An Optimal Transport Formulation of the Linear Feedback Particle Filter

Amirhossein Taghvaei, Prashant G. Mehta

Abstract—Feedback particle filter (FPF) is an algorithm to numerically approximate the solution of the nonlinear filtering problem in continuous time. The algorithm implements a feedback control law for a system of particles such that the empirical distribution of particles approximates the posterior distribution. However, it has been noted in the literature that the feedback control law is not unique. To find a unique control law, the filtering task is formulated here as an optimal transportation problem between the prior and the posterior distributions. Based on this formulation, a time stepping optimization procedure is proposed for the optimal control design. A key difference between the optimal control law and the one in the original FPF, is the replacement of noise term with a deterministic term. This difference serves to decreases the simulation variance, as illustrated with a simple numerical example.

I. INTRODUCTION

Nonlinear filtering is concerned with the problem of computing the posterior distribution of a hidden Markov process \(X_t\), given a time history of observations \(Z_t\). In the linear Gaussian setting, the solution is given by the Kalman-Bucy filter. For the general nonlinear non-Gaussian case, the problem is infinite dimensional and only approximate numerical solutions are possible.

The feedback particle filter (FPF) algorithm provides one such approximate numerical solution [20], [19]. In the FPF algorithm, the posterior distribution is approximated by empirical distribution of an ensemble of particles. The particles are realizations of a stochastic process, denoted in this paper as \(S_t\). The \(S_t\) process is simulated according to,

\[
dS_t = u_t(S_t) dt + K_t(S_t) dZ_t ,
\]

where \(u_t\) and \(K_t\) are control laws, designed such that \(S_t\) is distributed according to posterior distribution of \(X_t\), i.e.,

\[
S_t \sim P(X_t | Z_t), \quad \forall t \geq 0,
\]

where \(Z_t := \sigma\{Z_s : 0 \leq s \leq t\}\) is the filtration generated by observation. When this condition is true, the filter is said to be exact.

In the original FPF algorithm, a particular choice of \(u_t\) and \(K_t\) is proposed such that the filter is exact. Nevertheless, in its general form (1), there are infinitely many choices of \(u_t\) and \(K_t\) that all lead to exactness. Investigating the optimality properties of the choice made in FPF and selecting a unique optimal control law is the focus of the current work.

The optimality concept considered here is motivated by the optimal transportation literature [18], [9]. One can interpret feedback particle filter as transporting the initial distribution at time \(t = 0\) (prior) to the conditional distribution at time \(t\) (posterior). Clearly, there are infinitely many maps that transport one distribution into another. Selecting a unique and optimal map is the basic problem of optimal transportation and also informs the investigation of this paper (see Figure 1).

Based on the concept of optimal transportation, a time-stepping optimization procedure is proposed here to obtain a unique optimal control law, denoted as \(u^*_t\) and \(K^*_t\). In this procedure, a finite time interval is divided into discrete time steps \(\{t_0, t_1, \ldots, t_n\}\). Then a discrete time random process, \(S_{t_k}\) is constructed by initializing \(S_0\) according to the initial prior \(P(X_0)\), and sequentially evolving \(S_{t_k} \to S_{t_{k+1}}\) at each time-step with a map denoted by \(T_k\):

\[
S_{t_{k+1}} = T_k(S_{t_k}), \quad S_0 \sim P(X_0).
\]

The map \(T_k\) is obtained by solving an optimal transportation problem between the conditional probability distributions at time instants \(t_k\) and \(t_{k+1}\), i.e between \(P(X_{t_k} | Z_{t_k})\) and \(P(X_{t_{k+1}} | Z_{t_{k+1}})\). By construction, \(S_{t_k}\) is distributed according to \(P(X_{t_k} | Z_{t_k})\) for all \(k \in \{1, \ldots, n\}\). Finally, by taking the continuous-time limit, a continuous time process \(S_t\) and therefore the unique control law, \(u^*\) and \(K^*\) for (1) are obtained. Figure 2 illustrates the procedure.

The contributions of this paper are as follows: An optimal transport formulation of the feedback particle filter is introduced. The new formulation helps to solve the uniqueness issue for the FPF. For the linear Gaussian filtering problem, the optimal transport FPF is obtained and compared to the original FPF. Using a simple example, the advantage of optimal formulation in reducing simulation variance is analytically and numerically demonstrated.

There are two streams of literature that are relevant to
this work: (i) Particle filter algorithms that seek to replace the importance sampling step with a deterministic approach [6], [12], [5], [20]; (ii) Optimal transportation techniques for uncertainty propagation that includes synthesis of optimal transport maps for implementing the Bayes rules as a special case [4], [15], [8]. The present work integrates these two streams, that furthermore highlights the optimal transportation roots of the FPF algorithm.

The evaluation of the FPF algorithm and its comparison to other nonlinear filtering algorithms appears in [3], [16] [17]. For additional applications of optimal transportation see [7], [11], [14], [2]. The outline of the remainder of this paper is as follows: In Section II the linear FPF is introduced and the uniqueness issue is discussed. Section III includes a short background on the optimal transportation problem. In Section IV, the time stepping optimization procedure is proposed and applied to the linear Gaussian filtering problem. Finally Section V includes a comparison between the original FPF and the optimal transport FPF.

Notation: \( P \) denotes a probability distribution. \( X \sim P \) means the random variable \( X \) is distributed according to \( P \). \( \mathcal{N}(\mu, \Sigma) \) is a Gaussian probability distribution with mean \( \mu \) and covariance \( \Sigma \). \( \mathbf{A} \succ 0 \) means \( \mathbf{A} \) is a positive definite matrix. All sdes are to be interpreted in Itô sense.

II. EXACTNESS AND NON-UNIQUENESS

Linear Gaussian filtering: Consider the linear Gaussian filtering problem:

\[
\begin{align*}
\mathrm{d}X_t &= AX_t \mathrm{d}t + \mathrm{d}B_t, \\
\mathrm{d}Z_t &= CX_t \mathrm{d}t + \mathrm{d}W_t,
\end{align*}
\]

where \( X_t \in \mathbb{R}^d \) is the state at time \( t \), \( Z_t \in \mathbb{R}^m \) is the observation process, \( A, C \) are matrices of appropriate dimension, and \( \{B_t\}, \{W_t\} \) are two mutually independent Wiener processes taking values in \( \mathbb{R}^d \) and \( \mathbb{R}^m \), respectively. Without loss of generality, it is assumed that the covariance matrices associated with \( \{B_t\}, \{W_t\} \) are identity matrices. The initial condition is assumed to have a Gaussian distribution, \( \mathcal{N}(\hat{X}_0, \Sigma_0) \), with \( \Sigma_0 \succ 0 \). The initial condition and noise processes are also independent. The filtering problem is to compute the posterior distribution \( P(X_t|Z_t) \), where \( Z_t = \sigma(Z_s), s \leq t \).

Kalman filter: In the linear Gaussian case, the posterior distribution \( P(X_t|Z_t) \) is Gaussian \( \mathcal{N}(\hat{X}_t, \Sigma_t) \), whose mean and variance are given by the Kalman-Bucy filter,

\[
\begin{align*}
\dot{\hat{X}}_t &= A\hat{X}_t + K_t(dZ_t - C\hat{X}_t \mathrm{d}t), \\
\frac{\mathrm{d}}{\mathrm{d}t}\Sigma_t &= A\Sigma_t + \Sigma_t AT + I - \Sigma_t CT \Sigma_t,
\end{align*}
\]

where \( K_t := \Sigma_t C^T \) is the Kalman gain.

Feedback particle filter: The feedback particle filter for linear Gaussian problem (2a)-(2b) is given by,

\[
\begin{align*}
\mathrm{d}S_t &= AS_t \mathrm{d}t + \mathrm{d}B_t + \tilde{K}_t(dZ_t - \frac{CS_t + CS_t T}{2} \mathrm{d}t), \\
\end{align*}
\]

for \( i = 1, \ldots, N \), where \( \{B'_i\} \) are independent Wiener processes, \( S_0 \sim \mathcal{N}(\hat{X}_0, \Sigma_0) \), and \( \tilde{K}_i := \Sigma_i C^T \), and

\[
\tilde{S}_i = \frac{1}{N} \sum_{i=1}^{N} S_i, \quad \tilde{K}_i = \frac{1}{N} \sum_{i=1}^{N} (S_i - \hat{S}_i)(S_i - \hat{S}_i)^T.
\]

The sde (4) represents the mean-field limit of the finite-\( N \) algorithm, which is the focus of present work.

Theorem 1: (Exactness of linear FPF) Consider the linear Gaussian filtering problem (2a)-(2b), and the linear FPF (4). If \( P(X_0) = P(S_0) \), then

\[
P(S_t|Z_t) = P(X_t|Z_t), \quad \forall t > 0.
\]

Proof: The conditional distribution of \( S_t \) is a Gaussian because the sde (4) is linear and the initial condition is Gaussian. So, to prove (5), it is sufficient to show that the conditional mean and variance of \( S_t \) evolve according to Kalman filtering equations. The equation for the conditional mean is obtained by taking conditional mean for both sides of (4), using linearity and the fact that \( B'_i \) is zero mean.

\[
\mathrm{d}\tilde{S}_t = A\tilde{S}_t \mathrm{d}t + \tilde{K}_t(dZ_t - C\tilde{S}_t \mathrm{d}t).
\]

This equation is similar to (3a) except that the gain \( K_t = \Sigma_t C^T \) is replaced by \( \tilde{K}_t = \Sigma_t C^T \). In order to show the variances are equal, define \( E_t \) according to,

\[
E_t := S_t - \hat{S}_t.
\]

The equation for \( E_t \) is obtained by simply subtracting (6) from (4). This gives,

\[
\mathrm{d}E_t = (A - \frac{\Sigma_t C^T C}{2})E_t + \mathrm{d}B_t.
\]

The equation for the variance of \( E_t \) is now given by the Lyapunov equation,

\[
\frac{\mathrm{d}}{\mathrm{d}t}\Sigma_t = A\Sigma_t + \Sigma_t AT + I - \Sigma_t CT \Sigma_t.
\]
which is identical to (3b), starting with the same initial condition. Hence \( \tilde{X}_t = \Sigma_t \) for all \( t \geq 0 \). This also implies \( \tilde{K}_t = K_t \), which further implies that the sde for conditional mean (6) is identical to (3a). So \( \tilde{X}_t = \tilde{S}_t \).

**Remark 2: (Non-uniqueness)** The exactness proof above suggests a general procedure to construct an exact \( S \)-process. In particular, express \( S_t \) as sum of two terms,

\[
S_t = \tilde{S}_t + E_t.
\]

Let \( \tilde{S}_t \) evolve according to (3a) and let \( E_t \) evolve according to,

\[
dE_t = G_t E_t \, dt,
\]

where \( G_t \) is any solution to the matrix equation,

\[
G_t \tilde{\Sigma}_t + \tilde{\Sigma}_t G_t^T = A \tilde{\Sigma}_t + \tilde{\Sigma}_t A^T + I - \tilde{\Sigma}_t C^T C \tilde{\Sigma}_t.
\]

By construction, the equation for the variance is given by

\[
\frac{d\Sigma_t}{dt} = \Omega_t \Sigma_t + \Sigma_t \Omega_t^T + \frac{1}{2} \Sigma_t C^T C \Sigma_t.
\]

Remark 3: Consider affine maps of the form,

\[
T(x) = \tilde{Y} + F(x - \tilde{X}).
\]

The constraint \( T(X) \sim Y \) is satisfied for all matrices \( F \) such that

\[
F \Sigma_X F^T = \Sigma_Y.
\]  

The general solution of (12) is

\[
F = \frac{1}{\Sigma_Y} U \Sigma_X^{-\frac{1}{2}},
\]

for all orthogonal matrices \( U \). The optimal choice of \( U \) is the one that serves to make \( F \) a symmetric matrix.

**IV. OPTIMAL TRANSPORT SDE**

Consider the linear Gaussian filtering problem (2a)-(2b), over a finite time interval \( [0,T] \). The objective of this section is to design a unique optimal sde whose solution, denoted as \( S_t \) has conditional probability distribution equal to the conditional probability distribution of \( X_t \). That is,

\[
P(S_t | \mathcal{Z}_t) = P(X_t | \mathcal{Z}_t), \quad \text{for} \quad t \in [0,T].
\]

Recall that \( P(X_t | \mathcal{Z}_t) \) is Gaussian with mean and variance that evolve according to the Kalman filtering equations (3a)-(3b).

**A. Time stepping optimization procedure**

The following time stepping optimization procedure is proposed to obtain the desired sde:

1) Divide the time interval \([0,T]\) into \( n \in \mathbb{N} \) equal time steps with the time instants \( t_0 = 0 < t_1 < \ldots < t_n = T \).

2) Initialize a discrete time random process \( \{ S_{t_k} \}_{k=1}^n \) according to the initial distribution (prior) of \( X_0 \),

\[
S_{t_0} \sim P(X_0).
\]

3) For each time step \([t_k, t_{k+1}]\), evolve the process \( S_{t_k} \) according to the map \( T_{t_k} \),

\[
S_{t_{k+1}} = T_{t_k}(S_{t_k}), \quad \text{for} \quad k = 0, \ldots, n-1,
\]

where the map \( T_{t_k} \) is the optimal transport map between two probability measures, \( P(X_{t_k} | \mathcal{Z}_{t_k}) \) and \( P(X_{t_{k+1}} | \mathcal{Z}_{t_{k+1}}) \).

4) Take the limit as \( n \to \infty \) to obtain the continuous time process \( S_t \) and the desired sde,

\[
dS_t = u^*_t(S_t) \, dt + K_t(s) \, dZ_t.
\]

The procedure leads to the control laws \( u^*_t \) and \( K^*_t \) that depends upon \( P(X_t | \mathcal{Z}_t) \). Since \( P(X_t | \mathcal{Z}_t) \) is unknown, one can replace it with \( P(S_t | \mathcal{Z}_t) \), as the two are identical by construction. In practice, one would also need to approximate \( P(S_t | \mathcal{Z}_t) \) by simulating \( N \) realizations of the process \( S_t \) (see Remark 1).

The resulted sde (4) is referred to as the optimal transport sde. Next, the optimal transport sde is obtained for two cases: The scalar case in section IV-B, and the vector case in section IV-C.
B. Scalar Case

This section considers the linear Gaussian filtering problem (2a)-(2b) in the scalar case, where \( X_t, Z_t \in \mathbb{R} \). In this case, the matrices \( A \) and \( C \) are scalars, and denoted by \( a \) and \( c \) respectively. The proof of the following proposition appears in Appendix B.

**Proposition 2**: Consider the linear Gaussian filtering problem (2a)-(2b) in the scalar case. Assume the pair \((a,c)\) is observable, i.e \( c \neq 0 \). The optimal transport sde for this problem is given by,

\[
dS_t = aS_t \, dt + \frac{1}{2\sigma_t^2}(S_t - \hat{S}_t) \, dt + \tilde{K}_t (dZ_t - \frac{cS_t + c\hat{S}_t}{2} \, dt),
\]

where \( \tilde{K}_t = \hat{\Sigma}_t c, \hat{S}_t = E[S_t | \mathcal{Z}_t], \hat{\Sigma}_t = E[(S_t - \hat{S}_t)^2 | \mathcal{Z}_t] \). The sde (14) is exact, i.e,

\[
P(S_t | \mathcal{Z}_t) = P(X_t | \mathcal{Z}_t) \quad \text{for} \quad t > 0,
\]

if \( P(S_0) = P(X_0) \).

**Remark 4**: The Bayesian update law in the optimal transport sde (14) \( \tilde{K}_t (dZ_t - \frac{cS_t + c\hat{S}_t}{2} \, dt) \) is identical to FPF (4). The difference between (14) and (4) is the replacement of the stochastic term \( dB_t \) with a deterministic term \( \frac{1}{2\sigma_t^2}(S_t - \hat{S}_t) \, dt \). In section V, a numerical example is described to show that this difference serves to decrease the simulation variance.

C. Vector Case

The proof of the following Proposition appears in Appendix B.

**Proposition 3**: Consider the linear Gaussian filtering problem (2a)-(2b). Assume the pair \((A,C)\) is observable. Then the optimal transport sde for this problem is given by,

\[
dS_t = AS_t \, dt + \hat{K}_t (dZ_t - CS_t \, dt) + G_t (S_t - \hat{S}_t) \, dt,
\]

where \( \hat{K}_t := \hat{\Sigma}_t C^T, \hat{S}_t = E[S_t | \mathcal{Z}_t], \hat{\Sigma}_t = E[(S_t - \hat{S}_t)^2 | \mathcal{Z}_t] \), and \( G_t \) is the solution of the matrix equation,

\[
G_t \hat{\Sigma}_t + \hat{\Sigma}_t G_t^T = A \hat{\Sigma}_t + \hat{\Sigma}_t A^T + I - \hat{\Sigma}_t C^T C \hat{\Sigma}_t.
\]

The sde (15) is exact, i.e,

\[
P(S_t | \mathcal{Z}_t) = P(X_t | \mathcal{Z}_t) \quad \text{for} \quad t > 0,
\]

if \( P(S_0) = P(X_0) \).

**Remark 5**: The matrix equation (16) is the Lyapunov equation with \( \hat{\Sigma}_t > 0 \). A unique solution \( G_t \) exists and is furthermore a symmetric matrix. Next comparisons are drawn between the optimal transport sde (15) and the FPF (4). To aid the comparison \( G_t \) is expressed as,

\[
G_t = A + \frac{1}{2} \hat{\Sigma}_t^{-1} - \frac{1}{2} \hat{\Sigma}_t^{-1} C^T C + \Omega_t \hat{\Sigma}_t^{-1},
\]

where \( \Omega_t \) is the unique solution to the following matrix equation,

\[
\Omega_t \hat{\Sigma}_t^{-1} + \hat{\Sigma}_t^{-1} \Omega_t = A^T - A + \frac{1}{2} \hat{\Sigma}_t^{-1}(C^T C - C^T C) \hat{\Sigma}_t.
\]

Using (17), the optimal transport sde (15) has the form,

\[
dS_t = AS_t \, dt + \frac{1}{2} \hat{\Sigma}_t^{-1}(S_t - \hat{S}_t) \, dt + \hat{K}_t (dZ_t - \frac{CS_t + C\hat{S}_t}{2} \, dt) + \Omega_t \hat{\Sigma}_t^{-1}(S_t - \hat{S}_t) \, dt.
\]

Comparing to the original FPF (4) there are two differences.

1) The stochastic term \( dB_t \) is replaced with the deterministic term, \( \frac{1}{2} \hat{\Sigma}_t^{-1}(S_t - \hat{S}_t) \, dt \), which similar to scalar case.

2) In the vector case, there is an "extra" term \( \Omega_t \hat{\Sigma}_t^{-1}(S_t - \hat{S}_t) \, dt \), where \( \Omega_t \) is the skew symmetric matrix, solution to (18). The extra term does not effect the distribution \( S_t \). This term can be viewed as a correction term that serves to make the dynamics "symmetric" and hence optimal in the optimal transportation sense. This term was absent in scalar case because the only \( 1 \times 1 \) skew symmetric matrix is 0.

**Remark 6**: In practice, solving the Lyapunov equation (16) for \( G_t \) requires additional computational cost, which was absent in the original FPF.

---

![Figure 3](image-url)  
**Fig. 3.** Particle trajectories obtained by simulating the two models: (a) Monte-Carlo sde (20), (b) Optimal transport sde (21).
The aim of this section is to show that a deterministic approach, such as the one discussed in this paper, can reduce the simulation variance. This phenomenon is illustrated in the simplest possible setting of simulating a Brownian motion.

**Problem statement:** Consider a real-valued stochastic process \( X_t \in \mathbb{R} \),

\[
dX_t = dB_t,
\]

where the initial condition \( X_0 \) is distributed according to a Gaussian distribution \( \mathcal{N}(0,1) \), and \( \{B_t\} \) is standard Wiener process. Indeed, the probability distribution of \( X_t \) is Gaussian \( \mathcal{N}(0,1+t) \). The objective is to approximate \( P(X_t) \) using the Monte-Carlo and the optimal transport methods, and to compare the accuracy of these methods.

**Monte-Carlo:** In the first method, a system of \( N \) particles are simulated independently according to,

\[
dS^t_i = dB^t_i, \quad \text{for } i = 1, \ldots, N,
\]

where \( \{B^t_i\} \) are independent Wiener processes, and \( S^t_0 \) are i.i.d samples drawn from initial distribution \( \mathcal{N}(0,1) \).

**Optimal transport:** In the second method, a system of \( N \) particles are simulated according to the optimal transport sde,

\[
dS^t_i = \frac{1}{2\Sigma^t_i} (S^t_i - \hat{S}^{(N)}_t) \, dt,
\]

for \( i = 1, \ldots, N \), where \( \hat{S}^t_0 \) are i.i.d samples drawn from initial distribution \( \mathcal{N}(0,1) \), and

\[
\hat{S}^{(N)}_t := \frac{1}{N} \sum_{i=1}^N S^t_i, \quad \Sigma^t_i := \frac{1}{N} \sum_{i=1}^N (S^t_i - \hat{S}^{(N)}_t)^2.
\]

Note that in the limit as \( N \to \infty \) one obtains

\[
dS_t = \frac{1}{2\Sigma_t} (S_t - \hat{S}_t) \, dt
\]

which according to Remark 4 is the deterministic counterpart of (20).

**Particle trajectories:** The trajectories of particles obtained by simulating the Monte-Carlo model (20) and the optimal transport model (21) are depicted in Figure 3(a) and Figure 3(b), respectively.

**Estimating the mean:** The mean of \( X_t \) is estimated as the empirical mean of the particles,

\[
\mathbb{E}[X_t] \approx \hat{S}^{(N)}_t = \frac{1}{N} \sum_{i=1}^N S^t_i
\]

Figure 4(a) and Figure 4(b) depict the result for Monte-Carlo method and Optimal transport method respectively. As expected, the optimal transport method, has less "randomness". In fact, one can obtain an explicit formula for the empirical mean in this case.

\[
d\hat{S}^{(N)}_t = \frac{1}{N} \sum_{i=1}^N dS^t_i = \frac{1}{N} \sum_{i=1}^N \frac{1}{2\Sigma^t_i} (S^t_i - \hat{S}^{(N)}_t) \, dt = 0.
\]

So, the empirical mean remains constant, equal to its initial value,

\[
\hat{S}^{(N)}_t = \hat{S}^{(N)}_0.
\]

**Simulation variance:** Any estimate obtained using a simulation with finitely many particles is a random variable. The variance of this random variable is called simulation variance; c.f. [1]. Figure 5(a) and Figure 5(b) depict the mean and variance of \( \hat{S}^{(N)}_t \), obtained after averaging over 500 simulations, for the Monte-Carlo method and the optimal transportation method respectively. The figure shows that the simulation variance for the Monte-Carlo method increases with time, whereas for the optimal transport method, it remains constant. In this simple case, this observation can also be verified analytically:
A. Proof of Prop. 1

The other extension would be to consider the general model (21), and the second order term is uniformly bounded for all $t$.

Let $X \sim \mathcal{N}(\hat{X}, \Sigma_X)$, and $Y := T(X) \sim \mathcal{N}(\hat{Y}, \Sigma_Y)$. Consider the special case $\hat{X} = \hat{Y} = 0$. The transportation cost is,

$$E[|Y - X|^2] = E[X^TY] + E[Y^TY] - 2E[X^TY] = \text{Tr}(\Sigma_X) + \text{Tr}(\Sigma_Y) - 2E[X^TY].$$

Remark 7: The result shows that in the optimal transport model (21), $S_1^N, \ldots, S_N^N$ are correlated and the correlation serves to reduce the simulation variance.

VI. Conclusion

In this paper, an optimal transport formulation of the FPF algorithm is introduced. The optimal transport FPF is obtained by employing a time-step optimization procedure. In future, one would like to relate the current optimization procedure to an optimal control problem with an integral cost. The other extension would be to consider the general nonlinear non-Gaussian filtering problem.

APPENDIX

A. Proof of Prop. 1

The proof for the Monte-Carlo method, the simulation variance grows with time, whereas for the optimal transport method, it remains constant. In fact, the same result holds true for estimating the variance as shown in Figure 6(a) and Figure 6(b), and explicitly given by,

Monte-Carlo: $\text{Var}(\hat{S}_1^N) \approx \frac{3(1+t)^2}{N}$,

Optimal transport: $\text{Var}(\hat{S}_1^N) \approx \frac{3}{N}$.

B. Proof of Prop. 2 and Prop. 3

The key step in the proof is the following Lemma, Lemma 1: Consider the ode (3b). Let $\Sigma_t$ be the solution for $t \in [0, T]$. Then the following relationship holds,

$$\Sigma_t^2 = \Sigma_t^2 \Sigma_t^2 \Sigma_t^2 = I + G_t \Delta t + O(\Delta t^2), \quad (22)$$

where $G_t$ is the solution to the matrix equation,

$$G_t = \Sigma_t + \Sigma_t G_t = A \Sigma_t + \Sigma_t A^T + I - \Sigma_t C^T \Sigma_t, \quad (23)$$

and the second order term is uniformly bounded for all $t \in [0, T]$.

Proof: The solution $\Sigma_t$ is positive and bounded since the system is observable [13]. Fix $t \in [0, T]$, and define

$$F(s) := \Sigma_t^2 (\Sigma_{t+s}^2 \Sigma_{t+s})^{-1} \Sigma_{t+s}^2.$$
Since the solution to the Lyapunov equation (23) is unique, bounded for all \( N \) sians, the key step in the procedure is to obtain the optimal map is, where the number of particles \( k \) to obtain the sde, the time stepping procedure is used. The associated sde is,

\[
dS_t = d\tilde{X}_t + G_t(S_t - \tilde{X}_t) dt,
\]

where \( d\tilde{X}_t \) is given by (3a). Finally one obtains (15) by replacing \( \tilde{X} \) and \( \Sigma \) with \( \tilde{S} \) and \( \tilde{\Sigma} \) respectively, which are identical by exactness.

Proof of Prop 2 is a special case of this proof.

References

[1] S. Asmussen and P. W. Glynn. Stochastic Simulation: Algorithms and Analysis, volume 57. Springer Science & Business Media, 2007.
[2] J. Benamou and Y. Brenier. A computational fluid mechanics solution to the Monge–Kantorovich mass transfer problem. Numerische Mathematik, 84(3):375–393, 2000.
[3] K Berntorp. Feedback particle filter: Application and evaluation. In 18th Int. Conf. Information Fusion, Washington, DC, 2015.
[4] Y. Cheng and S. Reich. A McKean optimal transportation perspective on Feynman-Kac formulae with application to data assimilation. arXiv preprint arXiv:1311.6300, 2013.
[5] D. Crisan and J. Xiong. Approximate McKean-Vlasov representations for a class of SPDEs. Stochastics An International Journal of Probability and Stochastics Processes, 82(1):53–68, 2010.
[6] F. Daum, J. Huang, and A. Nourshin. Exact particle flow for nonlinear filters. In SPIE Defense, Security, and Sensing, pages 769704–769704, 2010.
[7] A. Dominitz and A. Tannenbaum. Texture mapping via optimal mass transport. Visualization and Computer Graphics, IEEE Transactions on, 16(3):419–433, 2010.
[8] T A El Moselhy and Y. M. Marzouk. Bayesian inference with optimal maps. Journal of Computational Physics, 231(23):7815–7850, 2012.
[9] L. C. Evans. Partial differential equations and Monge-Kantorovich mass transfer. Current developments in mathematics, pages 65–126, 1997.
[10] C. R. Givens, R. M. Shortt, et al. A class of Wasserstein metrics for probability distributions. Michigan Math. J, 31(2), 1984.
[11] R. Ma and T. P. Coleman. Generalizing the posterior matching scheme to higher dimensions via optimal transportation. In Communication, Control, and Computing (Allerton), 2011 49th Annual Allerton Conference on, pages 96–102, 2011.
[12] S. K. Mitter and N. J. Newton. A variational approach to nonlinear filters. In Stochastic Simulation: Algorithms and Optimization, 34(1):226–243, 1996.
[13] D. Ocone and E. Pardoux. Asymptotic stability of the optimal filter with respect to its initial condition. SIAM Journal on Control and Optimization, 42(5):1813–1833, 2003.
[14] M. Parno and Y. Marzouk. Transport map accelerated Markov chain Monte Carlo. arXiv preprint arXiv:1412.5492, 2014.

[15] S. Reich. A dynamical systems framework for intermittent data assimilation. BIT Numerical Mathematics, 51(1):235–249, 2011.

[16] P. M. Stano, A. K. Tilton, and R. Babuška. Estimation of the soil-dependent time-varying parameters of the hopper sedimentation model: The FPF versus the BPF. Control Engineering Practice, 24:67–78, 2014.

[17] A. K. Tilton, E. T. Hsiao-Wecksler, and P. G. Mehta. Filtering with rhythms: Application of the feedback particle filter to estimation of human gait cycle. In Proc. of the 2012 American Control Conference, pages 3433–3438, Montréal, Canada, June 2012.

[18] C. Villani. Topics in Optimal Transportation. Number 58. American Mathematical Soc., 2003.

[19] T. Yang, R. S. Laugesen, P. G. Mehta, and S. P. Meyn. Multivariable feedback particle filter. In Proc. of 51st IEEE Conf. Decision and Control, pages 4063–4070, Maui, HI, Dec 2012.

[20] T. Yang, P. G. Mehta, and S. P. Meyn. Feedback particle filter. IEEE Trans. Automatic Control, 58(10):2465–2480, October 2013.