilon_j(A(\lambda_j)\eta_{\lambda_{j-1}}^i + b(\lambda_j)). \quad (15)$$

The linearization of $H(\lambda_j)$ subsequently used to update $A(\lambda_j)$ is performed at $\bar{\eta}_{\lambda_{j-1}}$, which is the average of all particles $\{\eta_{\lambda_{j-1}}^i\}$ at time step λ_{j-1}.

For the LEDH, the functional mapping is

$$\eta_{\lambda_j}^i = f_{\lambda_j}^i(\eta_{\lambda_{j-1}}^i) = \eta_{\lambda_{j-1}}^i + \epsilon_j(A^i(\lambda_j)\eta_{\lambda_{j-1}}^i + b^i(\lambda_j)). \quad (16)$$

In the LEDH, the linearization of $H^i(\lambda_j)$ needed to update $A^i(\lambda_j)$ is performed at $\eta_{\lambda_{j-1}}^i$.

IV. PARTICLE FLOW WITH INVERTIBLE MAPPING

In the particle flow particle filtering framework we propose in this paper, the migrated particle η_1^i after the particle flow process is viewed as being drawn from a proposal distribution $q(\eta_1^i|x_{k-1}^i, z_k)$. In general, we cannot evaluate this proposal distribution, due to approximations in the filter implementation and the mismatch between the model assumptions of the embedded particle flow filter and the real scenario.

However, if the flow process defines an invertible deterministic mapping $\eta_1^i = T(\eta_0^i; z_k, x_{k-1}^i)$ between the i-th particle value before and after the flow, we can evaluate the proposal density as follows:

$$q(\eta_1^i|x_{k-1}^i, z_k) = \frac{p(\eta_0^i|x_{k-1}^i, z_k)}{|\dot{T}(\eta_0^i; z_k, x_{k-1}^i)|} = \frac{p(\eta_0^i|x_{k-1}^i)}{|\dot{T}(\eta_0^i; z_k, x_{k-1}^i)|}, \quad (17)$$

where $\dot{T}(\cdot) \in \mathbb{R}^{d \times d}$ is the Jacobian determinant of the mapping function $T(\cdot)$ for the i-th particle and $|\cdot|$ denotes the absolute value. The mapping can be different for each particle and can depend only on the measurements $z_{1:k}$ and the i-th particle's historical state values $x_{0:k-1}^i$, i.e. it cannot depend on the state values of other particles. We choose to restrict the dependence to the current measurement z_k and the previous state value x_{k-1}^i. The first equality of (17) is due to the invertible mapping between η_0^i and η_1^i. The second holds because η_0^i is generated solely through the dynamic model.

We can then evaluate the importance weight of each particle at time step k as:

$$w_k^i \propto \frac{p(\eta_1^i|x_{k-1}^i)p(z_k|\eta_1^i)|\dot{T}(\eta_0^i; z_k, x_{k-1}^i)|}{p(\eta_0^i|x_{k-1}^i)} w_{k-1}^i. \quad (18)$$

As noted in [3], the state update during the particle flow process is not in general an invertible mapping, so (18) does not hold. This motivated the development of complicated weight update procedures in [3] to approximate the importance weights. In this section, we propose modified particle flow procedures that possess the invertible mapping property, which allows us to perform efficient weight updates using (18).

A. Particle Flow Particle Filtering with the LEDH flow

The particle flow particle filter algorithm (PF-PF) based on the LEDH flow is presented in Algorithm 1. We show below that, under certain conditions, the function constructed by the

discretized particle flow in lines 11-21 of the algorithm leads to an invertible mapping. For the PF-PF (LEDH), we have:

$$
\begin{aligned}
|\dot{T}(\eta_0^i; z_k, x_{k-1}^i)| &= |\det(\frac{d\eta_1^i}{d\eta_0^i})| \\
&= |\det(\frac{d[\prod_{j=1}^{N_\lambda}(I + \epsilon_j A_j^i(\lambda))]\eta_0^i}{d\eta_0^i})| \\
&= \prod_{j=1}^{N_\lambda} |\det(I + \epsilon_j A_j^i(\lambda))| .
\end{aligned}
\tag{19}
$$

where $\det(\cdot)$ denotes the determinant. We prove at the end of Section IV-A that $0 < |\dot{T}(\eta_0^i; z_k, x_{k-1}^i)| < \infty$. Thus, the weight update expression for PF-PF (LEDH) is

$$
w_k^i \propto \frac{p(\eta_1^i|x_{k-1}^i)p(z_k|\eta_1^i)\prod_{j=1}^{N_\lambda}|\det(I + \epsilon_j A_j^i(\lambda))|}{p(\eta_0^i|x_{k-1}^i)}w_{k-1}^i
\tag{20}
$$

The particle flow procedure for each particle requires a predicted covariance estimate, i.e., the covariance matrix of the predictive posterior. The predicted covariance P can be obtained by using the Kalman covariance equations. The Kalman prediction step requires an estimated posterior covariance step in the previous time step, which can be estimated through the Kalman update step. When the dynamic model does not match the linear Gaussian scenario, the extended Kalman filter (EKF) or the unscented Kalman filter (UKF) [46] covariance prediction equations can be applied to estimate P.

We now prove that the function constructed by the discretized particle flow in Algorithm 1, $\eta_1^i = T(\eta_0^i; z_k, x_{k-1}^i)$, is invertible. We start with the following lemma.

Lemma IV.1. *For any $\lambda \in [0,1)$, if $h(\cdot, 0)$ is a C^1 function and $\bar{\eta}_\lambda^i$ is bounded, $\rho(A^i(\lambda))$ of $A^i(\lambda)$ defined by (13) is upper-bounded. Here $\rho(\cdot)$ denotes the spectral radius.*

Proof: Denote the largest eigenvalue of P by $\bar{p}$, and the smallest eigenvalue of $(\lambda H^i(\lambda)^T P H^i(\lambda) + R)$ by $\underline{r}(\lambda)$. Since P is positive definite, $\bar{p} > 0$. Since R is positive definite, for any non-zero $\mu \in \mathbb{R}^d$,

$$
\begin{aligned}
\mu^T &\left(\lambda H^i(\lambda)^T P H^i(\lambda) + R\right)\mu \\
&= \lambda \left(H^i(\lambda)\mu\right)^T P \left(H^i(\lambda)\mu\right) + \mu^T R\mu \\
&> 0 .
\end{aligned}
\tag{21}
$$

Thus, $(\lambda H^i(\lambda)^T P H^i(\lambda) + R)$ is positive definite. So, $\underline{r}(\lambda) > 0$.

Denote the operator norm induced by the Euclidean norm by $||\cdot||$. Since P and $\left(\lambda H^i(\lambda)^T P H^i(\lambda) + R\right)^{-1}$ are both positive semi-definite, $||P||$ and $||(\lambda H^i(\lambda)^T P H^i(\lambda) + R)^{-1}||$ are equal to the spectral radius, and we have

$$
||P|| = \bar{p} ,
\tag{22}
$$

$$
||(\lambda H^i(\lambda)^T P H^i(\lambda) + R)^{-1}|| = \frac{1}{\underline{r}(\lambda)} .
\tag{23}
$$

Algorithm 1: Particle flow particle filtering (LEDH).

1: Initialization: Draw $\{x_0^i\}_{i=1}^{N_p}$ from the prior $p_0(x)$. Set $\hat{x}_0$ and $\hat{P}_0$ to be the mean and covariance of $p_0(x)$, respectively;
2: Set $\{w_0^i\}_{i=1}^{N_p} = \frac{1}{N_p}$;
3: **for** $k = 1$ to K **do**
4:    **for** $i = 1, \dots, N_p$ **do**
5:       Apply EKF/UKF prediction to estimate P^i: $(x_{k-1}^i, P_{k-1}^i) \rightarrow (m_{k|k-1}^i, P^i)$;
6:       Calculate $\bar{\eta}^i = g_k(x_{k-1}^i, 0)$;
7:       Propagate particles $\eta_0^i = g_k(x_{k-1}^i, v_k)$;
8:       Set $\eta_1^i = \eta_0^i$ and $\theta^i = 1$;
9:       Calculate $\bar{\eta}_0^i = g_k(x_{k-1}^i, 0)$;
10:    **end for**
11:    Set $\lambda = 0$;
12:    **for** $j = 1, \dots, N_\lambda$ **do**
13:       Set $\lambda = \lambda + \epsilon_j$;
14:       **for** $i = 1, \dots, N_p$ **do**
15:          Set $\bar{\eta}_0 = \bar{\eta}_0^i$ and $P = P^i$;
16:          Calculate $A_j^i(\lambda)$ and $b_j^i(\lambda)$ from (13) and (14) with the linearization being performed at $\bar{\eta}^i$;
17:          Migrate $\bar{\eta}^i$: $\bar{\eta}^i = \bar{\eta}^i + \epsilon_j(A_j^i(\lambda)\bar{\eta}^i + b_j^i(\lambda))$;
18:          Migrate particles: $\eta_1^i = \eta_1^i + \epsilon_j(A_j^i(\lambda)\eta_1^i + b_j^i(\lambda))$;
19:          Calculate $\theta^i = \theta^i|\det(I + \epsilon_j A_j^i(\lambda))|$
20:       **end for**
21:    **end for**
22:    **for** $i = 1, \dots, N_p$ **do**
23:       Set $x_k^i = \eta_1^i$;
24:       $w_k^i = \frac{p(x_k^i|x_{k-1}^i)p(z_k|x_k^i)\theta^i}{p(\eta_0^i|x_{k-1}^i)}w_{k-1}^i$;
25:    **end for**
26:    **for** $i = 1, \dots, N_p$ **do**
27:       Normalize $w_k^i = w_k^i / \sum_{s=1}^{N_p} w_k^s$;
28:       Apply EKF/UKF update: $(m_{k|k-1}^i, P^i) \rightarrow (m_{k|k}^i, P_k^i)$;
29:    **end for**
30:    Estimate $\hat{x}_k = \sum_{i=1}^{N_p} w_k^i x_k^i$;
31:    (Optional) Resample $\{x_k^i, P_k^i, w_k^i\}_{i=1}^{N_p}$ to obtain $\{x_k^i, P_k^i, \frac{1}{N_p}\}_{i=1}^{N_p}$;
32: **end for**

We also have

$$
\begin{aligned}
||H^i(\lambda)|| &= \sqrt{\rho_{\max}(H^i(\lambda)^T H^i(\lambda))} \\
&\leq \sqrt{\text{Tr}(H^i(\lambda)^T H^i(\lambda))} \\
&= ||H^i(\lambda)||_F ,
\end{aligned}
\tag{24}
$$

where $\text{Tr}(\cdot)$ denotes the trace of a matrix, and $||\cdot||_F$ is the Frobenius norm. Similarly, $||H^i(\lambda)^T|| \leq ||H^i(\lambda)||_F$.

Since $h(\cdot)$ is a C^1 function, $H^i(\lambda)$ is continuous on $\bar{\eta}_\lambda^i \in \mathbb{R}^d$. Since $\bar{\eta}_\lambda^i$ is bounded on $\lambda \in [0,1]$, $H^i(\lambda)$ is bounded on $\lambda \in [0,1]$. Thus, there exists an $\bar{h} > 0$ such that $||H^i(\lambda)||_F \leq \bar{h}$ for any $\lambda \in [0,1]$.

For the square matrix $A^i(\lambda)$, its spectral radius is upper-bounded by its operator norm. Thus, by the sub-

multiplicativity of the operator norm,

$$
\begin{aligned}
\rho(A^i(\lambda)) &\leq ||A^i(\lambda)|| \\
&= || - \frac{1}{2} P H^i(\lambda)^T (\lambda H^i(\lambda) P H^i(\lambda)^T + R)^{-1} H^i(\lambda)|| \\
&\leq \frac{1}{2} ||P|| \cdot ||H^i(\lambda)^T|| \cdot ||(\lambda H^i(\lambda)^T P H^i(\lambda) + R)^{-1}|| \\
&\quad \cdot ||H^i(\lambda)|| \\
&\leq \frac{\bar{p}\bar{h}^2}{2\underline{r}(\lambda)} \, .
\end{aligned}
\tag{25}
$$

We now prove that with a sufficiently small step size, the mapping defined by (10), (11) and (15) is invertible.

Lemma IV.2. *For any* $\lambda_j \in [0, 1)$, *if* $h(\cdot, 0)$ *is a* C^1 *function and* $\bar{\eta}^i_{\lambda_j}$ *is bounded,* $f^i_{\lambda_j}$ *defined by* (13), (14) *and* (16) *is invertible, if* ϵ_j *is sufficiently small, specifically* $\epsilon_j < \dfrac{2\underline{r}(\lambda_j)}{\bar{p}\bar{h}^2}$.

Proof: For any $i \in \{1, \dots, N_p\}$ and $\lambda_j \in [0, 1)$, consider two values of $\eta^i_{\lambda_j}$, $\eta \neq \eta'$, Since $A^i(\lambda_j)$ and $b^i(\lambda_j)$ are the same for $\eta \neq \eta'$, from (12) and (16),

$$
f^i_{\lambda_j}(\eta) = \eta + \epsilon_j (A^i(\lambda_j)\eta + b^i(\lambda_j)) \, , \tag{26}
$$

$$
f^i_{\lambda_j}(\eta') = \eta' + \epsilon_j (A^i(\lambda_j)\eta' + b^i(\lambda_j)) \, . \tag{27}
$$

If $f^i_{\lambda_j}(\eta) = f^i_{\lambda_j}(\eta')$, then

$$
\eta - \eta' = -\epsilon_j A^i(\lambda_j)(\eta - \eta') \, . \tag{28}
$$

Equation (28) holds only if $\eta - \eta'$ is an eigenvector of $A^i(\lambda_j)$ and its corresponding eigenvalue ψ satisfies $\epsilon_j \psi = -1$. From Lemma IV.1, $\rho_{\max}(A^i(\lambda_j))$ is upper-bounded for $\lambda_j \in [0, 1)$. Thus $|\psi| \leq \dfrac{\bar{p}\bar{h}^2}{2\underline{r}(\lambda_j)}$. If we choose $\epsilon_j < \dfrac{2\underline{r}(\lambda_j)}{\bar{p}\bar{h}^2}$ then $\epsilon_j |\psi| < 1$ and $f^i_{\lambda_j}(\eta) \neq f^i_{\lambda_j}(\eta')$, implying that $f^i_{\lambda_j}$ is injective.

For this choice of ϵ_j, the equality (28) does not hold unless $\eta \neq \eta'$. Thus

$$
(I + \epsilon_j A^i(\lambda_j))(\eta - \eta') = 0 \, , \tag{29}
$$

only holds for $\eta = \eta'$, demonstrating that $\text{Null}(I + \epsilon_j A^i(\lambda_j)) = \{0\}$. Hence,

$$
\begin{aligned}
\dim(\text{range}(I + \epsilon_j A^i(\lambda_j))) &= d - \dim(\text{Null}(I + \epsilon_j A^i(\lambda_j))) \\
&= d
\end{aligned}
\tag{30}
$$

Since $\text{range}(I + \epsilon_j A^i(\lambda_j))$ is a subspace of $\mathbb{R}^d$,

$$
\text{range}(I + \epsilon_j A^i(\lambda_j)) = \mathbb{R}^d \, . \tag{31}
$$

Thus, $(I + \epsilon_j A^i(\lambda_j))$ has full rank and the mapping $f^i_{\lambda_j}(\eta^i_{\lambda_j}) = (I + \epsilon_j A^i(\lambda_j))\eta^i_{\lambda_j} + b^i(\lambda_j)$ is surjective.

Thus for the specified choice of ϵ_j, $f^i_{\lambda_j}$ is both injective and surjective, and hence invertible. ∎

Now we can establish the following theorem:

Theorem IV.3. *If* $h(\cdot, 0)$ *is a* C^1 *function,* $\eta^i_1 = T(\eta^i_0; z_k, x^i_{k-1})$ *in the particle flow particle filter with the LEDH flow defines an invertible mapping, if* $\epsilon_j < \dfrac{2\underline{r}(\lambda_j)}{\bar{p}\bar{h}^2}$ *for* $j \in \{1, \dots, N_\lambda\}$.

Proof: The theorem follows directly from Lemma IV.2 if $\bar{\eta}^i_{\lambda_j}$ is bounded for $j \in \{1, \dots, N_\lambda\}$. We prove this by induction. For $j = 1$, $\bar{\eta}^i_{\lambda_1} = \bar{\eta}^i_0$ is bounded as it is generated by propagating the sample mean of particles in the previous time step using $g(\cdot, 0)$ which is a bounded function.

For $j \in \{1, \dots, N_\lambda - 1\}$ assume it is true that $\bar{\eta}^i_{\lambda_j}$ is bounded. For the specified choice of ϵ_j, from Lemma IV.2, $f^i_{\lambda_j}$ is invertible. Thus, $\bar{\eta}^i_{\lambda_{j+1}} = f^i_{\lambda_j}(\bar{\eta}^i_{\lambda_j})$ is bounded.

By induction, $\bar{\eta}^i_{\lambda_j}$ is bounded for all $j \in \{1, \dots, N_\lambda\}$. This implies from Lemma IV.2 that $f^i_{\lambda_j}$ is invertible for $j \in \{1, \dots, N_\lambda\}$. Since the deterministic mapping

$$
\begin{aligned}
\eta^i_1 &= T(\eta^i_0; z_k, x^i_{k-1}) \\
&= f^i_{\lambda_{N_\lambda}}(\dots f^i_{\lambda_1}(\eta^i_0))
\end{aligned}
\tag{32}
$$

is a chain of invertible mappings, $T(\eta^i_0)$ is an invertible mapping. ∎

We now prove that $0 < |\dot{T}(\eta^i_0; z_k, x^i_{k-1})| < \infty$ for the specified choice of ϵ_j, which shows that the importance weight is finite for all η^i_1 generated by applying the constructed mapping $T(\eta^i_0)$.

Lemma IV.4. *If* $\epsilon_j < \dfrac{2\underline{r}(\lambda_j)}{\bar{p}\bar{h}^2}$ *for* $j \in \{1, \dots, N_\lambda\}$, *then* $0 < |\dot{T}(\eta^i_0; z_k, x^i_{k-1})| < \infty$.

Proof: From (31), we see that $(I + \epsilon_j A^i_j(\lambda))$ is invertible, $\forall j \in \{1, \dots, N_\lambda\}$. Thus, $\det(I + \epsilon_j A^i_j(\lambda)) \neq 0, \forall j \in \{1, \dots, N_\lambda\}$. From (19),

$$
|\dot{T}(\eta^i_0; z_k, x^i_{k-1})| = \prod_{j=1}^{N_\lambda} |\det(I + \epsilon_j A^i_j(\lambda))| > 0 \, . \tag{33}
$$

We denote the m-th column of $(I + \epsilon_j A^i_j(\lambda))$ by $u^m_j \in \mathbb{R}^d$, for $m \in \{1, \dots, d\}$. From Hadamard's inequality,

$$
\det(I + \epsilon_j A^i_j(\lambda)) \leq \prod_{m=1}^{d} ||u^m_j||_2 \, , \tag{34}
$$

where $||\cdot||_2$ is the Euclidean norm. Thus,

$$
\begin{aligned}
|\dot{T}(\eta^i_0; z_k, x^i_{k-1})| &= \prod_{j=1}^{N_\lambda} |\det(I + \epsilon_j A^i_j(\lambda))| \\
&\leq \prod_{j=1}^{N_\lambda} \prod_{m=1}^{d} ||u^m_j||_2 < \infty \, .
\end{aligned}
\tag{35}
$$
∎

B. Particle Flow Particle Filtering with the EDH flow

The particle flow particle filter algorithm based on the EDH flow with the invertible mapping property is presented in Algorithm 2. The PF-PF (EDH) is much more computationally efficient that the PF-PF (LEDH) by using common flow parameters $A_j(\lambda)$ and $b_j(\lambda)$ to perform flows for different particles. But this leads to statistical correlations between particles and the effect of the dependence can be challenging to characterize. Hence, the PF-PF (EDH) reduces the computational cost with a compromise on convergence properties of standard particle filters [40].

The mapping defined by (10) and (11) is the same as that defined by (13) and (14), if we replace $\bar{\eta}^i$ by $\bar{\eta}$ in Algorithm 1. From Theorem IV.3, this defines an invertible mapping between η_0^i and η_1^i for any i. Thus, the weight update equation (18) still holds for Algorithm 2.

Algorithm 2: Particle flow particle filtering (EDH).

1: Initialization: Draw $\{x_0^i\}_{i=1}^{N_p}$ from $p_0(x)$. Set $\hat{x}_0$ and $\hat{P}_0$ to be the mean and covariance of $p_0(x)$, respectively;
2: Set $\{w_0^i\}_{i=1}^{N_p} = \frac{1}{N_p}$;
3: **for** k = 1 to T **do**
4:     Apply EKF/UKF prediction to estimate P: $(\hat{x}_{k-1}, P_{k-1}) \rightarrow (m_{k|k-1}, P)$;
5:     **for** $i = 1, \ldots, N_p$ **do**
6:         Propagate particles $\eta_0^i = g_k(x_{k-1}^i, v_k)$;
7:         Set $\eta_1^i = \eta_0^i$;
8:     **end for**
9:     Calculate $\bar{\eta}_0 = g_k(\hat{x}_{k-1}, 0)$ and set $\bar{\eta} = \bar{\eta}_0$;
10:    Set $\lambda = 0$;
11:    **for** $j = 1, \ldots, N_\lambda$ **do**
12:       Set $\lambda = \lambda + \epsilon_j$;
13:       Calculate $A_j(\lambda)$ and $b_j(\lambda)$ from Equation (10) and (11) with the linearization being performed at $\bar{\eta}$;
14:       Migrate $\bar{\eta}$: $\bar{\eta} = \bar{\eta} + \epsilon_j(A_j(\lambda)\bar{\eta} + b_j(\lambda))$;
15:       **for** $i = 1, \ldots, N_p$ **do**
16:         Migrate particles: $\eta_1^i = \eta_1^i + \epsilon_j(A_j(\lambda)\eta_1^i + b_j(\lambda))$;
17:       **end for**
18:    **end for**
19:    **for** $i = 1, \ldots, N_p$ **do**
20:       Set $x_k^i = \eta_1^i$;
21:       $w_k^i = \frac{p(x_k^i|x_{k-1}^i)p(z_k|x_k^i)}{p(\eta_0^i|x_{k-1}^i)} w_{k-1}^i$;
22:    **end for**
23:    **for** $i = 1, \ldots, N_p$ **do**
24:       Normalize $w_k^i = w_k^i / \sum_{s=1}^{N_p} w_k^s$;
25:    **end for**
26:    Apply EKF/UKF update: $(m_{k|k-1}, P) \rightarrow (m_{k|k}, P_k)$;
27:    Estimate $\hat{x}_k = \sum_{i=1}^{N_p} w_k^i x_k^i$;
28:    (Optional) Resample $\{x_k^i, w_k^i\}_{i=1}^{N_p}$ to obtain $\{x_k^i, \frac{1}{N_p}\}_{i=1}^{N_p}$;
29: **end for**

For the PF-PF (EDH), we do not need to evaluate the Jacobian determinant. Based on the migration of particles shown in Line 16 of Algorithm 2,

$$
\begin{aligned}
|\dot{T}(\eta_0^i; z_k, x_{k-1}^i)| &= |\det(\frac{d\eta_1^i}{d\eta_0^i})| \\
&= |\det(\frac{d[\prod_{j=1}^{N_\lambda}(I + \epsilon_j A_j(\lambda))]\eta_0^i}{d\eta_0^i})| \\
&= \prod_{j=1}^{N_\lambda} |\det(I + \epsilon_j A_j(\lambda))|,
\end{aligned} \tag{36}
$$

where $\det(\cdot)$ denotes the determinant. From (36), we see that for the PF-PF (EDH), the determinant is the same for different particles, i.e., $|\dot{T}(\eta_0^i; z_k, x_{k-1}^i)| = |\dot{T}(\eta_0^{i'}; z_k, x_{k-1}^{i'})|$ for $i \neq i'$.

We can show that $0 < |\dot{T}(\eta_0^i; z_k, x_{k-1}^i)| < \infty$ using the same argument as in Lemma IV.4. Thus, the weight update expression for the PF-PF (EDH) is

$$
w_k^i \propto \frac{p(\eta_1^i|x_{k-1}^i)p(z_k|\eta_1^i)}{p(\eta_0^i|x_{k-1}^i)} w_{k-1}^i. \tag{37}
$$

C. Implementation and Complexity

Several numerical integration schemes are proposed and discussed in [35], [37], [51]. For algorithms involving particle flows, we adopt the $N_\lambda = 29$ exponentially spaced step sizes recommended in [35]. The constant ratio between step sizes is 1.2, i.e. $q = \frac{\epsilon_j}{\epsilon_{j-1}} = 1.2$, for $j = 2, 3, \ldots, N_\lambda$. The initial step size $\epsilon_1 = \frac{1-q}{1-q^{N_\lambda}} \approx 0.001$.

With the cost of increased computation, the eigenvalues of $A^i(\lambda)$ or $A(\lambda)$ can be evaluated and an adaptive step size used to ensure that the invertible mapping property is satisfied. In practice it is very unlikely that any of the pre-defined step sizes satisfies $\epsilon_j \psi = -1$ where ψ is any eigenvalue of $A^i(\lambda)$ or $A(\lambda)$. Denote by $\rho_{\max}(A)$ the largest magnitude of any eigenvalue of a matrix A. We have checked that the step-size choice described above leads to ϵ_j values that are always smaller than $\frac{1}{\rho_{\max}(A^i(\lambda))}$ or $\frac{1}{\rho_{\max}(A(\lambda))}$ in the simulation scenarios examined in Section V.

For the PF-PF (LEDH), the most computationally demanding part of the algorithm is the inverse operation in calculating $A^i(\lambda)$ and $b^i(\lambda)$. Since individual flow parameters are calculated for each particle, the computational complexity of the matrix inverse operations is $O(N_p S^3)$ (recall that S is the measurement dimension). There is an additional overhead in calculating the determinant in line 19 of Algorithm 1, but this is small compared to the inverse operations. Once the θ_i values in line 19 have been computed, the complexity of the weight update in line 24 is $O(N_p)$. The computational cost of the weight update is much lower than that of the GPFIS, which involves calculating matrix square roots and repeatedly solving the Sylvester equation.

For the PF-PF (EDH) introduced in Section IV-B, the most computationally intensive part of the flow is again the inversion operation in Equation (10) and (11), which has a computation complexity of $O(S^3)$. Since the calculation of the flow parameters is only performed at $\bar{\eta}$, the computational complexity of the inverse operation does not depend on the number of particles N_p. The weight update does not depend on the number of intermediate flow update steps N_λ as no determinant needs to be calculated to update the importance weight. The computational cost of the weight calculation is usually negligible compared to that of the flow; an exception is when the prior probability $p(\eta_1^i|x_{k-1}^i)$ is difficult to evaluate.

One possible concern with the proposed implementation is that flow parameters are calculated using the auxiliary flows by linearizing at $\bar{\eta}^i$ (LEDH) or $\bar{\eta}$ (EDH), as opposed to linearizing at the actual particle values $\{\eta^i\}$. We note that the linearization in the EDH is already performed at the mean or median of the particle cloud, so it is unlikely that using $\bar{\eta}$ introduces additional error for the PF-PF (EDH). For the PF-PF (LEDH), however, there is the potential for additional error beyond the linearization due to this mismatch. We have

compared the original LEDH filter with an LEDH filter that uses the auxiliary flows for flow parameter calculation using examples from Section V (for conciseness, these results are not shown in the paper). The constructed flows and the filtering performance were very similar, suggesting that the use of auxiliary flows introduces minimal error for these scenarios. This issue does warrant further exploration, and ideally a filter can be designed such that weight updates can be performed with small computational cost and the flows are calculated using the actual particle values.

V. SIMULATIONS AND RESULTS

We explore the performance of the PF-PF algorithms in two challenging simulation setups. The first is a multi-target acoustic tracking scenario with small measurement noise. The second is a high dimensional filtering problem in which the state evolves according to a multivariate Generalized Hyperbolic (GH) skewed-t distribution and the observations are count data derived via a Poisson process. Both scenarios lead to severe particle degeneracy for bootstrap particle filters due to either the highly informative measurements or the high dimensionality. In addition, we compare the performance of the PF-PF with the optimal filter in a simple linear Gaussian filtering example. Matlab code implementing the simulation experiments is available [1].

A. Multi-target acoustic tracking

1) Simulation setup: We constructed a multi-target tracking scenario with a relatively large state space and highly informative measurements, based on the simulation setup proposed in [52]. There are $C = 4$ targets moving independently in a region of size of 40 m $\times$40 m. Each follows a constant velocity model $x_k^{(c)} = Fx_{k-1}^{(c)} + v_k^{(c)}$, where $x_k^{(c)} = [\mathrm{x}_k^{(c)}, \mathrm{y}_k^{(c)}, \dot{\mathrm{x}}_k^{(c)}, \dot{\mathrm{y}}_k^{(c)}]$ are the position and velocity components of the c-th target.

$$F = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \text{ is the state transition matrix. } v_k^{(c)} \sim$$

$N(0, V)$ is the process noise.

At each time step, all targets emit sounds of amplitude Ψ. Attenuated sounds are measured by all sensors. Each sensor only records the sum of amplitudes. Thus, the measurement function for the s-th sensor located at R^s is additive:

$$\bar{z}^s(x_k) = \sum_{c=1}^{C} \frac{\Psi}{||(\mathrm{x}_k^{(c)}, \mathrm{y}_k^{(c)})^T - R^s||_2 + d_0}, \quad (38)$$

where $|| \cdot ||_2$ is the Euclidean norm, $d_0 = 0.1$ and $\Psi = 10$. There are $N_s = 25$ sensors located at grid intersections within the tracking area, as shown in Figure 1. The measurements are perturbed by Gaussian noise, i.e., the noisy measurement z_k^s from the s-th sensor is drawn from $N(\bar{z}^s(x_k), \sigma_w^2)$. σ_w^2 is set to 0.01. This leads to very informative measurements.

The initial target states are $[12, 6, 0.001, 0.001]^T$, $[32, 32, -0.001, -0.005]^T$, $[20, 13, -0.1, 0.01]^T$ and $[15, 35, 0.002, 0.002]^T$. 100 random trajectories are simulated

[1] http://networks.ece.mcgill.ca/sites/default/files/PFPF.zip

using a constant velocity model with the process covariance

matrix $\frac{1}{20} \begin{bmatrix} 1/3 & 0 & 0.5 & 0 \\ 0 & 1/3 & 0 & 0.5 \\ 0.5 & 0 & 1 & 0 \\ 0 & 0.5 & 0 & 1 \end{bmatrix}$. One set of measurements

is generated for each trajectory. We run each algorithm 5 times on each measurement set. Each execution starts with a different initial distribution. We implement the simulation using Matlab.
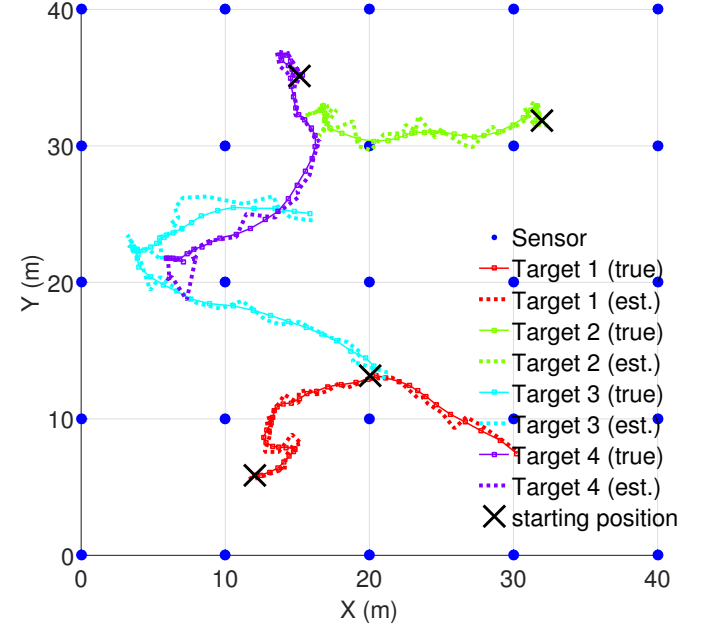

Fig. 1. One example of estimated trajectories using PF-PF (LEDH). The crosses mark the starting positions of the four targets and the solid lines show their true trajectories. Dotted lines indicate the estimated trajectories.

2) Parameter values for the filtering algorithms: The mean of the initial distributions for the filtering algorithms is sampled from a Gaussian centered at the true initial states. The standard deviation is 10 for positions and 1 for velocities. If the initial mean is outside of the tracking area, we reject and resample it. The covariance of the process noise for the

filters is set as $\begin{bmatrix} 3 & 0 & 0.1 & 0 \\ 0 & 3 & 0 & 0.1 \\ 0.1 & 0 & 0.03 & 0 \\ 0 & 0.1 & 0 & 0.03 \end{bmatrix}$. The entries are larger

than those used to generate the target trajectories, because we assume that there is more uncertainty about the model during tracking. Resampling is performed when the effective sample size (ESS) is less than $\frac{N_p}{2}$. The ESS at time step k is estimated as $\frac{1}{\sum_{i=1}^{N_p}(w_k^i)^2}$ after weight normalization. $N_p = 500$ particles are used in all Monte Carlo-based algorithms except for the BPF. The EKF is used to estimate the predictive covariance needed to calculate the flow parameters in Equation (13) and (14). The EDH and LEDH filter implementations adopt the redraw strategy in [50] at the beginning of each time step. $2d+1$ sigma points are generated for each particle in the UPF, where $d = 16$ is the state dimension. We use the ensemble square root filter (ESRF) [53], a popular implementation of the ensemble Kalman filter (EnKF), to construct the transport

map for the GSMC [41]. We set the diffusion term for the GPFIS algorithm to 0, as suggested in [3].

3) Experimental results: For this simulation scenario, we compare the PF-PF (LEDH) and PF-PF (EDH) algorithms proposed in this paper with the GPFIS algorithm [3], the BPF [1], and the EDH and LEDH particle flow algorithms [31], [50], the GSMC [41] and various Kalman-type filters [46], [53]. For this example, we do not compare with the SmHMC algorithm [26], because the target marginal posterior distribution is not log-concave, so the negative Hessian is not globally positive-definite, rendering implementation of SmHMC more challenging. We also do not compare with the block particle filter [16], because identifying a suitable partitioning of the state space is difficult, since many state variables contribute to each measurement.

The error metric we use in this multi-target tracking scenario with a fixed number of targets is the optimal mass transfer (OMAT) metric [54]. The OMAT metric $d_p(X, \hat{X})$ between two arbitrary sets $X = \{x_1, x_2, \ldots, x_C\}$ and $\hat{X} = \{\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_C\}$ is defined as

$$d_p(X, \hat{X}) = \left(\frac{1}{C} \min_{\pi \in \Pi} \sum_{c=1}^{C} d(x_c, \hat{x}_{\pi(c)})^p\right)^{1/p} \quad (39)$$

where the scalar p is a fixed parameter, Π is the set of possible permutations of $\{1, 2, \ldots, C\}$, and $d(x, \hat{x})$ is the Euclidean distance between x and $\hat{x}$. We set p to 1, so the OMAT metric assigns targets using the permutation that minimizes the Euclidean distance to the true target positions.

Figure 2 shows the average OMAT metric at each time step for the various tracking algorithms we compare. The PF-PF (LEDH) exhibits the smallest average tracking error, and reduces the average OMAT below 2 meters with just one time step. A sample of the estimated trajectories is shown in Figure 1. The PF-PF (LEDH) has much better performance than the LEDH flow algorithm which it uses to generate the proposal distribution. This demonstrates the benefits brought by the importance sampling step in the PF-PF (LEDH). We also observe that the invertible particle flow procedure we design based on the LEDH flow has similar performance to the LEDH filter. The EDH filter leads to much larger average tracking errors than the LEDH filter. The PF-PF (EDH) is much less accurate than the PF-PF (LEDH), indicating that the proposal distribution constructed using the EDH flow does not provide a good match to the posterior distribution. When the measurement function h varies significantly over the state space, it is important to perform local linearization and apply different mapping functions to different particles.

The GPFIS also has impressive tracking performance in the first 20 time steps. However, the estimation error increases in later time steps, possibly due to the fact that the weight update is approximate, unless the integration step size goes to 0. However, this is not computationally feasible, since even with 29 discrete time steps, GPFIS is the most computationally expensive algorithm, as shown in Table I. The BPF with 1 million particles has the second smallest average error in the later time steps, significantly smaller than BPF with 10^5 particles. All tested variants of Kalman-type filters, including the EKF,
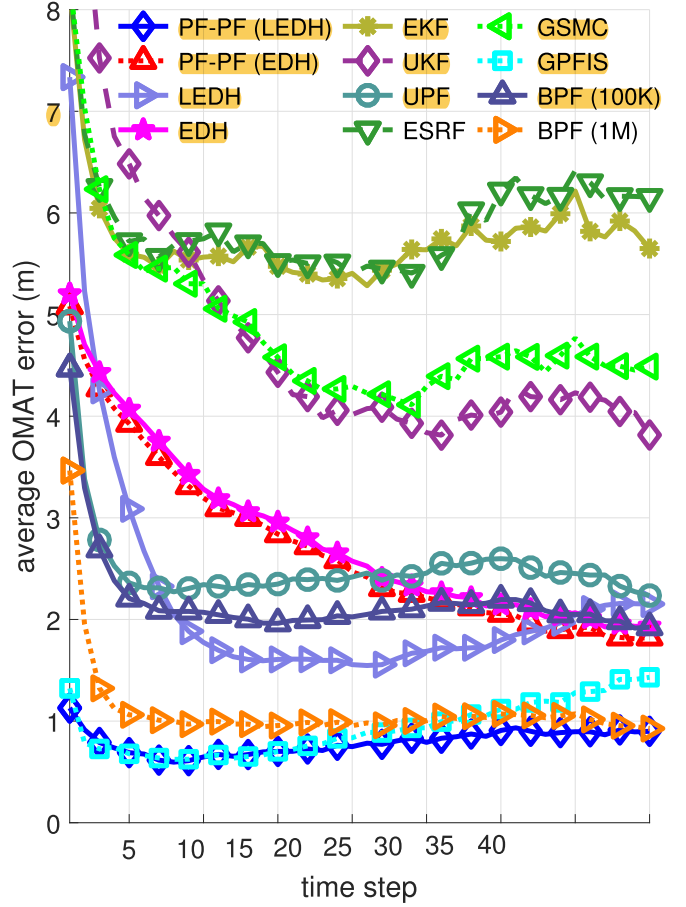

Fig. 2. Average OMAT errors at each time step in the multi-target acoustic tracking example.

the ESRF, and the UKF, exhibited very poor average tracking performance, probably because of the strong non-linearity of the measurement function. The UPF has smaller average estimation error than the UKF, but the average ESS is still small. The GSMC method has relatively poor performance, due to the fact that the ESRF it uses to construct the transport map does not generate very good proposal distributions. The boxplots of average (over time) OMAT (Figure 3) present similar performance relations. The PF-PF (LEDH) has the smallest median error as well as the first and third quartiles. There are also far fewer outliers, which are possible indicators of lost tracks, than for most of the other algorithms.

From Table I and Figure 4, we can see that the PF-PF algorithms and GPFIS provide the highest effective sample sizes among all tested algorithms with importance sampling. Even with one million particles, the average ESS of the bootstrap particle filter is still less than 7, and the cost of computation is more than 3 times the cost of the PF-PF (LEDH). The UPF has relatively high computational cost, as unscented transformations are performed for each sigma point, and the number of sigma points for each particle is proportional to the state dimension. The GSMC algorithm has a small effective sample size of 3.5. This shows that the transport map constructed by the ESRF is not effective for GSMC in this example. The PF-PF (LEDH) has a small

TABLE I
AVERAGE OMAT METRICS, ESS AND EXECUTION TIME PER STEP.
RESULTS ARE PRODUCED WITH AN INTEL I7-4770K 3.50GHz CPU AND
32GB RAM.

Algorithm	Particle num.	Avg. OMAT (m)	Avg. ESS	Exec. time (s)
PF-PF (LEDH)	500	0.79	45	0.9
PF-PF (EDH)	500	2.71	34	0.01
LEDH	500	2.19	N/A	0.8
EDH	500	2.81	N/A	0.01
EKF	N/A	5.74	N/A	0.00003
UKF	N/A	4.91	N/A	0.005
UPF	500	2.51	1.48	2.0
ESRF	500	5.90	N/A	0.01
GSMC	500	4.87	3.5	1.6
GPFIS	500	0.93	30	66.8
BPF	10^5	2.18	2.1	0.3
BPF	10^6	1.10	6.3	3.0

Fig. 3. Boxplots of average OMAT errors in the multi-target acoustic tracking example.

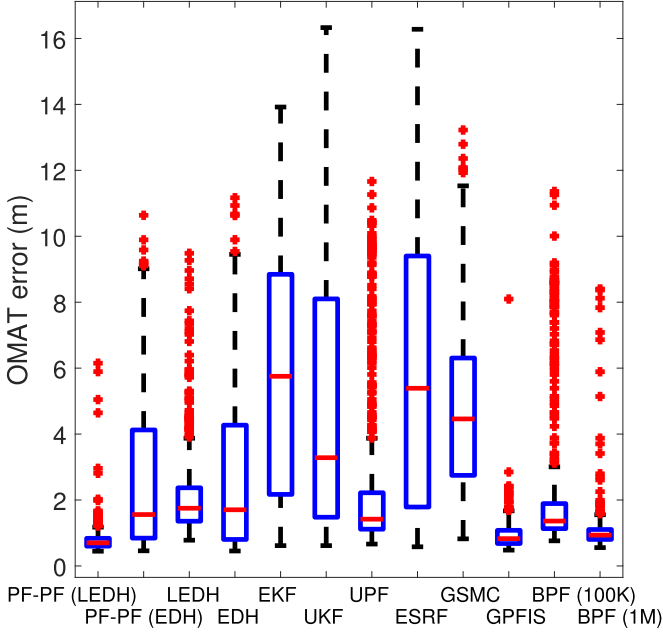Fig. 4. Average effective sample size at each time step in the multi-target acoustic tracking example.

increase of execution time compared with the LEDH, showing that the weight update is very efficient.

B. Large spatial sensor networks: linear Gaussian example

1) Simulation setup: To compare the accuracy of proposed filters relative to theoretical optimal accuracy, we examine the filters' performance in a simple linear Gaussian filtering problem with the spatial sensor network setup proposed in [26]. There are d sensors deployed uniformly on a two-dimensional grid $\{1, 2, \ldots, \sqrt{d}\} \times \{1, 2, \ldots, \sqrt{d}\}$, and d is set to 64 in this example. Each sensor collects measurements, independently of the other sensors, about the underlying state at its physical location. Denote the state at the c-th sensor's position at time k by $x_k^c \in \mathbb{R}$, and its measurement as $z_k^c \in \mathbb{R}$.
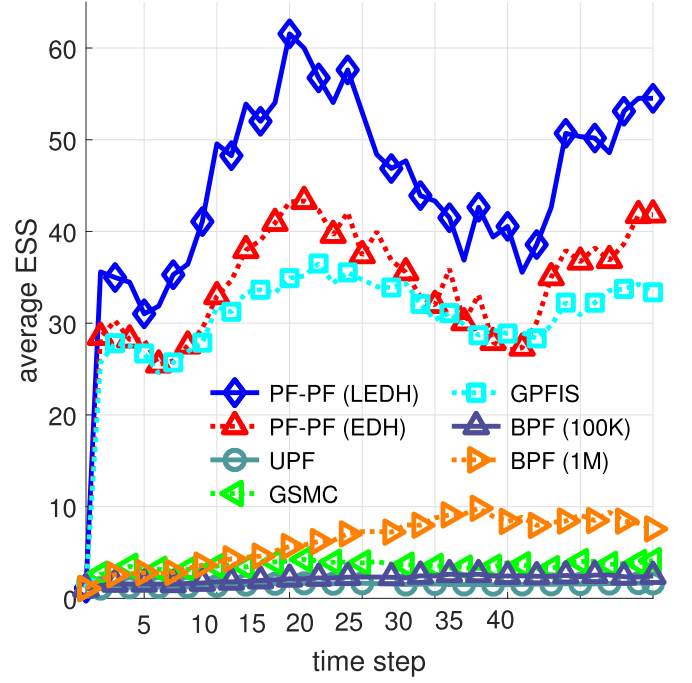
Then the full state at all sensor positions at time k is denoted by $x_k = [x_k^1, x_k^2, \ldots, x_k^d]' \in \mathbb{R}^d$, and all measurements at time k form the measurement vector $z_k = [z_k^1, z_k^2, \ldots, z_k^d]' \in \mathbb{R}^d$.

The dynamic model and the measurement model are, respectively:

$$x_k = \alpha x_{k-1} + v_k \,, \tag{40}$$

$$z_k = x_k + w_k \,, \tag{41}$$

where $\alpha = 0.9$, and $w_k \in \mathbb{R}^d$ is a zero-mean Gaussian random vector with covariance $\Sigma_z = \sigma_z^2 \mathbb{I}_{d \times d}$. $v_k \in \mathbb{R}^d$ is a zero-mean Gaussian random vector with covariance Σ. The (i, j)-th entry of Σ is

$$\Sigma_{i,j} = \alpha_0 e^{-\frac{||R^i - R^j||_2^2}{\beta}} + \alpha_1 \delta_{i,j} \,, \tag{42}$$

where $|| \cdot ||_2$ is the Euclidean norm, $R^i \in \mathbb{R}^2$ is the physical position of sensor i, and $\delta_{i,j}$ is the Kronecker delta symbol ($\delta_{i,i} = 1$ and $\delta_{i,j} = 0$ for $i \neq j$). This equation implies that the noise dependence increases when the spatial distance between two sensors decreases.

Following [26], we set $\alpha_0 = 3, \alpha_1 = 0.01, \beta = 20$. We vary the value of σ_z, to investigate the sensitivity of algorithms with respect to the level of measurement noise. All true states start with $x_0^c = 0$, for $c = 1, \ldots, d$. The experiment is executed 100 times for 10 time steps.

2) Parameter values for the filtering algorithms: We compare the proposed particle flow particle filter (PF-PF) algorithms with various Kalman-type filters and other particle filters with different measurement noise levels. In the linear Gaussian scenario, the Kalman filter (KF) provides the exact posterior distribution, thus it gives the optimal filtering accuracy. The EDH, the ESRF, and the UKF are all derived based

on the linear Gaussian assumptions. They thus provide near-optimal filtering performance, although their use of Monte Carlo samples or sigma points make their solutions deviate slightly from optimal. The BPF is the vanilla particle filter that uses the dynamic model to propose particles. The main goal of this experiment is to compare the PF-PF algorithms with those filters that provide optimal or near-optimal performance. We also demonstrate the challenges to particle filters from highly informative measurements by varying the measurement noise level.

For the algorithms employing particle flow, the step sizes are set to be the same as those reported in Section V-A2 and the Kalman filter covariance equations are used to estimate the predicted covariance. All filters are initialized with the same true state 0 in each state dimension.

3) Experimental results: Table II reports the average mean squared errors (MSEs) over 100 simulation trials and the execution times per time step with σ_z set to 2, 1, or 0.5.

TABLE II
AVERAGE MSE, ESS AND EXECUTION TIME PER STEP IN A 64-DIMENSIONAL LINEAR GAUSSIAN EXAMPLE OF THE LARGE SPATIAL SENSOR NETWORKS SIMULATION WITH DIFFERENT MEASUREMENT NOISE LEVELS, AMONG 100 SIMULATION TRIALS. RESULTS ARE PRODUCED WITH AN INTEL I7-4770K 3.50GHZ CPU AND 32GB RAM.

σ_z		2		1		0.5		Exec. time (s)
Algorithm	Particle num.	Avg. MSE	Avg. ESS	Avg. MSE	Avg. ESS	Avg. MSE	Avg. ESS	
PF-PF (LEDH)	200	0.61	28	0.25	23	0.10	19	1.8
PF-PF (EDH)	200	0.62	28	0.26	23	0.11	19	0.01
PF-PF (EDH)	10^4	0.53	1118	0.22	973	0.09	830	0.24
EDH	200	0.49	N/A	0.19	N/A	0.07	N/A	0.01
KF	N/A	0.49	N/A	0.18	N/A	0.07	N/A	0.0005
UKF	N/A	0.49	N/A	0.18	N/A	0.07	N/A	0.007
UPF	200	0.87	22	0.32	22	0.12	22	1.4
ESRF	200	0.50	N/A	0.19	N/A	0.07	N/A	0.001
GSMC	200	0.52	2.0	0.19	2.0	0.08	2.0	0.002
BPF	200	1.20	1.9	1.1	1.2	1.1	1.0	0.00006
BPF	10^5	0.54	49	0.32	3.2	0.29	1.3	0.24

We observe that the Kalman filter, the ESRF, the UKF, and the EDH filter have the smallest mean squared errors for different σ_z values. This is expected as the linear Gaussian models match their model assumptions. We also observe that with 200 particles, the GSMC algorithm has the smallest average MSEs among all particle filters. When the number of particles is increased to 10^4, the PF-PF (EDH) is still relatively efficient, and has average MSEs close to those of the Kalman filter.

We also note that for the BPF, as σ_z gets smaller, the average ESS drastically decreases, which indicates that smaller measurement noise leads to a more challenging situation for

many particle filters. However, for the PF-PF algorithms, the average effective sample size only decreases slightly as σ_z decreases from 2 to 0.5, showing that particle flows are able to propagate particles into the region of high posterior densities, even in the more challenging scenarios with highly informative measurements. When $\sigma_z = 0.5$, the PF-PF (EDH) and the GSMC with 200 particles are more accurate than the BPF with 10^5 particles.

We next investigate the sensitivity of the proposed algorithm to the approximation of the predictive covariance. We add noise to the Kalman filter-estimated covariance $\hat{P}$ to generate a noisy estimate $\tilde{P}$. To ensure that the noise-injected $\tilde{P}$ is positive definite (as a requirement for the covariance matrix), we first perform eigendecomposition of $\hat{P}$ and obtain the eigenvalues $D \in \mathbb{R}^d$ and the corresponding right eigenvectors $V \in \mathbb{R}^{d \times d}$. We then generate $\tilde{D} = \xi \circ D$, where $\xi \in \mathbb{R}^d$ is a random vector and $\circ$ denotes the dot product. The elements of ξ are independent and are distributed according to a log-normal distribution $\ln N(0, \sigma_p^2)$. Because $\hat{P}$ is positive definite, each element of D is positive. All elements of ϵ are also positive as they are generated from a log-normal distribution. Thus, all elements of $\tilde{D}$ are positive. We then generate $\tilde{P} = V \times \text{diag}(\tilde{D}) \times V^T$ where $\text{diag}(D)$ denotes a diagonal matrix with the elements of the vector D on the diagonal. With this procedure, $\tilde{P}$ is positive definite as all of its eigenvalues are positive. Numerical results show that the expected value of $\frac{|\hat{P} - \tilde{P}|}{|\hat{P}|}$, the ratio between the Euclidean norm of the error added to $\hat{P}$ and the Euclidean norm of $\hat{P}$, is approximately equal to σ_p. We report in Table III the MSE from 100 simulation trials with different values of σ_p.

TABLE III
AVERAGE MSE AND ESS IN THE 64-DIMENSIONAL LINEAR GAUSSIAN EXAMPLE OF THE LARGE SPATIAL SENSOR NETWORKS WITH $\sigma_z = 1$ AND DIFFERENT PREDICTIVE COVARIANCE ESTIMATION ERRORS, AMONG 100 SIMULATION TRIALS.

Algorithm	PF-PF (LEDH)		PF-PF (EDH)		PF-PF (EDH)	
Particle num.	200		200		10^4	
	Avg. MSE	Avg. ESS	Avg. MSE	Avg. ESS	Avg. MSE	Avg. ESS
$\sigma_p = 0$	0.25	23	0.26	23	0.22	973
$\sigma_p = 0.1$	0.39	26	0.39	26	0.34	1018
$\sigma_p = 0.2$	0.40	23	0.40	23	0.35	894
$\sigma_p = 0.5$	0.43	12	0.43	12	0.39	343
$\sigma_p = 1$	0.54	3.6	0.54	3.6	0.51	32

We observe that as we increase the amount of noise added into the estimated covariance, by increasing the standard deviation σ of the log-normal distribution, the PF-PF (EDH) tends to have higher mean squared error (MSE) and smaller effective sample size (ESS). However, even when $\sigma_p = 0.2$, which adds considerable noise to the eigenvalues of P, there is only a minor reduction in the ESS. It is only when σ_p reaches 0.5 or 1, indicating a very noisy predictive covariance matrix estimate, that the effective sample size decreases significantly.

Khan et al. investigates the effects of different covariance approximation techniques on the performance of the particle flow filter in [51]. Our experience with these techniques is that their impact on the overall tracking performance depends on the specific nature of the application. For consistency across our simulation results, we use the EKF covariance equations for the estimation of predictive covariance.

C. Large spatial sensor networks: Skewed-t dynamic model and count measurements

1) Simulation setup: The spatial placement of d sensors is the same as that introduced in V-B1. The dynamic model of the underlying state x_k in this example follows the multivariate Generalized Hyperbolic (GH) skewed-t distribution, which is a heavy-tailed distribution that is useful for modelling physical processes and financial markets with extreme behavior and asymmetric data [55]. We have

$$p(x_k|x_{k-1}) = K_{\frac{\nu+d}{2}}(\sqrt{(\nu + Q(x_k))(\gamma^T \Sigma^{-1} \gamma)})$$
$$\times \frac{e^{(x_k - \mu_k)^T \Sigma^{-1} \gamma}}{\sqrt{(\nu + Q(x_k))(\gamma^T \Sigma^{-1} \gamma)}^{-\frac{\nu+d}{2}} (1 + \frac{Q(x_k)}{\nu})^{\frac{\nu+d}{2}}} , \quad (43)$$

where $K_{\frac{\nu+d}{2}}$ is the modified Bessel function of the second kind of order $\frac{\nu+d}{2}$, $\mu_k = \alpha x_{k-1}$, $Q(x_k) = (x_k - \mu_k)^T \Sigma^{-1}(x_k - \mu_k)$, and the (i,j)-th entry of Σ is again defined by Equation (42). The parameters γ and ν determine the shape of the distribution. The covariance is given by:

$$\tilde{\Sigma} = \frac{\nu}{\nu - 2} \Sigma + \frac{\nu^2}{(2\nu - 8)(\frac{\nu}{2} - 1)^2} \gamma\gamma^T . \quad (44)$$

The measurements are count data with the following Poisson distribution

$$p(z_k|x_k) = \prod_{c=1}^{d} \mathcal{P}_0(z_k^c; m_1 e^{m_2 x_k^c}) , \quad (45)$$

where $\mathcal{P}_0(\cdot; m)$ is the Poisson(m) distribution. We set $m_1 = 1$ and $m_2 = \frac{1}{3}$. d is set to 144 or 400 to represent two high-dimensional filtering scenarios. Again, each scenario is executed 100 times for 10 time steps. We choose the parameter values of the simulation setup to be the same as those used in [26], because we would like to evaluate the proposed filters in the same simulation setups where the state-of-the-art SmHMC algorithm has been compared to other Langevin and Hamiltonian-based algorithms and exhibited the smallest estimation errors.

2) Parameter values for the filtering algorithms: We compare the proposed PF-PF algorithms with the Kalman-type filters and particle filters evaluated in Section V-B. In addition, we evaluate two filters specifically designed for high-dimensional nonlinear filtering: one is the SmHMC, which exhibits the smallest mean squared error (MSE) in [26], and the other is the block particle filter [16]. We do not compare with the GPFIS algorithm [3], due to its prohibitively large computational cost.

Parameter values for the SmHMC algorithm are set to those identified in [26], including the number of particles which is

200. We evaluate the performance for the PF-PFs using 200 particles, as well as a higher number of particles for the PF-PF (EDH) with the constraint that its computational time remains less than that of SmHMC with 200 particles. The step sizes of the particle flow-type algorithms are set to be the same as those reported in Section V-A2. Since the measurement noise depends on the state, the measurement covariance R is updated in each discretized particle flow step and before the EKF update. For the PF-PF (LEDH), R is updated using $\bar{\eta}^i$ for each particle; for the PF-PF (EDH), $\bar{\eta}$ is used. All filters are initialized with the same true state 0 in each state dimension, which is the scenario explored in [26].

3) Experimental results: Table IV reports the average mean squared errors (MSEs) over 100 simulation trials and the execution times per time step. We observe that the EDH and the LEDH filters have very similar average MSE errors. This suggests that computing the flow parameters separately for each particle does not provide additional gains in this setting, which is different from the result shown in Section V-A. Thus, the EDH is preferred over the LEDH as it is much less computationally demanding.

TABLE IV
AVERAGE MSE, ESS AND EXECUTION TIME PER STEP IN THE LARGE SPATIAL SENSOR NETWORKS SIMULATION WITH A SKEWED-t DYNAMIC MODEL AND COUNT MEASUREMENTS. RESULTS ARE PRODUCED WITH AN INTEL I7-4770K 3.50GHz CPU AND 32GB RAM. WHEN THERE IS A PARENTHESIS AFTER THE AVERAGE MSE VALUES, IT INDICATES THE NUMBER OF LOST TRACKS OUT OF 100 SIMULATION TRIALS WHERE THE LOST TRACKS ARE DEFINED AS THOSE WHOSE AVERAGE ESTIMATION ERRORS ARE GREATER THAN $\sqrt{d}$. THE AVERAGE MSE IS CALCULATED WITH THE SIMULATION TRIALS WHERE TRACKING IS NOT LOST.

d		144			400		
Algorithm	Particle num.	Avg. MSE	Avg. ESS	Exec. time (s)	Avg. MSE	Avg. ESS	Exec. time (s)
PF-PF (LEDH)	200	0.95	6.7	7.8	1.04	3.4	110
PF-PF (EDH)	200	0.96	6.6	0.05	1.05	3.4	0.5
PF-PF (EDH)	10^4	0.82	81	1.6	0.89	20	4.8
LEDH	200	0.71	N/A	6.8	0.62	N/A	88
EDH	200	0.69	N/A	0.05	0.60	N/A	0.5
EDH	10^4	0.69	N/A	0.6	0.60	N/A	2.5
SmHMC	200	0.83	N/A	15	0.73	N/A	87
Block PF	10^4	1.7	N/A	10	1.6	N/A	29
EKF	N/A	2.5 (29)	N/A	0.002	3.4 (18)	N/A	0.03
UKF	N/A	2.4 (34)	N/A	0.05	3.8 (27)	N/A	1.2
UPF	200	2.2 (34)	3.0	12	4.6 (43)	1.4	236
ESRF	200	2.3 (23)	N/A	0.01	2.8 (15)	N/A	0.05
GSMC	200	2.4 (22)	1.1	0.02	3.5 (23)	1.0	0.06
BPF	10^5	1.8	6.3	0.7	4.0 (1)	1.3	2.3
BPF	10^6	1.3 (1)	26	6.8	3.3 (1)	1.5	23

We also observe that the PF-PF (EDH) and the PF-PF (LEDH) lead to larger average MSEs than the EDH or the LEDH with the same number of particles. The EDH and LEDH filters use particle flow to generate approximations of the posterior distribution; the PF-PF algorithms perform subsequent importance sampling to modify this approximation. The importance sampling makes the filter statistically consistent, but in high dimensions, it can introduce a high variance in the weights, leading to poorer performance in state estimation. This sampling error can be reduced by increasing the number of particles, and we see that with 10000 particles, the average MSE of PF-PF (EDH) has significantly decreased, as the effective sample size increases considerably. Even with 10000 particles, the PF-PF (EDH) is more computationally efficient than the SmHMC with 200 particles. The estimation errors are similar for the PF-PF (EDH) and SmHMC when $d = 144$, although the average error of the PF-PF (EDH) is considerably higher when $d = 400$. The Block PF has relatively large estimation errors, possibly due to the intrinsic bias introduced from the blocking step as stated in [16]. The Kalman-type filters frequently struggle to track the state at all, leading to lost tracks. The ESRF in particular has a high number of lost tracks when the state dimension is 400, as the sample predictive covariance is often close to singular. The GSMC method and the UPF, which use the ESRF and the UKF, respectively to construct the proposal distributions, also exhibit poor performance. Even with a million particles, the BPF performs relatively poorly compared to most other algorithms.

We evaluate the effective sample size (ESS) for the PF-PF algorithms and other particle filters. The standard ESS estimate is not meaningful for SmHMC, because due to its MCMC structure there are no weights associated with the particles; for the block PF we can only calculate an ESS for each block, so the value is not comparable. The ESS values indicate why the PF-PF filters perform worse in this setting compared to the acoustic tracking example. Only a very small fraction of the particles have a significant weight. There is however, a substantial improvement compared to the BPF, the UPF and the GSMC; the BPF requires 100 times more particles to achieve comparable ESS values.

VI. CONCLUSIONS

In this paper, we have presented particle flow particle filtering algorithms with efficient importance weight computation. We proved that the embedded particle flows possess the invertible mapping property, which is crucial for achieving straightforward weight updates. The weight updates add only a small computation cost to particle flow filters they build upon.

We have evaluated the proposed algorithms' performance in three scenarios. In the multi-target tracking simulation setup, the PF-PF (LEDH) with 500 particles leads to the smallest tracking error and maintains particle clouds with the highest effective sample size at most time steps. This demonstrates that the PF-PF (LEDH) is capable of producing better particle representations of posterior distributions than other filtering algorithms with much higher computational cost in this highly

informative measurement setting. In a linear Gaussian filtering example, we show that the PF-PF algorithms approach the optimal accuracy with a reasonable number of particles in different settings with various levels of measurement noise. We also discuss the sensitivity of the PF-PF with respect to the estimation errors of the predictive covariance. In the large spatial sensor network setting where the state dimension is high and the models are non-Gaussian, the EDH filter provides the smallest average MSE and is computationally efficient. The error introduced by incorporating importance sampling in the proposed PF-PF (EDH) algorithm outweighs the approximation error in the EDH filter.

The proposed PF-PF algorithms are computationally efficient particle filters that can perform well in high-dimensional settings, but the last simulation motivates the development of improved mechanisms for using the particle flow procedures to construct a consistent filter. An important future research direction is the construction of improved invertible particle flows, which may be achieved by employing stochastic particle flows [35], [39] or performing linearization at the actual particle locations. It can be significantly more difficult to construct stochastic particle flows with the invertible mapping property, but the diffusion of particles may lead to more diverse sets of particles and hence better proposal distributions. Other directions include the identification of new particle flows based on mixture models to allow the construction of proposal distributions that can match multi-modal distributions, the incorporation of iterative importance sampling [56] to increase the effective sample size in high-dimensional filtering scenarios, the integration of quasi Monte Carlo [57] with particle flow particle filters for faster convergence, and convergence studies of particle filters with invertible particle flow.

REFERENCES
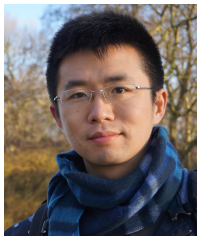
[1] N. Gordon, D. Salmond, and A. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," in *IEE Proc. F Radar and Signal Processing*, vol. 140, no. 2, Apr. 1993, pp. 107–113.

[2] P. Bickel, B. Li, and T. Bengtsson, "Sharp failure rates for the bootstrap particle filter in high dimensions," in *Pushing the limits of contemporary statistics: Contributions in honor of Jayanta K. Ghosh.* Institute of Mathematical Statistics, 2008, pp. 318–329.

[3] P. Bunch and S. Godsill, "Approximations of the optimal importance density using Gaussian particle flow importance sampling," *J. Amer. Statist. Assoc.*, vol. 111, no. 514, pp. 748–762, 2016.

[4] A. Doucet, S. Godsill, and C. Andrieu, "On sequential Monte Carlo sampling methods for Bayesian filtering," *Stat. Comput.*, vol. 10, no. 3, pp. 197–208, 2000.

[5] J. Cornebise, E. Moulines, and J. Olsson, "Adaptive methods for sequential importance sampling with application to state space models," *Stat. Comput.*, vol. 18, no. 4, pp. 461–480, 2008.

[6] M. Pitt and N. Shephard, "Filtering via simulation: Auxiliary particle filters," *J. Am. Statist. Assoc.*, vol. 94, no. 446, pp. 590–599, Jun. 1999.

[7] A. Doucet, N. d. Freitas, K. P. Murphy, and S. J. Russell, "Rao-Blackwellised particle filtering for dynamic Bayesian networks," in *Proc. Conf. Uncertainty in Artificial Intelligence (UAI)*, San Francisco, CA, 2000, pp. 176–183.

[8] R. van der Merwe, A. Doucet, N. De Freitas, and E. Wan, "The unscented particle filter," in *Proc. Neural Info. Proc. Sys. (NIPS)*, Denver, CO, Dec. 2000, pp. 584–590.

[9] T. Bengtsson, P. Bickel, and B. Li, "Curse-of-dimensionality revisited: Collapse of the particle filter in very large scale systems," in *Probability and Statistics: Essays in Honor of David A. Freedman*, D. Nolan and T. Speed, Eds. Beachwood, OH: Institute of Mathematical Statistics, 2008, vol. 2, pp. 316–334.

[10] C. Snyder, T. Bengtsson, P. Bickel, and J. Anderson, "Obstacles to high-dimensional particle filtering," *Mon. Weather Rev.*, vol. 136, no. 12, pp. 4629–4640, 2008.

[11] A. Beskos, D. Crisan, and A. Jasra, "On the stability of sequential Monte Carlo methods in high dimensions," *Ann. Appl. Probab.*, vol. 24, no. 4, pp. 1396–1445, Aug. 2014.

[12] P. J. van Leeuwen, "Nonlinear data assimilation in geosciences: an extremely efficient particle filter," *Q. J. Royal Met. Soc.*, vol. 136, pp. 1991–1999, 2010.

[13] M. Ades and P. J. van Leeuwen, "The equivalent-weights particle filter in a high-dimensional system," *Q. J. Royal Met. Soc.*, vol. 141, no. 1, pp. 484–503, 2015.

[14] P. M. Djurić, T. Lu, and M. F. Bugallo, "Multiple particle filtering," in *Proc. Intl. Conf. Acoustics, Speech and Signal Proc. (ICASSP)*, vol. 3, Apr. 2007, pp. 1181–1184.

[15] P. M. Djurić and M. F. Bugallo, "Particle filtering for high-dimensional systems," in *Proc. Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, Dec 2013, pp. 352–355.

[16] P. Rebeschini and R. van Handel, "Can local particle filters beat the curse of dimensionality?" *Ann. Appl. Probab.*, vol. 25, no. 5, pp. 2809–2866, 10 2015.

[17] A. Beskos, D. Crisan, A. Jasra, K. Kamatani, and Y. Zhou, "A stable particle filter for a class of high-dimensional state-space models," *Adv. Appl. Probab.*, vol. 49, no. 1, p. 2448, 2017.

[18] C. Berzuini, N. G. Best, W. R. Gilks, and C. Larizza, "Dynamic conditional independence models and Markov chain Monte Carlo methods," *J. Am. Stat. Assoc.*, vol. 92, no. 440, pp. 1403–1412, 1997.

[19] W. R. Gilks and C. Berzuini, "Following a moving target–Monte Carlo inference for dynamic Bayesian models," *J. R. Stat. Soc. B*, vol. 63, no. 1, pp. 127–146, 2001.

[20] S. Godsill and T. Clapp, "Improvement strategies for monte carlo particle filters," in *Sequential Monte Carlo Methods in Practice*. New York: Springer, 2001, ch. 7, pp. 139–158.

[21] C. Musso, N. Oudjane, and F. Le Gland, "Improving regularised particle filters," in *Sequential Monte Carlo Methods in Practice*. New York: Springer, 2001, ch. 10, pp. 247–271.

[22] A. Golightly and D. J. Wilkinson, "Bayesian sequential inference for nonlinear multivariate diffusions," *Stat. Comput.*, vol. 16, no. 4, pp. 323–338, 2006.

[23] V. Maroulas and P. Stinis, "Improved particle filters for multi-target tracking," *J. Computational Physics*, vol. 231, no. 2, pp. 602–611, 2012.

[24] K. Kang, V. Maroulas, and I. D. Schizas, "Drift homotopy particle filter for non-gaussian multi-target tracking," in *Proc. Intl. Conf. Information Fusion*, Salamanca, Spain, July 2014, pp. 1–7.

[25] A. Brockwell, P. D. Moral, and A. Doucet, "Sequentially interacting Markov chain Monte Carlo methods," *Ann. Stat.*, vol. 38, no. 6, pp. 3387–3411, 2010.

[26] F. Septier and G. W. Peters, "Langevin and Hamiltonian based sequential MCMC for efficient Bayesian filtering in high-dimensional spaces," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 2, pp. 312–327, Mar. 2016.

[27] U. D. Hanebeck, K. Briechle, and A. Rauh, "Progressive Bayes: a new framework for nonlinear state estimation," in *Proc. SPIE Multisensor, Multisource Information Fusion: Architectures, Algorithms, and Applications*, vol. 5099, Orlando, FL, Apr. 2003, pp. 256–267.

[28] F. Daum and J. Huang, "Nonlinear filters with log-homotopy," in *Proc. SPIE Signal and Data Processing of Small Targets*, San Diego, CA, Sep. 2007, p. 669918.

[29] ——, "Particle flow for nonlinear filters with log-homotopy," in *Proc. SPIE Signal and Data Processing Small Targets*, Orlando, FL, Apr. 2008, p. 696918.

[30] F. Daum, J. Huang, A. Noushin, and M. Krichman, "Gradient estimation for particle flow induced by log-homotopy for nonlinear filters," in *Proc. SPIE Conf. Signal Proc., Sensor Fusion, Target Recog.*, Orlando, FL, Apr. 2009, p. 733602.

[31] F. Daum, J. Huang, and A. Noushin, "Exact particle flow for nonlinear filters," in *Proc. SPIE Conf. Signal Proc., Sensor Fusion, Target Recog.*, Orlando, FL, Apr. 2010, p. 769704.

[32] F. Daum and J. Huang, "Exact particle flow for nonlinear filters: Seventeen dubious solutions to a first order linear underdetermined PDE," in *Proc. Asilomar Conf. Signals, Systems and Computers (ASILOMAR)*, Pacific Grove, CA, Nov. 2010, pp. 64–71.

[33] F. Daum, J. Huang, and A. Noushin, "Coulomb's law particle flow for nonlinear filters," in *Proc. SPIE Conf. Signal Proc., Sensor Fusion, Target Recog.*, San Diego, CA, Sep. 2011, p. 81370B.

[34] F. Daum and J. Huang, "Small curvature particle flow for nonlinear filters," in *Proc. SPIE Signal and Data Processing of Small Targets*, Baltimore, MD, May 2012, p. 83930A.

[35] ——, "Particle flow with non-zero diffusion for nonlinear filters," in *Proc. SPIE Conf. Signal Proc., Sensor Fusion, Target Recog.*, Baltimore, MD, May 2013, p. 87450P.

[36] ——, "Renormalization group flow and other ideas inspired by physics for nonlinear filters, Bayesian decisions, and transport," in *Proc. SPIE Conf. Signal Proc., Sensor Fusion, Target Recog.*, Baltimore, MD, May 2014, p. 90910I.

[37] ——, "Seven dubious methods to mitigate stiffness in particle flow with non-zero diffusion for nonlinear filters, Bayesian decisions, and transport," in *Proc. SPIE Conf. Signal Proc., Sensor Fusion, Target Recog.*, Baltimore, MD, May 2014, p. 90920C.

[38] M. A. Khan and M. Ulmke, "Non-linear and non-Gaussian state estimation using log-homotopy based particle flow filters," in *Proc. Sensor Data Fusion: Trends, Solutions, Applications (SDF)*, Bonn, Germany, Oct. 2014, pp. 1–6.

[39] F. E. de Melo, S. Maskell, M. Fasiolo, and F. Daum, "Stochastic particle flow for nonlinear high-dimensional filtering problems," *arXiv:1511.01448*, 2015.

[40] D. Crisan and A. Doucet, "A survey of convergence results on particle filtering methods for practitioners," *IEEE Trans. Signal Process.*, vol. 50, pp. 736–746, Mar. 2002.

[41] S. Reich, "A guided sequential Monte Carlo method for the assimilation of data into stochastic dynamical systems," in *Recent Trends in Dynamical Systems*. Basel, Switzerland: Springer, 2013, vol. 35, pp. 205–220.

[42] J. Heng, A. Doucet, and Y. Pokern, "Gibbs flow for approximate transport with applications to Bayesian computation," *arXiv:1509.08787*, 2015.

[43] Y. Li, L. Zhao, and M. J. Coates, "Particle flow auxiliary particle filter," in *Proc. Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, Cancun, Mexico, Dec. 2015, pp. 157–160.

[44] ——, "Particle flow for particle filtering," in *Proc. Intl. Conf. Acoustics, Speech and Signal Proc. (ICASSP)*, Shanghai, China, Mar. 2016, pp. 3979–3983.

[45] Y. Li and M. J. Coates, "Fast particle flow particle filter via clustering," in *Proc. Intl. Conf. Information Fusion*, Heidelberg, Germany, July 2016, pp. 2022–2027.

[46] S. J. Julier and J. K. Uhlmann, "New extension of the Kalman filter to nonlinear systems," in *Proc. SPIE Conf. Signal Proc., Sensor Fusion, Target Recog.*, Orlando, FL, Apr. 1997, pp. 182–193.

[47] G. Evensen, "The ensemble Kalman filter: theoretical formulation and practical implementation," *Ocean Dynamics*, vol. 53, no. 4, pp. 343–367, 2003.

[48] P. J. van Leeuwen, "Nonlinear data assimilation for high-dimensional systems," in *Nonlinear Data Assimilation*, P. J. van Leeuwen, C. Yuan, and S. Reich, Eds. Switzerland: Springer, 2015, ch. 1, pp. 1–73.

[49] F. W. Warner, *Foundations of differentiable manifolds and Lie groups*. Berlin, Germany: Springer, 1983.

[50] T. Ding and M. J. Coates, "Implementation of the daum-huang exact-flow particle filter," in *Proc. IEEE Statistical Signal Processing Workshop (SSP)*, Ann Arbor, MI, Aug. 2012, pp. 257–260.

[51] M. A. Khan and M. Ulmke, "Improvements in the implementation of log-homotopy based particle flow filters," in *Proc. Intl. Conf. Information Fusion*, July 2015, pp. 74–81.

[52] O. Hlinka, O. Sluciak, F. Hlawatsch, P. M. Djuric, and M. Rupp, "Distributed Gaussian particle filtering using likelihood consensus," in *Proc. Intl. Conf. Acoustics, Speech and Signal Proc. (ICASSP)*, Prague, Czech Republic, May 2011, pp. 3756–3759.

[53] G. Evensen, "Sampling strategies and square root analysis schemes for the EnKF," *Ocean Dynamics*, vol. 54, no. 6, pp. 539–560, 2004.

[54] D. Schuhmacher, B. T. Vo, and B. N. Vo, "A consistent metric for performance evaluation of multi-object filters," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3447–3457, Aug 2008.

[55] D. Zhu and J. W. Galbraith, "A generalized asymmetric Student-t distribution with application to financial econometrics," *J. Econom.*, vol. 157, no. 2, pp. 297–305, 2010.

[56] M. Morzfeld, M. S. Day, R. W. Grout, G. S. H. Pau, S. A. Finsterle, and J. B. Bell, "Iterative importance sampling algorithms for parameter estimation problems," *arXiv:1608.01958*, 2016.

[57] M. Gerber and N. Chopin, "Sequential quasi Monte Carlo," *J. R. Stat. Soc. Ser. B (Stat. Method.)*, vol. 77, no. 3, pp. 509–579, 2015.

Yunpeng Li (S'16) is a Ph.D. candidate at the Department of Electrical and Computer Engineering at McGill University, Canada. He received the B.A. and M.S. Eng. degrees from the Beijing University of Posts and Telecommunications, China, in 2009 and 2012, respectively. He joined the Machine Learning Research Group at University of Oxford, U.K., as a Postdoctoral Researcher in Machine Learning in April 2017. He has conducted research internships at HP Labs China (Spring 2012) and McGill University (Summer 2011, Summer 2010). His research interests include Monte Carlo methods in high-dimensional spaces, microwave breast cancer detection, and Bayesian inference.

Mark Coates (SM'04) received the B.E. degree in computer systems engineering from the University of Adelaide, Australia, in 1995, and a Ph.D. degree in information engineering from the University of Cambridge, U.K., in 1999. He joined McGill University (Montreal, Canada) in 2002, where he is currently an Associate Professor in the Department of Electrical and Computer Engineering. He was a research associate and lecturer at Rice University, Texas, from 1999-2001. In 2012-2013, he worked as a Senior Scientist at Winton Capital Management, Oxford, UK. He was an Associate Editor of IEEE Transactions on Signal Processing from 2007-2011 and a Senior Area Editor for IEEE Signal Processing Letters from 2012-2015. In 2006, his research team received the NSERC Synergy Award in recognition of their successful collaboration with Canadian industry, which has resulted in the licensing of software for anomaly detection and Video-on-Demand network optimization. Coates' research interests include communication and sensor networks, statistical signal processing, and Bayesian and Monte Carlo inference.