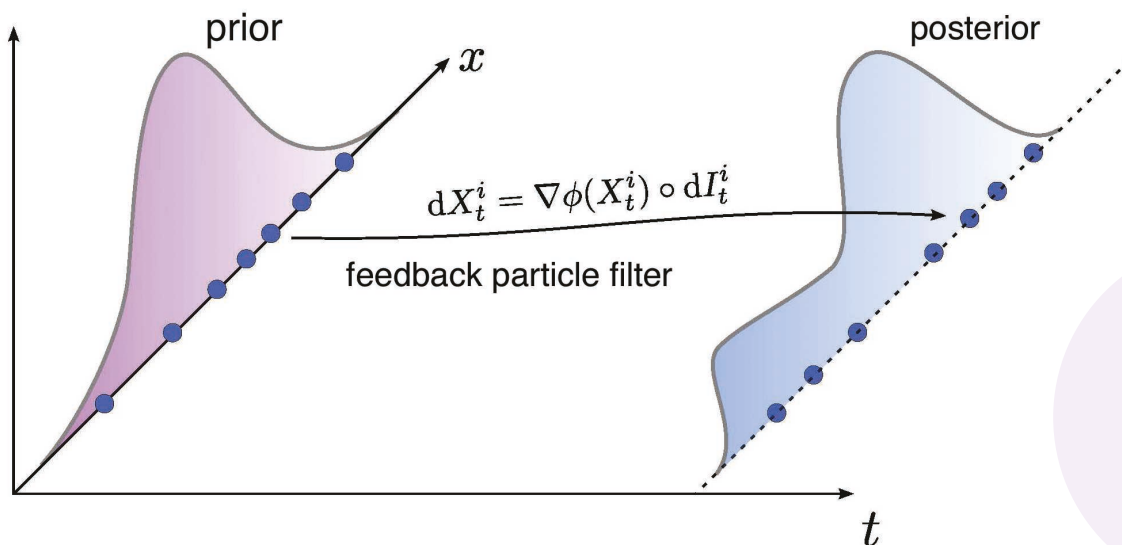




Optimal Transportation Methods in Nonlinear Filtering

THE FEEDBACK PARTICLE FILTER

AMIRHOSSEIN TAGHVAEI
and PRASHANT G. MEHTA

How data became one of the most powerful tools to fight an epidemic is a question that a recent (10 June 2020) *The New York Times* article poses in its title. Indeed, the spread of COVID-19 involves dynamically evolving hidden data (for example, the number of infected people, the number of

asymptomatic people) that must be deduced from noisy and partially observed data (for example, the number of daily deaths, the number of daily hospitalizations, and the number of daily positive tests). The underlying mathematics for posing and solving this and several other partially observed dynamic problems is familiar to control theorists.

A mathematical abstraction of these types of problems commonly involves the definition of two stochastic

Digital Object Identifier 10.1109/MCS.2021.3076391
Date of current version: 19 July 2021



processes, (X, Z) , where in a continuous-time setting $X = \{X_t \in \mathbb{S} : t \geq 0\}$ is the hidden signal process and $Z = \{Z_t \in \mathbb{O} : t \geq 0\}$ is the observed or measured process. For the sake of exposition, the state space $\mathbb{S}$ and the observation space $\mathbb{O}$ are assumed to be the Euclidean spaces, and the two processes are modeled as solutions of a stochastic differential equation (SDE):

$$dX_t = a(X_t)dt + \sigma(X_t)dB_t, \quad X_0 \sim (\text{prior}), \quad (1)$$

$$dZ_t = h(X_t)dt + dW_t, \quad Z_0 = 0, \quad (2)$$

where $a(\cdot)$, $\sigma(\cdot)$, $h(\cdot)$ are given smooth functions of their arguments, and the signal (or process) noise $B = \{B_t : t \geq 0\}$ and the measurement (or observation) noise $W = \{W_t : t \geq 0\}$ are assumed to be independent Wiener processes (w.p.). For example, in the models of disease spread, Z_t may indicate the number (cumulative) of positive tests up to time t . In this case, $Z_{t_2} - Z_{t_1}$ is the number (increment) of positive tests during the time interval $[t_1, t_2]$, and dZ_t can be thought of as the infinitesimal increment over the infinitesimal time dt . Given the models for the stochastic processes (X, Z) , the mathematical problem of stochastic filtering is to estimate the conditional distribution of the state X_t , given the observations up to time t . The conditional distribution $P(X_t | \mathcal{Z}_t)$ is referred to as the posterior distribution, where $\mathcal{Z}_t$ is the time history (filtration) of the observations up to time t .

There are many solution approaches under different modeling assumptions. The most classical of these approaches is the method of least squares, which was invented at the turn of the 19th century. It remains popular in identification of the (static) model parameters (see [1] for an application of these methods to parameter estimation in disease modeling). For the dynamic case, when the models are linear [that is, $a(x) = Ax$, $\sigma(x) = \sigma$ and $h(x) = Hx$] and the distributions (of the noise processes and the prior) are Gaussian, Kalman and Bucy derived a recursive algorithm [2] known as the Kalman–Bucy filter:

$$d\hat{X}_t = \underbrace{A\hat{X}_t dt}_{\text{dynamics}} + \underbrace{K_t(dZ_t - H\hat{X}_t dt)}_{\text{control}},$$

where $\hat{X}_t := E(X_t | \mathcal{Z}_t)$ is the conditional mean and K_t is the Kalman gain. Each of the two terms on the right-hand side have an intuitive explanation. The first term accounts for the effect of the dynamics due to the signal model. The second term implements the effect of conditioning because of the most recent observation (increment) dZ_t . The second term is referred to as the correction or the Bayes' update step of the Kalman filter. It is remarkable that the Bayes' update formula in the Kalman filter takes the form of a feedback control law, where

$$[\text{control}] = [\text{gain}] \cdot [\text{error}]$$

and

$$[\text{error}] = [\text{Observation}] - [\text{prediction}].$$

Note that $H\hat{X}_t dt$ is the filter prediction of the new observation dZ_t . The formula is so simple that it can and should be a part of any introductory undergraduate controls class, as an example of proportional gain feedback control law! Of course, this straightforward formula has had an enormous impact in many applications (such as target tracking and surveillance, air traffic management, weather surveillance, ground mapping, geophysical surveys, remote sensing, autonomous navigation, and robotics).

The Kalman filter has many extensions, for example, to problems involving additional uncertainties in the signal and the observation models. The resulting algorithms are referred to as the interacting multiple model (IMM) [3] and the probabilistic data association (PDA) filters [4], respectively. In the PDA filter, the Kalman gain is allowed to vary based on an estimate of the instantaneous uncertainty in the observations. In the IMM filter, multiple Kalman filters are run in parallel, and their outputs combined to form an estimate. Arguably, the structural aspects of the Kalman filter have been as important as the algorithm itself in the design, integration, testing, and operation of the overall system. As a simple illustration of this, consider, for example, the Kalman filter gain. The gain is known to scale proportionally to the signal-to-noise ratio of the observations. In practice, the gain is often tuned or adapted in an online manner to trade off performance for robustness. Without such structural features, it is a challenge to create scalable, cost-effective, robust solutions.

Summary

A feedback particle filter (FPF) is a Monte Carlo algorithm used to approximate the solution of a stochastic filtering problem. In contrast to conventional particle filters, the Bayesian update step in the FPF is implemented via a mean-field type feedback control law.

The objective of this article is to situate the development of the FPF within the framework of optimal transportation theory. Starting from the simplest setting of the Bayes' update formula, a coupling viewpoint is introduced to construct particle filters. It is shown that the conventional importance sampling resampling particle filter implements an independent coupling. Design of optimal couplings is introduced first for the Gaussian settings and subsequently extended to derive the FPF algorithm. The final half of the article provides a review of some of the salient aspects of the FPF algorithm, including the feedback structure, algorithms for gain function design, and a comparison with conventional particle filters. The comparison serves to illustrate the benefit of feedback in particle filtering.

A limitation of the Kalman filter is that it gives an exact solution only in the linear Gaussian settings. Beginning in the early 1990s (spurred, in part, by computational advances), simulation-based Monte Carlo (MC) algorithms became popular for the purposes of numerically approximating the posterior distribution in more general settings. These classes of algorithms, referred to as particle filters, approximate the posterior distribution using a population of N particles $\{X_t^i : t \geq 0, 1 \leq i \leq N\}$. One can interpret each of the particles as independent samples drawn from the posterior. Alternatively, one can interpret the empirical distribution of the population as approximating the posterior distribution.

Like the Kalman filter, the particle filter is also an recursive algorithm. The signal model is used to simulate the effect of the dynamics. The Bayesian update step is implemented using techniques such as importance sampling and resampling. Although these techniques are easily described, they bear little resemblance to the feedback control structure of the Kalman filter.

The focus of this article is on the feedback particle filter (FPF) algorithm (see “Summary”). An FPF represents an exact solution of the nonlinear non-Gaussian filtering problem in (1) and (2), where the state space $\mathbb{S}$ can in general be a Riemannian manifold. In applications, Euclidean spaces and matrix Lie groups are most common. The distinguishing feature of the FPF is that the Bayesian update step is implemented via a feedback control law of the form

$$[\text{control}] = [\text{gain}] \cdot [\text{error}],$$

where

$$[\text{error}] = [\text{Observation}] - \left(\frac{1}{2} [\text{Part.predict.}] + \frac{1}{2} [\text{Pop.predict.}] \right).$$

The terms [Part. predict.] and [Pop. predict.] refer to the prediction—regarding the next [Observation] “ dZ_t ”—as made by the particle and the population, respectively (see Figure 1). Because the control for each particle depends also on the population (and thus the empirical distribution), this is an example of a mean-field-type control law [5], [6]. For the linear Gaussian problem in the Euclidean state space, the [gain] of the FPF is exactly the Kalman gain. In non-Gaussian settings, the gain solves a certain linear partial differential equation (PDE), which is known as the weighted Poisson equation. The exact formula for the FPF control and gain appears in the main body of the article.

At the turn of the decade (beginning in 2010), the FPF algorithm was introduced by our research group at the University of Illinois [7]–[9]. The algorithm can be viewed as a modern extension to the Kalman filter, a viewpoint stressed in [10]. Like the Kalman filter, the FPF is easily extended to handle additional uncertainties in signal and measurement models. These extensions, namely, the

PDA-FPF and the IMM FPF, appear in the prior works [11] and [12].

From a historical perspective, the FPF is a part of a broader class of exact and approximate interacting particle algorithms, specifically, the ensemble Kalman filter (EnKF), which is widely used for data assimilation in weather prediction and other types of geophysical applications [13]. Closely related to and predating the work on the FPF, the first interacting particle representation of the continuous-time nonlinear filtering problem in (1) and (2) appears in [14]. In linear Gaussian settings, the update formula for the FPF is known as the square root form of the EnKF [15], [16].

The objective of this article is to situate the development of the FPF and its related controlled interacting particle system algorithms (for example, the EnKF) within the framework of optimal transportation theory. The key notion is that of “coupling” between two distributions—prior and posterior in the Bayesian settings of this article. Optimal transportation theory is then applied to design the optimal coupling. In practice, this requires a solution of certain PDEs, such as the Poisson equation that arises in the FPF algorithm. The coupling viewpoint has several advantages, which are described in the main body of this article.

THE COUPLING VIEWPOINT

The heart of any simulation-based recursive particle filter algorithm is the Bayes’ update formula

$$[\text{posterior}] \propto [\text{likelihood}] \cdot [\text{prior}].$$

The notation $p_0(\cdot)$, $p_1(\cdot)$, and $\ell(\cdot)$ is used to denote the prior, posterior, and likelihood distributions, respectively. The expressions for these in the linear Gaussian example appear

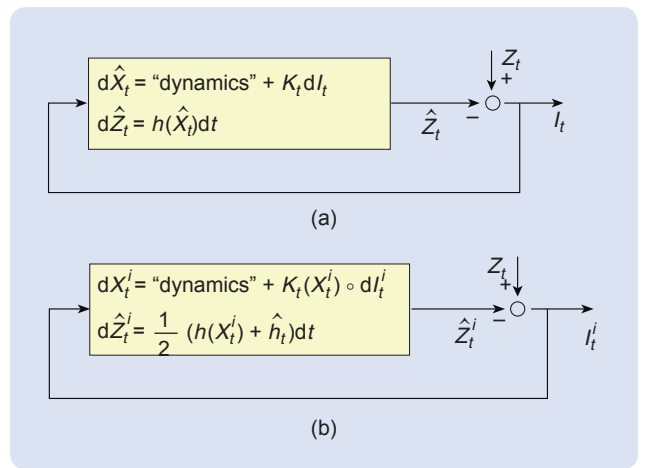


FIGURE 1 The feedback control structure of a (a) Kalman filter (KF) and (b) feedback particle filter (FPF). $\hat{X}_t$ in the KF is the estimate (conditional mean) of the hidden state. X_t^i in the FPF is a sample from the posterior (conditional distribution) of the hidden state. In both algorithms, the Bayesian update is implemented via a gain times error control law.

The distinguishing feature of the FPF is that the Bayesian update step is implemented via a feedback control law.

in the “Optimal Coupling for Gaussian Distributions” section. In any particle filter algorithm, one also needs to simulate the effect of dynamics. This step is straightforward using the signal model directly. Therefore, the focus of this section is only on the update formula. One of the challenges in simulation settings is that an analytic expression of the prior distribution is not available. Instead, the prior is approximated in terms of N independent samples $\{X_0^i : 1 \leq i \leq N\}$:

$$p_0(x) \approx \frac{1}{N} \sum_{i=1}^N \delta_{X_0^i}(x),$$

where δ_z is the Dirac distribution at $z \in \mathbb{S}$. The expression on the right-hand side is referred to as the empirical distribution of the population. Alternatively, one can think of X_0^i as an independent sample drawn from the prior. In a numerical implementation of the update formula, the problem is to convert the sample of N particles $\{X_0^i : 1 \leq i \leq N\}$ from the prior distribution $p_0(\cdot)$ to a sample of N particles $\{X_1^i : 1 \leq i \leq N\}$ from the posterior distribution $p_1(\cdot)$. The algorithmic problem is depicted in Figure 2 and is expressed as

Input: samples $\{X_0^i : 1 \leq i \leq N\} \sim p_0$, likelihood function ℓ ,
Output: samples $\{X_1^i : 1 \leq i \leq N\} \sim p_1$.

The task of converting samples from one distribution $p_0(\cdot)$ to samples from another distribution $p_1(\cdot)$ is viewed as the problem of finding a coupling $\pi(\cdot, \cdot)$ [17]–[19]. By definition, a coupling is any joint probability distribution that satisfies the marginal constraints $\int \pi(x, x') dx' = p_1(x)$ and $\int \pi(x, x') dx = p_0(x')$. It is convenient to express $\pi(x, x') = \mathcal{T}(x|x')p_0(x')$, where $\mathcal{T}(\cdot|\cdot)$ is referred to as the transition kernel. Once the coupling is at hand, new samples are generated using the transition kernel.

Given this viewpoint, the MC algorithm simulates the following stochastic update law for the system of particles:

$$X_1^i \sim \mathcal{T}(\cdot|X_0^i). \quad (3)$$

This means that a new sample X_1^i is generated by sampling from the distribution $\mathcal{T}(\cdot|X_0^i)$. The sampling algorithm (3) ensures that if the probability distribution of X_0^i is p_0 , then the probability distribution of X_1^i is p_1 . The associated algorithmic task is expressed as

Input: samples $\{X_0^i : 1 \leq i \leq N\} \sim p_0$, likelihood function ℓ ,
Output: coupling between p_0 and p_1 .

Sequential Importance Resampling Particle Filter

There are infinitely many couplings between two distributions. The simplest possible choice is an independent coupling, where $\pi(x, x') = p_1(x)p_0(x')$. For independent coupling, the transition kernel $\mathcal{T}(x|x') = p_1(x)$. The sequential importance resampling (SIR) particle filters [20], [21] numerically implement the independent coupling in two steps:

- 1) A weighted distribution of the particles is first formed according to $\sum_{i=1}^N w_i \delta_{X_0^i}$, where the weights $w_i = \ell(X_0^i) / \sum_{j=1}^N \ell(X_0^j)$. The weighted distribution is an approximation of the posterior distribution p_1 . This step is called importance sampling.
- 2) N particles are independently sampled from the weighted distribution $X_1^i \sim \sum_{i=1}^N w_i \delta_{X_0^i}$ by sampling from a multinomial distribution with parameter vector $(N, \{w_i\}_{i=1}^N)$. This step is called resampling.

Theoretically, it is shown that the empirical approximation with particles becomes exact in the limit as $N \rightarrow \infty$ with error rate $O(N^{-1/2})$ [22], [23]. However, both empirically and theoretically, it was discovered that particle filters can suffer from a large simulation variance that worsens as the dimension of the problem increases [24]–[26]. To maintain the same mean-squared error, a particle filter requires a number of particles that scales exponentially with the dimension. This issue is referred to as the *curse of dimensionality* (CoD). The issues with the stochastic independent coupling in SIR filters has motivated the investigation of other forms of coupling that are also optimal in some sense [27], [28]. In the simulation literature, this is referred to as the design of proposal distributions.

Optimal Transport Coupling

Optimal transportation theory provides a principled approach for identifying a coupling. Given two distributions, p_0 and p_1 , the optimal transportation problem is

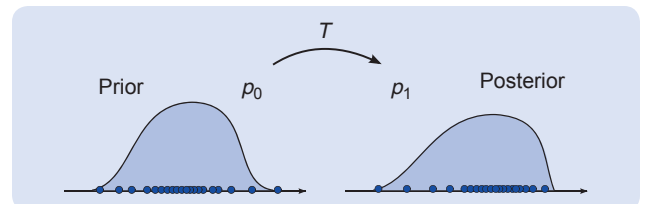


FIGURE 2 The coupling viewpoint of the filtering problem. The task of a particle filter is to convert a sample of N particles from the prior distribution to a sample of N particles from the posterior distribution. This task is viewed as finding a coupling between the prior and the posterior distributions.

Optimal transportation theory provides a principled approach for identifying a coupling.

$$\min_{\pi \in \Pi_{p_0, p_1}} \mathbb{E}_{(X_0, X_1) \sim \pi} [|X_0 - X_1|^2], \quad (4)$$

where Π_{p_0, p_1} is the set of all the couplings with marginals fixed to p_0 and p_1 . The optimal cost is referred to as the L^2 -Wasserstein distance between p_0 and p_1 [29]. The optimization problem (4) is known as the Kantorovich formulation of the optimal transportation problem. By a famous result due to Brenier, if the two distributions admit density with respect to a Lebesgue measure, then the optimal coupling is unique and deterministic of the form $\pi(x, x') = \delta_{\nabla \Phi(x)}(x) p_0(x')$, where Φ is a convex function [29, Th. 2.12]. The function Φ is obtained by solving the Monge–Ampère PDE [30]. A numerical solution to the Monge–Ampère PDE based only on the samples is a challenging problem. In the following, these results are described for the special case when the prior and the posterior distributions are both Gaussian.

Optimal Coupling for Gaussian Distributions

In Gaussian settings, the prior and the likelihood are both Gaussian distributions:

$$p_0(x) \propto \exp\left(-\frac{1}{2}(x - m_0)^\top \Sigma_0^{-1}(x - m_0)\right),$$

$$l(x) \propto \exp\left(-\frac{|y - Hx|^2}{2}\right).$$

In this case, a simple completion of square helps show that the posterior is also a Gaussian distribution

$$p_1(x) \propto \exp\left(-\frac{1}{2}(x - m_1)^\top \Sigma_1^{-1}(x - m_1)\right).$$

This yields the following update formula for the mean and the variance:

$$m_1 = m_0 + K(Y - Hm_0), \quad \Sigma_1 = \Sigma_0 - KH\Sigma_0,$$

where $K = \Sigma_0 H^\top (H\Sigma_0 H^\top + I)^{-1}$ is the Kalman gain. This is the update formula for the discrete-time Kalman filter.

The coupling design problem is to couple the Gaussian prior p_0 and the Gaussian posterior p_1 . The optimal coupling [solution to the optimal transportation problem (4)] between two Gaussian distributions is explicitly known and is an affine map of the form

$$T(x) = F(x - m_0) + m_1, \quad (5)$$

where F is the (unique) symmetric matrix solution to the matrix equation $F\Sigma_0 F = \Sigma_1$. The explicit form of the solution is

$$F = \Sigma_0^{-\frac{1}{2}} \left(\Sigma_0^{-\frac{1}{2}} \Sigma_1 \Sigma_0^{-\frac{1}{2}} \right)^{\frac{1}{2}} \Sigma_0^{-\frac{1}{2}}.$$

Note that if $X \sim \mathcal{N}(m_0, \Sigma_0)$ then $T(X) \sim \mathcal{N}(m_1, \Sigma_1)$ because: 1) the mean $\mathbb{E}[T(X)] = m_1$, 2) the variance $\mathbb{E}[(T(X) - m_1)(T(X) - m_1)^\top] = F\Sigma_0 F = \Sigma_1$, and 3) an affine transformation of a Gaussian random variable is again Gaussian. The optimal transport map (5) yields the following algorithm for sampling:

$$X_1^i = T(X_0^i) = F(X_0^i - m_0) + m_1.$$

Given $X_0^i \sim p_0$, $X_1^i \sim p_1$. The optimal coupling depends on the statistics of both the prior and posterior distributions. In a simulation-based setting, one has only a population of particles $\{X_0^i : 1 \leq i \leq N\}$. The transport map must also be approximated from the particles. One such approximation is

$$\text{particle update: } X_1^i = F^{(N)}(X_0^i - m_0^{(N)}) + m_0^{(N)} + K^{(N)}(Y - Hm_0^{(N)}), \quad (6)$$

where $m_0^{(N)} = (1/N)\Sigma_{i=1}^N X_0^i$ is an empirical approximation of the mean m_0 , $\Sigma_0^{(N)} = (1/(N-1))\Sigma_{i=1}^N (X_0^i - m_0^{(N)})(X_0^i - m_0^{(N)})^\top$ is an empirical approximation of the variance, $F^{(N)}$ is the unique symmetric matrix solution to the matrix equation $F^{(N)}\Sigma_0^{(N)}F^{(N)} = \Sigma_0^{(N)} - K^{(N)}H\Sigma_0^{(N)}$, and $K^{(N)} = \Sigma_0^{(N)}H^\top(H\Sigma_0^{(N)}H^\top + I)^{-1}$.

The update formula (6) is compared to the discrete-time EnKF update

$$\text{particle update (EnKF): } X_1^i = X_0^i + K^{(N)}(Y - HX_0^i + W^i), \quad (7)$$

where $\{W^i : 1 \leq i \leq N\}$ are independent copies of the observation noise. The EnKF update is an example of a stochastic coupling in contrast to the deterministic optimal coupling (6). The EnKF update does not require solving for $F^{(N)}$. This makes it simpler to implement numerically. However, the presence of noise W^i in the update law introduces an additional source of error in any finite- N implementation [31].

Besides (7), there are several other forms of the EnKF update. One particular update, which has been crucial in successful application of EnKF in geoscience, is the ensemble square-root Kalman filter (EnSRKF) [32]. This update is

an example of a deterministic coupling that also avoids the need to compute $F^{(N)}$ [17, Sec. 7.1]. In the sequel, the continuous-time version of the EnSRKF arises as a special case of the FPF.

The systems in (6) and (7) are both examples of interacting particle systems. The interaction arises because of the terms involving the empirical quantities $m_0^{(N)}$ and $\Sigma_0^{(N)}$. In the limit as $N \rightarrow \infty$, these converge to their respective statistics, and the particles become independent of each other [33]–[37]. This phenomenon is referred to as the propagation of chaos [38], [39]. The ($N = \infty$) limit is referred to as the mean-field limit, and for (6), it is identified by a single equation:

$$\text{mean-field update: } \bar{X}_1 = F(\bar{X}_0 - m_0) + m_0 + K(Y - Hm_0). \quad (8)$$

As a practical matter, one first designs a coupling (5) that is used to define the mean-field process (8). Subsequently, the mean-field terms are approximated empirically to define a particle system (6). A finite- N implementation of the particle system yields a practical algorithm to solve the filtering task.

A synopsis of this section is presented in “Summary of the Linear Gaussian Example.” An excellent exposition of the coupling approach to discrete-time filtering appears in [17] and [18]. Other examples of couplings include an approximation of the Rosenblatt transport map [40], [41] and Gibbs flow [42].

FEEDBACK PARTICLE FILTER

The coupling viewpoint is employed next to introduce and describe the FPF algorithm. An FPF is an MC solution to the continuous-time nonlinear filtering problem in (1) and (2). A construction of the FPF proceeds in the following two steps:

- 1) Construct a stochastic process $\bar{X} = \{\bar{X}_t \in \mathbb{S} : t \geq 0\}$ such that the conditional distribution (given $\mathcal{Z}_t$) of $\bar{X}_t$ is equal to the posterior distribution of X_t .
- 2) Simulate N stochastic processes $\{X_t^i \in \mathbb{S} : t \geq 0, 1 \leq i \leq N\}$ to empirically approximate the conditional distribution of $\bar{X}_t$.

$$\underbrace{\mathbb{E}[f(X_t) | \mathcal{Z}_t]}_{\text{exactness condition}} \stackrel{\text{Step 1}}{=} \mathbb{E}[f(\bar{X}_t) | \mathcal{Z}_t] \stackrel{\text{Step 2}}{\approx} \frac{1}{N} \sum_{i=1}^N f(X_t^i)$$

for all bounded functions f . The process $\bar{X}$ is referred to as the mean-field process, and the N processes are referred to as particles. The construction ensures that the filter is exact in the mean-field ($N = \infty$) limit.

Mean Process

The mean-field process $\bar{X}$ is modeled as a solution of a controlled SDE:

$$d\bar{X}_t = \underbrace{a(\bar{X}_t)dt + \sigma(X_t)d\bar{B}_t}_{\text{dynamics}} + \underbrace{u_t(\bar{X}_t)dt + K_t(\bar{X}_t)dZ_t}_{\text{control}}, \quad \bar{X}_0 \stackrel{d}{=} X_0, \quad (9)$$

Summary of the Linear Gaussian Example

Prior: Gaussian $\mathcal{N}(m_0, \Sigma_0)$

Observation model: $Y = HX + W$

Likelihood function: $l(x) = \exp\left(-\frac{|y - Hx|^2}{2}\right)$

Posterior: Gaussian $\mathcal{N}(m_1, \Sigma_1)$

Optimal transport map: $T(x) = F(x - m_0) + m_1$

Mean-field process: $\bar{X}_1 = F(\bar{X}_0 - m_0) + m_0 + K(Y - Hm_0)$

Particle system: $X_1^i = F^{(N)}(X_0^i - m_0^{(N)}) + m_0^{(N)} + K^{(N)}(Y - Hm_0^{(N)})$,
for $i = 1, \dots, N$.

where $\bar{B} = \{\bar{B}_t : t \geq 0\}$ is a w.p., independent of $\bar{X}_0$. The first two terms in (9) are a copy of the dynamics in (1). The other two terms are control laws (transition kernels) that must be designed to implement the filtering update step: The mathematical control objective is to design $\{u_t(\cdot) : t \geq 0\}$ and $\{K_t(\cdot) : t \geq 0\}$ such that the conditional distribution (given $\mathcal{Z}_t$) of $\bar{X}_t$ is equal to the posterior distribution of X_t . The control is regarded as implementing the transition kernel of a coupling. As in the simpler discrete-time setting in the preceding section, there are infinitely many couplings and associated transition kernels that satisfy the exactness criteria. This is not surprising. The exactness condition specifies only the marginal distribution of $\bar{X}_t$ at times $t \geq 0$. This is clearly not enough to uniquely identify a stochastic process (for instance, the joint distributions at two time instants are not specified).

The procedure from the preceding section is suitably adapted to design the optimal coupling. The optimality criterion is the Kantorovich form (4) of the optimal transportation problem. The details appear in “Optimal Transport Construction of Stochastic Processes,” where it is shown that the optimal K_t is of the gradient form as follows: $K_t(x) = \nabla \phi_t(x)$, where ϕ_t solves the partial differential equation (PDE) $-\nabla \cdot (p_t \nabla \phi_t) = p_t(h - \hat{h}_t)$, ∇ is the gradient operator, $\nabla \cdot$ is the divergence operator, $\hat{h}_t := \mathbb{E}[h(\bar{X}_t) | \mathcal{Z}_t] = \int h(x)p_t(x)dx$, and p_t is the conditional density of $\bar{X}_t$. The optimal solution for u_t is

$$u_t(x) = -\frac{1}{2}(h(x) + \hat{h}_t)K_t(x) + \frac{1}{2}K_t \nabla K_t(x) + \xi_t(x),$$

where ξ_t is the (unique such) divergence-free vector field (that is, $\nabla \cdot (p_t \xi_t) = 0$), such that u_t is of a gradient form. An intuitive explanation of the three terms is as follows. The first term is gain $K_t(x)$ times the average of the particle prediction $h(x)$ and the population prediction $\hat{h}_t$. Together with the stochastic term $K_t dZ_t$, the first term yields the gain times error structure of the FPF. The second term is the so-called Wong–Zakai correction term, from which it follows that the gain times error formula is expressed in its Stratanovich form. The geometric significance of the Stratanovich form is

described after the FPF formula has been formally presented. The significance of the third term follows. The divergence-free choices of ξ_t parameterize a manifold of couplings, all of which yield the same (exact) posterior. Therefore, the choice of ξ_t affects *only* the (Wasserstein) optimality, but not the exactness property of the filter. In the FPF, ξ_t is set to zero for all $t \geq 0$. Although such a solution is optimal only in the scalar (1D) settings, it avoids the need to solve an additional PDE to compute the optimal ξ_t . The resulting algorithm is referred to as the FPF.

Feedback Particle Filter

The mean-field process $\bar{X}$ evolves according to the SDE:

$$d\bar{X}_t = \underbrace{a(\bar{X}_t)dt + \sigma(\bar{X}_t)d\bar{B}_t}_{\text{dynamics}} + \underbrace{K_t(\bar{X}_t) \circ \left(dZ_t - \frac{h(\bar{X}_t) + \hat{h}_t}{2} dt \right)}_{\text{Bayes' update: feedback control law}},$$

$$\bar{X}_0 \sim p_0, \quad (10)$$

where the symbol $\circ$ denotes the fact that the SDE is expressed in its Stratanovich form [43, Sec. 3.3]. The Itô form of the FPF includes the standard Wong–Zakai correction term that arises on account of the dependence of the gain $K_t(\cdot)$ on the state X_t [44, eq. 2]. Because the gain also depends upon the density, the interpretation of the Stratanovich form in the general case is more involved, as discussed at length in [45].

Optimal Transport Construction of Stochastic Processes

DETERMINISTIC PATH

Let $\mathcal{P}_2(\mathbb{R}^d)$ be the space of everywhere positive probability densities on $\mathbb{R}^d$ with a finite second moment. Given a smooth path $\{\rho_t \in \mathcal{P}_2(\mathbb{R}^d) : t \geq 0\}$, the problem is to construct a stochastic process $\{\bar{X}_t; t \geq 0\}$ such that the probability density of $\bar{X}_t$ (denoted as $\bar{p}_t$) equals ρ_t for all $t \geq 0$. The exactness condition is expressed as

$$\bar{p}_t = \rho_t, \quad \forall t \geq 0. \quad (S1)$$

Now there are infinitely many stochastic processes that satisfy the exactness condition. This is because the exactness condition specifies only the one-time marginal distribution, which is clearly not enough to uniquely identify the stochastic process, for example, the two-time joint distributions are not specified. A unique choice is made by prescribing an additional optimality criterion based on the optimal transportation theory. To make these considerations concrete, assume that the given path $\{\rho_t : t \geq 0\}$ evolves according to the partial differential equation (PDE)

$$\frac{\partial \rho_t}{\partial t} = \mathcal{V}(\rho_t),$$

where $\mathcal{V}(\cdot)$ is an operator (for example, the Laplacian) that acts on probability densities. This necessarily restricts the operator $\mathcal{V}$, for example, $\int \mathcal{V}(\rho)(x)dx = 0$ for all $\rho \in \mathcal{P}_2(\mathbb{R}^d)$. The following model is assumed for the process $\{\bar{X}_t : t \geq 0\}$:

$$\frac{d}{dt} \bar{X}_t = u_t(\bar{X}_t), \quad \bar{X}_0 \sim p_0, \quad (S2)$$

where $u_t(\cdot)$ is a control law that must be designed. Using the continuity equation, the exactness condition (S1) will be satisfied if

$$-\nabla \cdot (\bar{p}_t u_t) = \mathcal{V}(\bar{p}_t), \quad \forall t \geq 0. \quad (S3)$$

The nonuniqueness issue is now readily seen. The first-order PDE (S3) admits infinitely many solutions. A unique solution u_t is selected by optimizing the coupling between $\bar{X}_t$ and $\bar{X}_{t+\Delta t}$ in the limit as $\Delta t \rightarrow 0$. The leading term in the transportation cost $\mathbb{E}[|X_{t+\Delta t} - X_t|^2]$ is of order $O(\Delta t^2)$ whereby

$$\lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t^2} \mathbb{E}[|X_{t+\Delta t} - X_t|^2] = \int_{\mathbb{R}^d} |u_t(x)|^2 \bar{p}_t(x) dx.$$

Therefore, for each fixed $t \in [0, 1]$, the control law u_t is obtained by solving the constrained optimization problem

$$\min_{u_t} \int_{\mathbb{R}^d} |u_t(x)|^2 \bar{p}_t(x) dx, \quad \text{subject to} \quad -\nabla \cdot (\bar{p}_t u_t) = \mathcal{V}(\bar{p}_t). \quad (S4)$$

The cost is the infinitesimal form of the L^2 -Wasserstein distance, and the constraint expresses the exactness condition. By a standard calculus of variation argument, the solution of (S4) is obtained as $u_t^* = \nabla \phi_t$, where ϕ_t solves the second-order PDE $-\nabla \cdot (\bar{p}_t \nabla \phi_t) = \mathcal{V}(\bar{p}_t)$. The resulting stochastic process $\bar{X}_t$ evolves according to

$$\frac{d\bar{X}_t}{dt} = \nabla \phi_t(\bar{X}_t), \quad \bar{X}_0 \sim p_0,$$

$$\phi_t \text{ solves the PDE } -\nabla \cdot (\bar{p}_t \nabla \phi_t) = \mathcal{V}(\bar{p}_t).$$

As a concrete example, suppose that the given path is a solution of the heat equation $\partial \rho_t / \partial t = \Delta \rho_t$. So $\mathcal{V}(\cdot)$ is the Laplacian operator. The solution of the second-order PDE is easily obtained as $\phi_t = \log(\bar{p}_t)$. The optimal transport process $\bar{X}_t$ then evolves according to

$$\frac{d}{dt} \bar{X}_t = -\nabla \log(\rho_t(\bar{X}_t)), \quad \bar{X}_0 \sim p_0.$$

This process should be compared to the stochastic differential equation (SDE)

$$dX_t = dB_t, \quad X_0 \sim p_0, \quad (S5)$$

The gain function is $K_t(x) := \nabla \phi_t(x)$, where ϕ_t is the solution of the Poisson equation

$$\text{Poisson equation: } -\frac{1}{p_t(x)} \nabla \cdot (p_t(x) \nabla \phi_t(x)) = (h(x) - \hat{h}_t), \quad \forall x \in \mathbb{R}^d. \quad (11)$$

The operator on the left-hand side of (11) is the *probability-weighted Laplacian*. It is denoted as $\Delta_\rho := (1/\rho) \nabla \cdot (\rho \nabla)$, where at time t the probability density ρ is the conditional density p_t . Equation (11) is referred to as the Poisson equation of nonlinear filtering [46]. The Stratanovich form of the update formula provides for an intrinsic (that is, coordinate independent) description of the filter. It was shown that the FPF is an exact filter, not

only in the Euclidean settings, but also when the state space $\mathbb{S}$ is a Riemannian manifold (for example, a matrix Lie group [47]). For a manifold with boundaries, the Poisson equation is supplemented with the Neumann boundary conditions.

Particles

A finite- N algorithm is obtained by empirically approximating the mean-field control law. The particles $\{X_t^i : t \geq 0, 1 \leq i \leq N\}$ evolve according to

$$dX_t^i = a(X_t^i)dt + \sigma(X_t^i)dB_t^i + K_t^{(N)}(X_t^i) \circ \left(dZ_t - \frac{h(X_t^i) + \hat{h}_t^{(N)}}{2} dt\right), \quad X_0^i \stackrel{\text{i.i.d.}}{\sim} p_0,$$

where $\{B_t : t \geq 0\}$ is a w.p. The SDE (S5) is a well-known stochastic coupling whose one-point marginal evolves according to the solution of the heat equation.

Stochastic Path

In the filtering problem, the path of the posterior probability densities is stochastic (because it depends upon random observations $\{Z_t : t \geq 0\}$). Therefore, the discussion in the preceding section is not directly applicable. Suppose that the stochastic path $\{p_t(\cdot) \in \mathcal{P}_2(\mathbb{R}^d) : t \geq 0\}$ is governed by a stochastic PDE

$$dp_t = \mathcal{H}(p_t)dt,$$

where $\mathcal{H}(\cdot)$ is an operator that acts on probability densities, and $\{t : t \geq 0\}$ is a w.p. Consider the following SDE model:

$$d\bar{X}_t = u_t(\bar{X}_t)dt + K_t(\bar{X}_t)dI_t, \quad \bar{X}_0 \sim p_0,$$

where, compared to the deterministic form of (S2), an additional stochastic term is now included. The problem is to identify control laws $u_t(\cdot)$ and $K_t(\cdot)$ such that the conditional distribution of $\bar{X}_t$ equals p_t . This exactness condition, counterpart of (S3), is now

$$-\nabla \cdot (\bar{p}_t K_t) = \mathcal{H}(\bar{p}_t), \quad (\text{S6a})$$

$$-\nabla \cdot (\bar{p}_t u_t) + \frac{1}{2}(\nabla \cdot (\bar{p}_t K_t) K_t + \bar{p}_t K_t \nabla K_t) = 0. \quad (\text{S6b})$$

These equations are obtained by writing the time evolution of the conditional probability density of $\bar{X}_t$ [44, Prop. 1]. As in the deterministic setting, the solution is not unique. The unique optimal control law is obtained by requiring that the coupling between $\bar{X}_t$ and $\bar{X}_{t+\Delta t}$ is optimal in the limit as $\Delta t \rightarrow 0$. In contrast to the deterministic setting, the leading term in the transportation cost $\mathbb{E}[|X_{t+\Delta t} - X_t|^2]$ is the order of $O(\Delta t)$, whereby

$$\lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \mathbb{E}[|X_{t+\Delta t} - X_t|^2] = \int_{\mathbb{R}^d} |K_t(x)|^2 \bar{p}_t(x) dx. \quad (\text{S7})$$

Therefore, for each fixed $t \in [0, 1]$, the control law K_t is obtained by solving the constrained optimization problem

$$\min_{K_t} \int_{\mathbb{R}^d} |K_t(x)|^2 \bar{p}_t(x) dx, \quad \text{subject to } -\nabla \cdot (\bar{p}_t K_t) = \mathcal{H}(\bar{p}_t). \quad (\text{S8})$$

As before, the solution of the optimization problem (S8) is given by $K_t^* = \nabla \phi_t$, where ϕ_t solves the second-order PDE $-\nabla \cdot (\bar{p}_t \nabla \phi_t) = \mathcal{H}(\bar{p}_t)$. It remains to identify the control law for u_t . For this purpose, the second-order term in the infinitesimal Wasserstein cost is used:

$$\lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} (\mathbb{E}[|X_{t+\Delta t} - X_t|^2] - \Delta t \int_{\mathbb{R}^d} |K_t^*|^2 \bar{p}_t dx) = \int_{\mathbb{R}^d} |u_t|^2 \bar{p}_t dx.$$

The right-hand side is minimized subject to the constraint (S6b). Remarkably, the optimal solution is expressed as

$$u_t^* = -\frac{1}{2\bar{p}_t} \mathcal{H}(\bar{p}_t) \nabla \phi_t + \frac{1}{2} \nabla^2 \phi_t \nabla \phi_t + \xi_t,$$

where ξ_t is the (unique such) divergence-free vector field (that is, $\nabla \cdot (p_t \xi_t) = 0$), such that u_t is of a gradient form. The resulting optimal transport process is

$$d\bar{X}_t = \nabla \phi_t(\bar{X}_t) \circ \left(dI_t - \frac{1}{2\bar{p}_t} \mathcal{H}(\bar{p}_t) dt\right) + \xi_t(\bar{X}_t) dt, \quad \bar{X}_0 \sim p_0. \quad (\text{S9})$$

It is also readily shown that the process $\{\bar{X}_t : t \geq 0\}$ is exact for any choice of divergence-free vector field ξ_t . The most convenient choice is obtained by simply setting $\xi_t \equiv 0$. The resulting filter is exact and also (infinitesimally) optimal to the first order [see (S7)].

For the special case of the nonlinear filtering problem, $\mathcal{H}(p) = (h - \hat{h})p$ where $\hat{h} = \int h(x)p(x)dx$ and $dI_t = dZ_t - \hat{h}_t dt$ is the increment of the innovation process. The optimal transport stochastic process (S9) is then given by the Stratanovich form

$$d\bar{X}_t = \nabla \phi_t(\bar{X}_t) \circ \left(dZ_t - \frac{1}{2}(h + \hat{h}_t)dt\right) + \xi_t(\bar{X}_t)dt, \quad \bar{X}_0 \sim p_0.$$

The control law in (10) represents the particular suboptimal choice $\xi_t \equiv 0$.

Benefit of Feedback

In this section, a simple example is used to illustrate the phenomena of the curse of dimensionality (CoD) in particle filters (PFs). A comparison with the feedback PF (FPF) is provided to see how the curse is mitigated using feedback control. The example considered is

$$dX_t = 0, \quad X_0 \sim \mathcal{N}(0, \sigma_0^2 I_d), \quad (\text{S10a})$$

$$dZ_t = X_t dt + \sigma_w dW_t, \quad (\text{S10b})$$

for $t \in [0, 1]$, where X_t is a d -dimensional process, $\sigma_w, \sigma_0 > 0$, and I_d is a $d \times d$ identity matrix. The posterior distribution at time $t = 1$ is a Gaussian distribution $\mathcal{N}(m_1, \Sigma_1)$ with mean $m_1 = (\sigma_0^2 / (\sigma_0^2 + \sigma_w^2)) Z_1$ and variance $\Sigma_1 = (\sigma_0^2 \sigma_w^2 / (\sigma_0^2 + \sigma_w^2)) I_d$. Consider the following MC approaches to approximate the posterior distribution:

- 1) Sequential importance resampling PF: Sample $\{X_0^i : 1 \leq i \leq N\}$ from the initial distribution. Form the weighted distribution and generate new samples from the weighted distribution.

$$X_t^i \sim \sum_{j=1}^N w_j \delta_{X_0^j}, \quad w_i = \frac{e^{-\frac{|Z_1 - X_0^i|^2}{2\sigma_w^2}}}{\mathbb{E}\left[e^{-\frac{|Z_1 - X_0^i|^2}{2\sigma_w^2}} \middle| \mathcal{Z}_1\right]}, \quad X_0^i \sim \mathcal{N}(0, \sigma_0^2 I_d). \quad (\text{S11})$$

- 2) FPF: Simulate the particles according to

$$dX_t^i = \frac{1}{\sigma_w^2} \Sigma_t^{(N)} \left(dZ_t - \frac{X_t^i + m_t^{(N)}}{2} dt \right), \quad X_0^i \sim \mathcal{N}(0, \sigma_0^2 I_d) \quad (\text{S12})$$

for $t \in [0, 1]$, where $m_t^{(N)}$ is the empirical mean of the particles, and $\Sigma_t^{(N)}$ is the empirical variance of the particles.

for $i = 1, \dots, N$, where $\{B_t^i : t \geq 0, 1 \leq i \leq N\}$ are mutually independent standard w.p., $\hat{h}_t^{(N)} := (1/N) \sum_{i=1}^N h(X_t^i)$ is the empirical approximation of $\hat{h}_t$, and $K_t^{(N)}$ is the output of an algorithm that approximates the solution to (11) at each fixed time t :

Gain function approximation:

$$K_t^{(N)} := \text{Algorithm}(\{X_t^i : 1 \leq i \leq N\}; h).$$

The notation is suggestive of the fact that the algorithm is adapted to the ensemble $\{X_t^i : 1 \leq i \leq N\}$ and the function h ; the density $p_t(x)$ is not known in an explicit manner. Two examples are presented in the sidebars to illustrate the FPF algorithm in practice. In “Benefit of Feedback,” analytical and numerical comparisons are provided to show how feedback control can help ameliorate the CoD. In “Example: Feedback Particle Filters for SIR Models,” the FPF algorithm is applied to an epidemiological SIR model. At this point, it is instructive to specialize the FPF to the linear Gaussian case where the solution of the Poisson equation is explicitly known.

The mean-squared error (MSE) in approximating the exact conditional mean m_1 of X_1 is defined as

$$\text{MSE} := \mathbb{E} \left[\frac{1}{N} \sum_{i=1}^N |X_1^i - m_1|^2 \right].$$

The following is proved in [37].

Proposition 1

Consider the filtering problem (S10) with dimension d . Then,

- 1) For the PF (S11),

$$\text{MSE}_{\text{PF}}(f) = \frac{\sigma^2}{N} \left(3(2^d) - \frac{1}{2} \right) \geq \frac{\sigma^2}{N} 2^{d+1}.$$

- 2) For the FPF (S12),

$$\text{MSE}_{\text{FPF}}(f) \leq \frac{\sigma^2}{N} (3d^2 + 2d).$$

These results are consistent with the extensive studies on importance sampling-based PFs [24], [26], [S1], [S2]. In these articles, it is shown that if $(\log N \log d)/d \rightarrow 0$ then the largest importance weight $\max_{1 \leq i \leq N} w^i \rightarrow 1$ in probability. Consequently, to prevent weight collapse, the number of particles must grow exponentially with the dimension. This phenomenon is referred to as the CoD for the PFs.

A numerical comparison of the MSE as a function of N and d is depicted in Figure S1(a) and (b). The expectation is approximated by averaging more than $M = 1000$ independent simulations. It is observed that, to have the same error, the importance-sampling-based approach requires the number of samples N to grow exponentially with the dimension d [whereas the growth using the FPF for this numerical example is $O(d^{1/2})$]. The scaling with dimension

Feedback Particle Filters for a Linear Gaussian Setting

Suppose that $a(x) = Ax$, $h(x) = Hx$, and that p_t is a Gaussian density with mean $\bar{m}_t$ and variance $\bar{\Sigma}_t$. Then the solution of the Poisson equation is known in an explicit form [44, Sec. D]. The resulting gain function is constant and equal to the Kalman gain:

$$K_t(x) \equiv \bar{\Sigma}_t H^\top, \quad \forall x \in \mathbb{R}^d. \quad (12)$$

Therefore, the mean-field process (10) for the linear Gaussian problem is

$$d\bar{X}_t = A\bar{X}_t dt + d\bar{B}_t + \bar{\Sigma}_t H^\top \left(dZ_t - \frac{H\bar{X}_t + H\bar{m}_t}{2} dt \right), \quad \bar{X}_0 \sim p_0.$$

Given the explicit form of the gain function (12), the empirical approximation of the gain is simply $K_t^{(N)} = \Sigma_t^{(N)} H^\top$, where $\Sigma_t^{(N)}$ is the empirical covariance of the particles. Therefore, the evolution of the particles is

$$dX_t^i = AX_t^i dt + dB_t^i + K_t^{(N)} \left(dZ_t - \frac{HX_t^i + Hm_t^{(N)}}{2} dt \right), \quad X_0^i \stackrel{\text{i.i.d.}}{\sim} p_0 \quad (13)$$

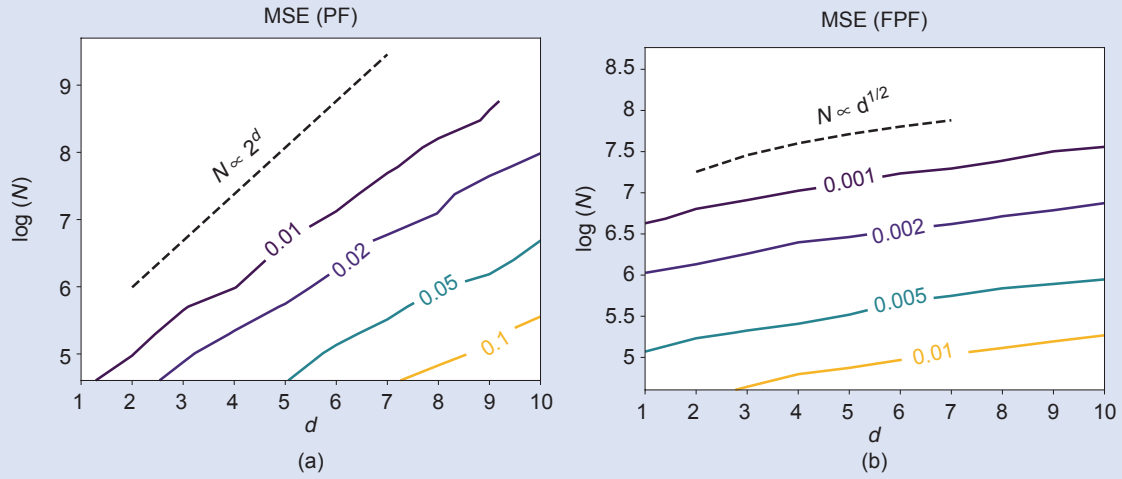


FIGURE S1 Overcoming the curse of dimensionality of particle filters (PFs). The solid lines correspond to the level sets of the mean-squared error (MSE) for the filtering problem in (S10). To have the same error, the PF requires the number of samples N to grow exponentially with the dimension d , whereas the growth using the feedback PF (FPF) for this numerical example is $O(d^{1/2})$. (a) The importance sampling PF in (S11) and (b) the FPF in (S12).

depicted in Figure S1(b) suggests that the $O(d^2)$ bound for the MSE in the linear FPF is loose. This is because of the conservative nature of approximations used in deriving the inequality [37]. The overall conclusions of the study are consistent with the other numerical results reported in the literature [S3].

REFERENCES

[S1] P. Bickel, B. Li, and T. Bengtsson, "Sharp failure rates for the bootstrap particle filter in high dimensions," in *Pushing the Limits of Con-*

temporary Statistics: Contributions in Honor of Jayanta K. Ghosh, B. Clarke and S. Ghosa, Eds. Beachwood OH: Institute of Mathematical Statistics, 2008, pp. 318–329.

[S2] C. Snyder, T. Bengtsson, P. Bickel, and J. Anderson, "Obstacles to high-dimensional particle filtering," *Monthly Weather Rev.*, vol. 136, no. 12, pp. 4629–4640, 2008. doi: 10.1175/2008MWR2529.1.

[S3] S. C. Surace, A. Kutschireiter, and J.-P. Pfister, "How to avoid the curse of dimensionality: Scalability of particle filters with and without importance weights," *SIAM Rev.*, vol. 61, no. 1, pp. 79–91, 2019. doi: 10.1137/17M1125340.

for $i = 1, \dots, N$, where $m_i^{(N)}$ is the empirical mean of the particles. The empirical quantities are computed as

$$m_i^{(N)} := \frac{1}{N} \sum_{i=1}^N X_i, \quad \Sigma_i^{(N)} := \frac{1}{N-1} \sum_{i=1}^N (X_i - m_i^{(N)})(X_i - m_i^{(N)})^\top.$$

The linear Gaussian FPF (13) is identical to the square-root form of the EnKF [16, eq. 3.3]. The main difficulty in implementing an FPF in the general nonlinear settings is the gain function approximation. The two algorithms for this problem are presented in the following section.

GAIN FUNCTION APPROXIMATION

The exact gain function is a solution of (11). In practice, this problem is solved numerically:

Input: samples $\{X^i : 1 \leq i \leq N\} \stackrel{\text{i.i.d.}}{\sim} \rho$,
observation function $h(\cdot)$,
Output: gain function $\{K(X^i) : 1 \leq i \leq N\}$,

where ρ is the (posterior) density at time t . The explicit dependence on time t is suppressed in this section. The problem is illustrated in Figure 3.

Constant Gain Approximation

The simplest approximation is the *constant gain approximation* formula, where the gain K , is approximated by its expected value (which represents the best least-square approximation of the gain by a constant). Remarkably,

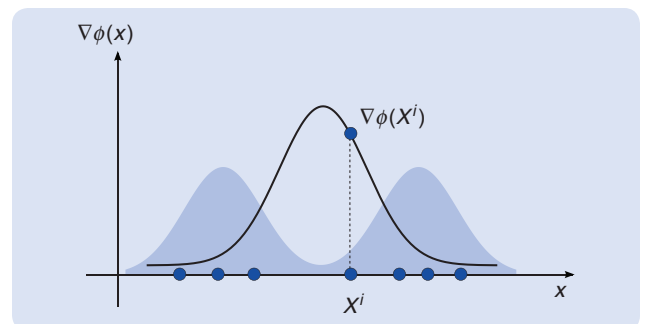


FIGURE 3 The gain function approximation problem in the feedback particle filter. The exact gain function $K(x) = \nabla\phi(x)$, where ϕ solves the Poisson equation (11); the numerical problem approximates $\nabla\phi(x)|_{x=X^i}$ using only the particles $\{X^i : 1 \leq i \leq N\}$ sampled from density ρ (depicted as the shaded region).

Example: Feedback Particle Filter for SIR Models

The basic mathematical model of epidemiological disease propagation is the susceptible-infected-recovered (SIR) ordinary differential equation model

$$\begin{aligned}\dot{S}_t &= -\beta S_t I_t, \\ \dot{I}_t &= \beta S_t I_t - \alpha I_t, \\ \dot{R}_t &= \alpha I_t,\end{aligned}$$

where S_t , I_t , and R_t are the susceptible, infected, and recovered population fractions, respectively, at time t . The parameters β and α are the transmission and recovery rate parameters, respectively. In an epidemic, one observes the number of newly infected people over a time increment (daily). For the study, this is modeled as

$$dZ_t = (\beta I_t S_t) dt + \sigma_W dW_t, \quad (S16)$$

where $W = \{W_t : t \geq 0\}$ is the standard w.p., and σ_W is the standard deviation (std dev) parameter. Given the observations, the filtering objective is to estimate the population sizes and possibly also the model parameters. In this study, the recovery rate parameter α is assumed known, while the transmission rate parameter β is estimated. In a filtering setup, this requires a model that is assumed to be of the form

$$d\beta_t = \sigma_B dB_t,$$

where $B = \{B_t : t \geq 0\}$ is a standard w.p. and σ_B is the std dev parameter.

The model and filter are simulated using the Euler discretization scheme for time integration. The simulation parameters are as follows: time-discretization step-size $\Delta t = 1$; std dev for the observation noise $\sigma_W = 0.1$; std dev for the process noise $\sigma_B = 0.1$, initial distribution $I(0) \sim \text{unif}[0, 0.1]$ and $S(0) = 1 - I(0)$; recovery rate $\alpha = 0.1$; and the transmission rate β is fixed to be 0.1 but assumed unknown to the filtering algorithm. The feedback particle filter (FPF) is simulated using $N = 100$ particles. Two gain function approximation algorithms are implemented: the constant gain and diffusion

map approximations. For the diffusion map approximation, the heuristic $\epsilon = 10 \text{med}(\{|X^i - X^j|^2; 1 \leq i, j \leq N\}) (\log(N))^{-1}$ is used, where $\text{med}(\cdot)$ denotes the statistical median. The simulation parameters and their values are tabulated in Table S1. Figure S2 depicts the numerical results for the synthetic observation data generated using the model. Although the results depicted in the figure are illustrative as an application of FPF to the SIR models, additional work is necessary for its use in prediction with real COVID-19 data. This is because of the following reasons:

- 1) The observation model (S16) is not accurate. In real-world settings, one observes only a certain unknown (and possibly a time-varying and delayed) fraction of the newly infected population. This leads to fundamental issues with the identifiability of the transmission rate parameter β [S4], [S5]. An accurate estimation of β (or the closely associated nondimensional reproduction number R_0) is important to capture the initial growth of the epidemic [S6].
- 2) The three-state SIR dynamic model is rather simplistic. This is because of several reasons: 1) The model assumes a homogeneous, well-mixed population (while in practice, there is strong evidence of heterogeneities [S7] as well as spatial network effects [S8]); 2) the model is based on the underlying assumption of Markovian transitions between the epidemiological states, which is contradicted by the experimental data on delay distributions [S9]; and 3) even in the simplistic settings of the SIR model, the transmission rate parameter β is strongly time varying. It is affected by both the individual choices (for example, mask wearing) of the large number of agents as well as population-level government mandates (for example, lockdowns).

These difficulties notwithstanding, ensemble Kalman filter-based solutions to the COVID-19 data-assimilation problem appear in [S10] and [S11]. However, much work remains to be done on this important problem of immense societal importance. In a post-COVID reality, it is not inconceivable that the surveillance and monitoring of infectious diseases such

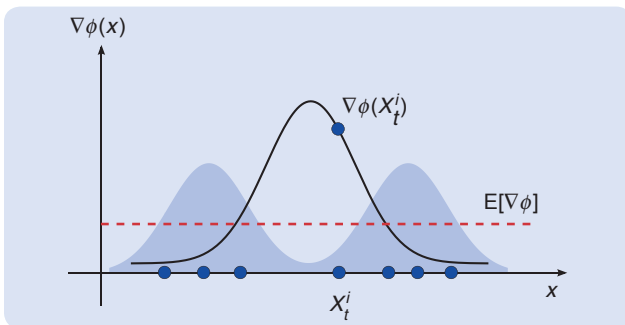


FIGURE 4 The constant gain approximation in the feedback particle filter. The gain function is approximated by its expected value according to (14).

the expected value admits a closed-form expression, which is then readily approximated empirically using the particles

Constant gain approximation:

$$\begin{aligned}\mathbb{E}[K_t(X_t) | \mathcal{Z}_t] &= \int_{\mathbb{R}^d} (h(x) - \hat{h}_t) \\ x p_t(x) dx &\approx \frac{1}{N} \sum_{i=1}^N (h(X_t^i) - \hat{h}_t^{(N)}) X_t^i.\end{aligned} \quad (14)$$

Figure 4 depicts the constant gain approximation. With the constant gain approximation, the FPF algorithm simplifies to an EnKF algorithm [10]. The constant gain

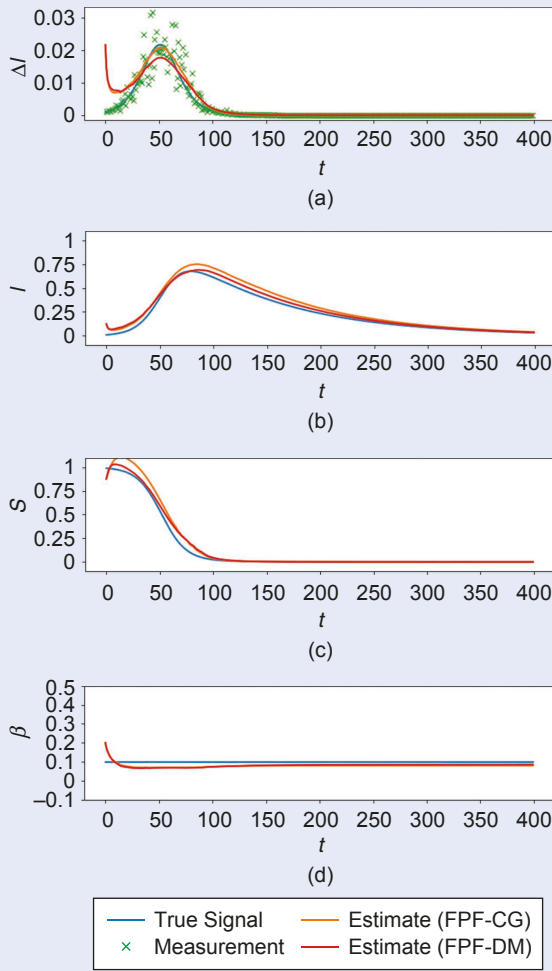


FIGURE S2 An application of the feedback particle filter (FPF) on the SIR epidemiological model. The observation is the number of new confirmed cases each day ΔI [depicted in (a)]. The size of infected population is $I(t)$, and the size of susceptible population is $S(t)$. The infection transmission rate β is assumed to be unknown and is estimated. The estimation algorithm is an FPF with constant gain (FPF-CG) approximation and an FPF with diffusion map (FPF-DM) approximation.

TABLE S1 The simulation parameters for the application of the feedback particle filter to the epidemiological example.

Parameter	Notation	Value
Time step-size	Δt	1
Observation noise	σ_w	0.1
Process noise	σ_B	0.1
Number of particles	N	100
Recovery rate	α	0.1
Transmission rate	β	0.1

as seasonal flu will be as pervasive and commonplace as weather tracking is today.

REFERENCES

- [S4] W. C. Roda, M. B. Varughese, D. Han, and M. Y. Li, "Why is it difficult to accurately predict the COVID-19 epidemic?" *Infect. Dis. Model.*, vol. 5, pp. 271–281, Mar. 2020. doi: 10.1016/j.idm.2020.03.001.
- [S5] S. L. Wu et al., "Substantial underestimation of SARS-CoV-2 infection in the united states," *Nature Commun.*, vol. 11, no. 1, pp. 1–10, 2020. doi: 10.1038/s41467-020-18272-4.
- [S6] R. Li et al., "Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2)," *Science*, vol. 368, no. 6490, pp. 489–493, 2020. doi: 10.1126/science.abb3221.
- [S7] M. G. M. Gomes et al., "Individual variation in susceptibility or exposure to SARS-CoV-2 lowers the herd immunity threshold," 2020. [Online]. Available: <https://www.medrxiv.org/content/10.1101/2020.04.27.20081893v3>
- [S8] P. E. Paré, C. L. Beck, and T. Basar, "Modeling, estimation, and analysis of epidemics over networks: An overview," *Annu. Rev. Control*, vol. 50, pp. 345–360, Sept. 2020. doi: 10.1016/j.arcontrol.2020.09.003.
- [S9] S. Y. Olmez et al., "A data-informed approach for analysis, validation, and identification of COVID-19 models," 2020. [Online]. Available: <https://www.medrxiv.org/content/early/2020/10/06/2020.10.03.20206250.1>
- [S10] R. Engbert, M. M. Rabe, R. Kliegl, and S. Reich, "Sequential data assimilation of the stochastic seir epidemic model for regional COVID-19 dynamics," 2020. [Online]. Available: <https://doi.org/10.1101/2020.04.13.20063768>
- [S11] G. Evensen et al., "An international assessment of the COVID-19 pandemic using ensemble data assimilation," 2020. [Online]. Available: <https://doi.org/10.1101/2020.06.11.20128777>

formula (14) was known in the EnKF literature prior to the FPF derivation [16], [48], and there have been a number of studies to improve upon this formula [9], [44], [49]–[53]. The following describes the diffusion map approximation, which appears to be the most promising approach in general settings.

Diffusion Map-Based Algorithm

The notation $e^{\epsilon \Delta_\rho}$ is used to denote the semigroup associated with the probability-weighted Laplacian Δ_ρ [54]. As explained in "Poisson Equation and Its Approximations" (and more fully in [55]), (11) is equivalently expressed as the fixed-point equation

$$\phi = e^{\epsilon \Delta_\rho} \phi + \int_0^\epsilon e^{s \Delta_\rho} (h - \hat{h}) ds, \quad (15)$$

where $\epsilon > 0$ is arbitrary. For small values of ϵ , there is a well-known approximation of the exact semigroup $e^{\Delta_\rho \epsilon}$ in terms of the so-called diffusion map

$$T_\epsilon f(x) := \frac{1}{n_\epsilon(x)} \int_{\mathbb{R}^d} \frac{g_\epsilon(x-y)}{\sqrt{\int g_\epsilon(y-z) \rho(z) dz}} f(y) \rho(y) dy,$$

where $g_\epsilon(x) := e^{-(|x|^2/4\epsilon)}$ is the Gaussian kernel in $\mathbb{R}$, and $n_\epsilon(x)$ is the normalization factor chosen so that $\int T_\epsilon 1(x) dx = 1$

Poisson Equation and Its Approximations

The Poisson equation (11) of nonlinear filtering is a linear partial differential equation. Its finite-dimensional counterpart is a familiar linear problem:

$$Ax = b, \quad (\text{S13})$$

where A is an $n \times n$ (strictly) positive definite symmetric matrix, and the right-hand side b is a given $n \times 1$ vector. The problem is to obtain the unknown $n \times 1$ vector x . For this purpose, the following equivalent formulations of the finite-dimensional problem are first introduced:

- 1) x is the solution of the weak form

$$y^T Ax = y^T b, \quad \forall y \in \mathbb{R}^n.$$

- 2) For any $t > 0$, x is the solution to the fixed-point equation

$$x = e^{-tA}x + \int_0^t e^{-sA} b ds.$$

- 3) x is the solution of the optimization problem

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} x^T Ax - x^T b.$$

When n is large, these formulations are useful to numerically approximate the solution of (S13):

- 1) For each fixed $y \in \mathbb{R}^n$, the weak form is a single equation. By restricting y to a suitable low-dimensional subspace $S \subset \mathbb{R}^n$, the number of linear equations is reduced for the purposes of obtaining an approximate solution (possibly also in S).
- 2) The fixed-point equation form is useful because e^{-tA} is a contraction for positive-definite A . So, a good initial guess for x can readily be improved by using the Banach iteration.
- 3) The optimization form is useful to develop alternate (for example, search-type) algorithms to obtain the solution to (S13).

Next, the Poisson equation (11) expressed succinctly as

$$-\Delta_\rho \phi = h - \hat{h},$$

where $\Delta_\rho = (1/\rho)\nabla \cdot (\rho\nabla)$ is the probability-weighted Laplacian. Functional analytic considerations require the introduction of function spaces: $L^2(\rho)$ is the space of square integrable functions with respect to ρ with inner product $\langle f, g \rangle_{L^2} = \int f(x)g(x)\rho(x)dx$, $H^1(\rho)$ is the Hilbert space of functions in $L^2(\rho)$ whose first derivative (defined in the weak sense) is also in $L^2(\rho)$, and $H_0^1(\rho) = \{\psi \in H^1(\rho) \mid \int \psi(x)\rho(x)dx = 0\}$. These definitions are important because $H_0^1(\rho)$ is the natural space for the solution ϕ (11). The operator $-\Delta_\rho$ is symmetric (self-adjoint) and positive definite because

$$-\langle f, \Delta_\rho g \rangle_{L^2} = \langle \nabla f, \nabla g \rangle_{L^2} = -\langle \Delta_\rho f, g \rangle_{L^2}, \quad \forall f, g \in H_0^1(\rho).$$

One requires an additional technical condition—the Poincaré inequality—to conclude that the operator is strictly positive definite. Assuming that the Poincaré inequality holds, it is also readily shown that Δ_ρ^{-1} is well defined, that is, a unique solution $\phi \in H_0^1(\rho)$ exists for a given $h \in L^2(\rho)$ [44, Th. 2]. For the purposes of numerical approximation, entirely analogous to the finite-dimensional case, the following equivalent formulations of the Poisson equation are introduced:

- 1) ϕ is a solution of the weak form

$$\langle \nabla \psi, \nabla \phi \rangle_{L^2} = \langle \psi, h - \hat{h} \rangle_{L^2} \quad \forall \psi \in H_0^1(\rho). \quad (\text{S14})$$

- 2) ϕ is a solution of the fixed-point equation

$$\phi = e^{t\Delta_\rho} \phi + \int_0^t e^{s\Delta_\rho} (h - \hat{h}) ds.$$

- 3) ϕ is the solution of the optimization problem

$$\min_{\phi \in H_0^1(\rho)} \frac{1}{2} \langle \nabla \phi, \nabla \phi \rangle_{L^2} + \langle \phi, h - \hat{h} \rangle_{L^2}. \quad (\text{S15})$$

These formulations have been used to develop numerical algorithms for gain function approximation:

- 1) Instead of $\psi \in H_0^1(\rho)$ in the weak form (S14), a relaxation is considered, whereby $\psi \in S = \text{span}\{\psi_1, \dots, \psi_M\}$, a finite-dimensional subspace of $H_0^1(\rho)$. The resulting algorithm is referred to as the *Galerkin algorithm for gain function approximation* [44]. The constant gain formula (14) is obtained by considering S to be the subspace spanned by the coordinate functions.
- 2) The semigroup $e^{t\Delta_\rho}$ is approximated with the diffusion map operator T_ϵ , as described previously in the article. This approximation yields the diffusion map-based algorithm for gain function approximation tabulated in “Diffusion Map-Based Algorithm for Gain Function Approximation.”
- 3) The optimization formulation (S15) is useful to explore nonlinear parameterizations of the gain function, for example, using neural networks. A preliminary investigation of this appears in [S12]. The related deep learning-inspired techniques for solving partial differential equations using neural networks appear in [S13].

REFERENCES

- [S12] S. Y. Olmez, A. Taghvaei, and P. G. Mehta, “Deep FPF: Gain function approximation in high-dimensional setting,” 2020, arXiv:2010.01183.
- [S13] E. Weinan and B. Yu, “The deep ritz method: A deep learning-based numerical algorithm for solving variational problems,” *Commun. Math. Statist.*, vol. 6, no. 1, pp. 1–12, 2018.

[56]. It is straightforward to approximate the diffusion map empirically in terms of the particles

$$T_\epsilon^{(N)}f(x) = \frac{1}{n_\epsilon^{(N)}(x)} \sum_{i=1}^N \frac{g_\epsilon(x - X^i)}{\sqrt{\sum_{j=1}^N g_\epsilon(X^i - X^j)}} f(X^i),$$

where $n_\epsilon^{(N)}(x)$ is the normalization factor. Upon approximating (15) using the empirical approximation $T_\epsilon^{(N)}$ for $e^{\epsilon\Delta_\rho}$, the diffusion map-based algorithm is obtained. The algorithm is summarized in “Diffusion Map-Based Algorithm for Gain Function Approximation.”

Error: Bias-Variance Tradeoff

The error in diffusion map approximation comes from two sources: 1) the bias error due to the diffusion map approximation of the semigroup and 2) the variance error due to empirical approximation in terms of particles. The error is analyzed in [15], where it is shown that

Root mean square error =

$$\left(\mathbb{E} \left[\frac{1}{N} \sum_{i=1}^N |K(X^i) - K_{\text{exact}}(X^i)|^2 \right] \right)^{\frac{1}{2}} \leq \underbrace{O(\epsilon)}_{\text{bias}} + \underbrace{O\left(\frac{1}{\epsilon^{1+\frac{d}{2}} N^{\frac{1}{2}}}\right)}_{\text{variance}}. \quad (16)$$

Diffusion Map-Based Algorithm for Gain Function Approximation

Input: $\{X^i : 1 \leq i \leq N\}$, $\{h(X^i) : 1 \leq i \leq N\}$, Kernel bandwidth ϵ

Output: $\{K^i : 1 \leq i \leq N\}$

- 1) $g_{ij} := e^{-\frac{|X^i - X^j|^2}{4\epsilon}}$ for $i, j = 1$ to N
- 2) $k_{ij} := \frac{g_{ij}}{\sqrt{\sum_l g_{il}} \sqrt{\sum_l g_{jl}}}$ for $i, j = 1$ to N
- 3) $d_i = \sum_j k_{ij}$ for $i = 1$ to N
- 4) $\tau_{ij} := \frac{k_{ij}}{d_i}$ for $i, j = 1$ to N
- 5) $\pi_i = \frac{d_i}{\sum_j d_j}$ for $i = 1$ to N
- 6) $\hat{h} = \sum_{i=1}^N \pi_i h(X^i)$
- 7) $\Phi = (0, \dots, 0) \in \mathbb{R}^N$
- 8) Solve the fixed-point problem $\Phi = \tau\Phi + \epsilon(h - \hat{h})$ iteratively
- 9) $r_i = \Phi_i + \epsilon h_i$ for $i = 1$ to N
- 10) $s_{ij} = \frac{1}{2\epsilon} \tau_{ij} \left(r_j - \sum_{k=1}^N \tau_{ik} r_k \right)$ for $i, j = 1$ to N
- 11) $K^i = \sum_j s_{ij} X^j$ for $i = 1$ to N

The error due to bias converges to zero as $\epsilon \rightarrow 0$, and the error due to variance converges to zero as $N \rightarrow \infty$. There is a tradeoff between the two errors. To reduce bias, one must reduce ϵ ; however, for any fixed value of N , one can reduce ϵ only up to a point where the variance starts increasing. The bias-variance tradeoff is illustrated in Figure 5. If ϵ is large, the error due to bias dominates; however, if ϵ is small, the error due to variance dominates. As a final point, there is a remarkable and somewhat unexpected relationship between the diffusion map and the constant gain approximations. Specifically, in the limit as $\epsilon \rightarrow \infty$, the diffusion map gain converges to the constant gain. This suggests a systematic procedure to improve upon the constant gain by detuning the value of ϵ away from the $[\epsilon = \infty]$ limit. For any fixed N , a finite value of ϵ is chosen to

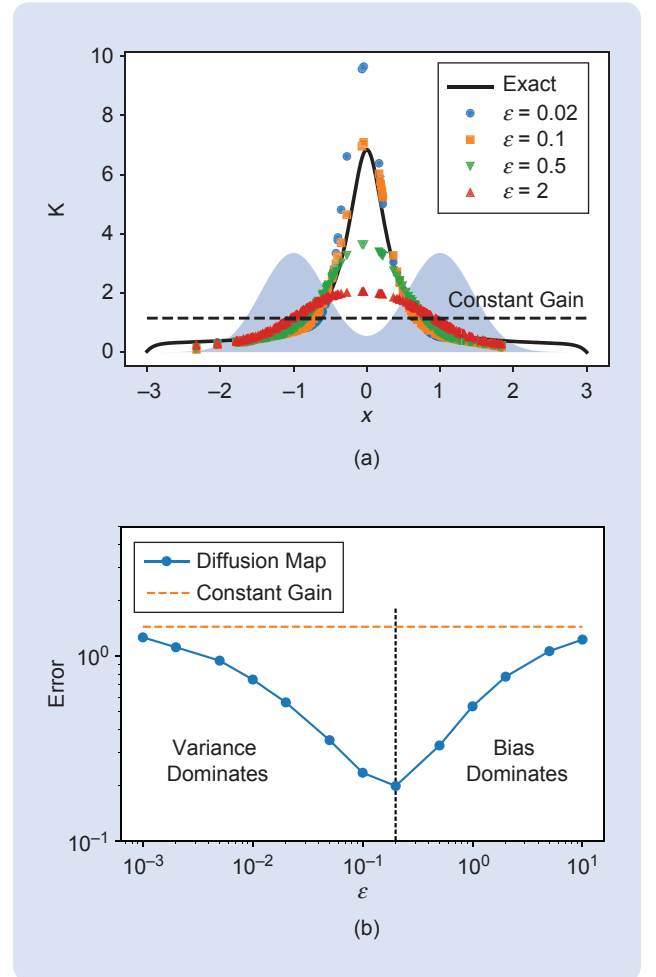


FIGURE 5 The bias-variance tradeoff in diffusion map-based gain function approximation. (a) The dashed line is the constant gain solution in (14). As $\epsilon \rightarrow \infty$, the diffusion map gain converges to the constant gain. The shaded area in the background is the density function ρ taken as the sum of two Gaussians, $\mathcal{N}(-1, \sigma^2)$ and $\mathcal{N}(+1, \sigma^2)$, with $\sigma^2 = 0.2$. The exact gain function $K(x)$ is computed for $h(x) = x$ by using an integral formula [55, eq. 4.6]. (b) The root-mean-square error is computed as an empirical approximation of (16) by averaging 1000 simulations for $N = 200$ particles.

The error due to bias converges to zero as $\epsilon \rightarrow 0$, and the error due to variance converges to zero as $N \rightarrow \infty$.

minimize the root mean square error according to the bias-variance tradeoff. Based on this, a rule of thumb for choosing the ϵ value appears in [55, Remark 5.1].

SOME FINAL REMARKS

In the past decade, the coupling perspective to data-assimilation problems has been enormously valuable, with outstanding theoretical contributions and application impacts. Given the limited scope of this article (with its narrow focus on the FPF algorithm), it is not possible to do justice to the depth and breadth of this exciting new area in one article. The reader is referred to [17] and [18] for an excellent introduction to the subject.

A few important remarks are also necessary. The continuous-time formulation is stressed in this article for the reasons of mathematical elegance and beauty. In practice, discrete-time formulations are much more common. The coupling viewpoint also applies to these settings [17] and was used in the article to introduce the main ideas. Next, optimal couplings are almost always difficult to compute. The most popular forms of couplings used, in practice, are suboptimal. This is true for the classical EnKF and the FPF algorithms. A discussion and exactness and optimality for FPF appears in "Optimal Transport Construction of Stochastic Processes."

As a final point, closely related to the coupling viewpoint is the gradient flow interpretation of the Bayes' update formula (see [46] for an FPF-specific exposition and also [57] and [58] for related algorithms).

There are several directions for future work. It is an open problem to fully carry out stability and error analysis of the finite- N FPF particle system with the diffusion map-based gain function approximation. It will be very useful to characterize the CoD in these general settings. It is also of interest to construct optimization-type formulations that directly yield a finite- N algorithm without the need for empirical approximation as an intermediate step. Such constructions may lead to better error properties by design. Finally, apart from the optimal transportation formulation stressed in this article, one may consider alternative approaches for control design. One possible direction is based on the Schrödinger bridge problem [18], [59].

AUTHOR INFORMATION

Amirhossein Taghvaei (ataghvae@uci.edu) received the Ph.D. degree in mechanical science and engineering and the M.S. degree in mathematics from the University of

Illinois at Urbana-Champaign. He is a postdoctoral researcher at the University of California, Irvine, California, 92697, USA. In the fall of 2021, he will be an assistant professor in the Department of Aeronautics and Astronautics, University of Washington, Seattle. His research interests include mean-field optimal control, optimal transportation theory, and machine learning.

Prashant G. Mehta received the Ph.D. degree in applied mathematics from Cornell University, Ithaca, New York, in 2004. He is a professor in the Department of Mechanical Science and Engineering at the University of Illinois at Urbana-Champaign (UIUC), Champaign, Illinois, 61801, USA. Prior to joining UIUC, he was a research engineer at the United Technologies Research Center (UTRC). He cofounded the university-originated start-up Rithmio, whose technology portfolio was acquired by Bosch Sensortec in 2017. He received the Outstanding Achievement Award at UTRC in 2004 for his contributions to the modeling and control of combustion instabilities in jet engines. His students received best student article awards at the IEEE Conference on Decision and Control in 2007, 2009, and 2019 and were finalists for these awards in 2010 and 2012. He serves on the editorial board of *IEEE Transactions on Automatic Control*. He previously served on the editorial boards of *ASME Journal of Dynamic Systems, Measurement, and Control* and *Systems and Control Letters*. His research interests include nonlinear filtering and mean-field games. He is a Member of IEEE.

REFERENCES

- [1] S. B. Bastos and D. O. Cajueiro, "Modeling and forecasting the early evolution of the COVID-19 pandemic in Brazil," 2020, arXiv:2003.14288.
- [2] R. E. Kalman and R. S. Bucy, "New results in linear filtering and prediction theory," *J. Basic Eng.*, vol. 83, no. 1, pp. 95–108, 1961. doi: 10.1115/1.3658902.
- [3] H. A. P. Blom, "The continuous time roots of the interacting multiple model filter," in *Proc. 51st IEEE Conf. Decis. Control*, Maui, Hawaii, Dec. 2012, pp. 6015–6021.
- [4] Y. Bar-Shalom, F. Daum, and J. Huang, "The probabilistic data association filter," *IEEE Control Syst. Mag.*, vol. 29, no. 6, pp. 82–100, Dec. 2009.
- [5] A. Bensoussan, J. Frehse, P. Yam, *Mean field Games and Mean Field Type Control Theory*, vol. 101. New York: Springer-Verlag, 2013.
- [6] R. Carmona and F. Delarue, *Probabilistic Theory of Mean Field Games with Applications I-II*. New York: Springer-Verlag, 2018.
- [7] T. Yang, P. G. Mehta, and S. P. Meyn, "A mean-field control-oriented approach for particle filtering," in *Proc. Amer. Control Conf.*, San Francisco, June 2011, pp. 2037–2043. doi: 10.1109/ACC.2011.5991422.
- [8] T. Yang, P. G. Mehta, and S. P. Meyn, "Feedback particle filter with mean-field coupling," in *Proc. 50th IEEE Conf. Decis. Control*, Orlando, FL, Dec. 2011, pp. 7909–7916. doi: 10.1109/CDC.2011.6160950.
- [9] T. Yang, P. G. Mehta, and S. P. Meyn, "Feedback particle filter," *IEEE Trans. Autom. Control*, vol. 58, no. 10, pp. 2465–2480, 2013. doi: 10.1109/TAC.2013.2258825.
- [10] A. Taghvaei, J. De Wiljes, P. G. Mehta, and S. Reich, "Kalman filter and its modern extensions for the continuous-time nonlinear filtering

- problem," *Trans. ASME, J. Dyn. Syst. Meas. Control*, vol. 140, no. 3, p. 030904, 2018. doi: 10.1115/1.4037780.
- [11] T. Yang, G. Huang, and P. G. Mehta, "Joint probabilistic data association-feedback particle filter for multiple target tracking applications," in *Proc. Amer. Control Conf. (ACC)*, 2012, pp. 820–826. doi: 10.1109/ACC.2012.6315551.
- [12] T. Yang, H. A. Blom, and P. G. Mehta, "Interacting multiple model-feedback particle filter for stochastic hybrid systems," in *Proc. 52nd IEEE Conf. Decis. Control*, 2013, pp. 7065–7070. doi: 10.1109/CDC.2013.6761009.
- [13] G. Evensen, "Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics," *J. Geophys. Res., Oceans*, vol. 99, no. C5, pp. 10,143–10,162, 1994. doi: 10.1029/94JC00572.
- [14] D. Crisan and J. Xiong, "Approximate McKean-Vlasov representations for a class of SPDEs," *Stochastics*, vol. 82, no. 1, pp. 53–68, 2010. doi: 10.1080/17442500902723575.
- [15] S. Reich, "A dynamical systems framework for intermittent data assimilation," *BIT Numerical Anal.*, vol. 51, no. 1, pp. 235–249, 2011. doi: 10.1007/s10543-010-0302-4.
- [16] K. Bergemann and S. Reich, "An ensemble Kalman-Bucy filter for continuous data assimilation," *Meteorologische Zeitschrift*, vol. 21, no. 3, pp. 213–219, 2012. doi: 10.1127/0941-2948/2012/0307.
- [17] S. Reich and C. Cotter, *Probabilistic forecasting and Bayesian data assimilation*. Cambridge, U.K.: Cambridge Univ. Press, 2015.
- [18] S. Reich, "Data assimilation: The Schrödinger perspective," *Acta Numerica*, vol. 28, pp. 635–711, June 2019. doi: 10.1017/S0962492919000011.
- [19] Y. Cheng and S. Reich, "A McKean optimal transportation perspective on Feynman-Kac formulae with application to data assimilation," 2013, arXiv:1311.6300.
- [20] N. J. Gordon, D. J. Salmond, and A. F. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," in *Proc. Inst. Elect. Eng., F (Radar and Signal Processing)*, vol. 140, no. 2, pp. 107–113, 1993. doi: 10.1049/ip-f-2.1993.0015.
- [21] A. M. Doucet and A. Johansen, "A tutorial on particle filtering and smoothing: Fifteen years later," in *Handbook Nonlinear Filtering*, 2009, vol. 12, pp. 656–704. [Online]. Available: https://www.cs.ubc.ca/~arnaud/doucet_johansen_tutorialPF.pdf
- [22] P. Del Moral and A. Guionnet, "On the stability of interacting processes with applications to filtering and genetic algorithms," *Annales de l'Institut Henri Poincaré (B) Probab. Statist.*, vol. 37, no. 2, pp. 155–194, 2001. doi: 10.1016/S0246-0203(00)01064-5.
- [23] O. Cappé, E. Moulines, and T. Rydén, "Inference in hidden Markov models," in *Proc. EUFLAT Conf.*, 2009, pp. 14–16.
- [24] T. Bengtsson, P. Bickel, and B. Li, "Curse of dimensionality revisited: Collapse of the particle filter in very large scale systems," in *IMS Lecture Notes - Monograph Series in Probability and Statistics: Essays in Honor of David F. Freedman*, D. Nolan and T. Speed, Eds. Beachwood, OH: Institute of Mathematical Sciences, 2008, vol. 2, pp. 316–334.
- [25] A. Beskos, D. Crisan, A. Jasra, and N. Whiteley, "Error bounds and normalising constants for sequential Monte Carlo samplers in high dimensions," *Advances Appl. Probab.*, vol. 46, no. 1, pp. 279–306, 2014. doi: 10.1239/aap/1396360114.
- [26] 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, 2015. doi: 10.1214/14-AAP1061.
- [27] P. Del Moral, "Feynman-Kac formulae," in *Feynman-Kac Formulae*, J. Gani, C. C. Heyde, and T. G. Kurtz, Eds. New York: Springer-Verlag, 2004, pp. 47–93.
- [28] A. Bain and D. Crisan, *Fundamentals of Stochastic Filtering*, vol. 3. New York: Springer-Verlag, 2009.
- [29] C. Villani, *Topics in Optimal Transportation*. Providence, RI: American Mathematical Society, 2003.
- [30] L. C. Evans, "Partial differential equations and Monge-Kantorovich mass transfer," *Curr. Developments Math.*, vol. 1997, no. 1, pp. 65–126, 1997. doi: 10.4310/CDM.1997v1997.n1.a2.
- [31] A. Taghvaei and P. G. Mehta, "An optimal transport formulation of the linear feedback particle filter," in *Proc. Amer. Control Conf. (ACC)*, 2016, pp. 3614–3619.
- [32] J. Whitaker and T. M. Hamill, "Ensemble data assimilation without perturbed observations," *Monthly Weather Rev.*, vol. 130, no. 7, pp. 1913–1924, 2002. doi: 10.1175/1520-0493(2002)130<1913:EDAWPO>2.0.CO;2.
- [33] J. Mandel, L. Cobb, and J. D. Beezley, "On the convergence of the ensemble Kalman filter," *Appl. Math.*, vol. 56, no. 6, pp. 533–541, 2011. doi: 10.1007/s10492-011-0031-2.
- [34] P. Del Moral and J. Tugaut, "On the stability and the uniform propagation of chaos properties of ensemble Kalman-Bucy filters," *Ann. Appl. Probab.*, vol. 28, no. 2, pp. 790–850, Apr. 2018. doi: 10.1214/17-AAP1317.
- [35] P. Del Moral, A. Kurtzmann, and J. Tugaut, "On the stability and the uniform propagation of chaos of a class of extended ensemble Kalman-Bucy filters," *SIAM J. Control Optim.*, vol. 55, no. 1, pp. 119–155, 2017. doi: 10.1137/16M1087497.
- [36] J. de Wiljes, S. Reich, and W. Stannat, "Long-time stability and accuracy of the ensemble Kalman-Bucy filter for fully observed processes and small measurement noise," *SIAM J. Appl. Dyn. Syst.*, vol. 17, no. 2, pp. 1152–1181, 2018. doi: 10.1137/17M1119056.
- [37] A. Taghvaei and P. G. Mehta, "An optimal transport formulation of the ensemble Kalman filter," *IEEE Trans. Autom. Control*, early access, 2019. doi: 10.1109/TAC.2020.3015410.
- [38] A. Sznitman, "Topics in propagation of chaos," in *Ecole d'Été de Probabilités de Saint-Flour XIX-1989*, P. L. Hennequin, Eds. New York: Springer-Verlag, 1991, pp. 165–251.
- [39] S. T. Rachev and L. Rüschendorf, *Mass Transportation Problems: Volume I: Theory*, vol. 1. New York: Springer Science & Business Media, 1998.
- [40] T. A. El Moselhy and Y. M. Marzouk, "Bayesian inference with optimal maps," *Comput. Phys.*, vol. 231, no. 23, pp. 7815–7850, 2012. doi: 10.1016/j.jcp.2012.07.022.
- [41] D. A. Mesa, J. Tantiengloc, M. Mendoza, S. Kim, and T. P. Coleman, "A distributed framework for the construction of transport maps," *Neural Comput.*, vol. 31, no. 4, pp. 613–652, 2019. doi: 10.1162/neco_a_01172.
- [42] J. Heng, A. Doucet, and Y. Pokern, "Gibbs flow for approximate transport with applications to Bayesian computation," 2015, arXiv:1509.08787.
- [43] B. Oksendal, *Stochastic Differential Equations: An Introduction with Applications*. Heidelberg, Germany: Springer Science & Business Media, 2013.
- [44] T. Yang, R. S. Laugesen, P. G. Mehta, and S. P. Meyn, "Multivariable feedback particle filter," *Automatica*, vol. 71, pp. 10–23, Sept. 2016. doi: 10.1016/j.automatica.2016.04.019.
- [45] S. Pathiraja, S. Reich, and W. Stannat, "McKean-Vlasov SDES in nonlinear filtering," 2020, arXiv:2007.12658.
- [46] R. S. Laugesen, P. G. Mehta, S. P. Meyn, and M. Raginsky, "Poisson's equation in nonlinear filtering," *SIAM J. Control Optim.*, vol. 53, no. 1, pp. 501–525, 2015. doi: 10.1137/13094743X.
- [47] C. Zhang, A. Taghvaei, and P. G. Mehta, "Feedback particle filter on Riemannian manifolds and matrix Lie groups," *IEEE Trans. Autom. Control*, vol. 63, no. 8, pp. 2465–2480, 2017. doi: 10.1109/TAC.2017.2771336.
- [48] G. Evensen, *Data Assimilation. The Ensemble Kalman Filter*. New York: Springer-Verlag, 2006.
- [49] K. Berntorp and P. Grover, "Data-driven gain computation in the feedback particle filter," in *Proc. 2016 Amer. Control Conf. (ACC)*, pp. 2711–2716.
- [50] Y. Matsuura, R. Ohata, K. Nakakuki, and R. Hirokawa, "Suboptimal gain functions of feedback particle filter derived from continuation method," in *Proc. AIAA Guidance, Navigation, Control Conf.*, 2016, p. 1620.
- [51] A. Radhakrishnan, A. Devraj, and S. Meyn, "Learning techniques for feedback particle filter design," in *Proc. Conf. Decis. Control (CDC)*, 2016, pp. 648–653.
- [52] A. Radhakrishnan and S. Meyn, "Feedback particle filter design using a differential-loss reproducing kernel Hilbert space," in *Proc. 2018 Annu. Amer. Control Conf. (ACC)*, pp. 329–336.
- [53] K. Berntorp, "Comparison of gain function approximation methods in the feedback particle filter," in *Proc. 21st Int. Conf. Inf. Fusion (FUSION)*, 2018, pp. 123–130.
- [54] D. Bakry, I. Gentil, and M. Ledoux, *Analysis and Geometry of Markov Diffusion Operators*, vol. 348. Cham, Switzerland: Springer Science & Business Media, 2013.
- [55] A. Taghvaei, P. G. Mehta, and S. P. Meyn, "Diffusion map-based algorithm for gain function approximation in the feedback particle filter," *SIAM/ASA J. Uncertainty Quantification*, vol. 8, no. 3, pp. 1090–1117, 2020. doi: 10.1137/19M124513X.
- [56] R. R. Coifman and S. Lafon, "Diffusion maps," *Appl. Comput. Harmon. Anal.*, vol. 21, no. 1, pp. 5–30, 2006. doi: 10.1016/j.acha.2006.04.006.
- [57] A. Halder and T. T. Georgiou, "Gradient flows in filtering and Fisher-Rao geometry," in *Proc. 2018 Annu. Amer. Control Conf. (ACC)*, 2018, pp. 4281–4286.
- [58] A. Garbuno-Inigo, F. Hoffmann, W. Li, and A. M. Stuart, "Interacting Langevin diffusions: Gradient structure and ensemble Kalman sampler," *SIAM J. Appl. Dyn. Syst.*, vol. 19, no. 1, pp. 412–441, 2020. doi: 10.1137/19M1251655.
- [59] Y. Chen, T. T. Georgiou, and M. Pavon, "On the relation between optimal transport and Schrödinger bridges: A stochastic control viewpoint," *J. Optim. Theory Appl.*, vol. 169, no. 2, pp. 671–691, 2016. doi: 10.1007/s10957-015-0803-z.