arXiv:2011.13804v1 [math.OC] 27 Nov 2020

A UNIFICATION OF WEIGHTED AND UNWEIGHTED PARTICLE FILTERS*

EHSAN ABEDI†, SIMONE CARLO SURACE†, AND JEAN-PASCAL PFISTER†

Abstract. Particle filters (PFs), which are successful methods for approximating the solution of the filtering problem, can be divided into two types: weighted and unweighted PFs. It is well-known that weighted PFs suffer from the weight degeneracy and curse of dimensionality. To sidestep these issues, unweighted PFs have been gaining attention, though they have their own challenges. The existing literature on these types of PFs is based on distinct approaches. In order to establish a connection, we put forward a framework that unifies weighted and unweighted PFs in the continuous-time filtering problem. We show that the stochastic dynamics of a particle system described by a pair process, representing particles and their importance weights, should satisfy two necessary conditions in order for its distribution to match the solution of the Kushner-Stratonovich equation. In particular, we demonstrate that the bootstrap particle filter (BPF), which relies on importance sampling, and the feedback particle filter (FPF), which is an unweighted PF based on optimal control, arise as special cases from a broad class, and that there is a smooth transition between the two. The freedom in designing the PF dynamics opens up potential ways to address the existing issues in the aforementioned algorithms, namely weight degeneracy in the BPF and gain estimation in the FPF.

Key words. Nonlinear filtering, Kushner-Stratonovich equation, Fokker-Planck equation, Stochastic differential equations, McKean-Vlasov processes, Interacting particle systems, Importance sampling.

AMS subject classifications. 60G35, 60H15, 65C35, 65C05, 35Q84.

1. INTRODUCTION.

1.1. Filtering Problem. The goal of filtering is to dynamically estimate a latent variable from noisy observations. We consider the continuous-time nonlinear filtering problem (in one dimension for the sake of simplicity), where $X_t \in \mathbb{R}$ is the *hidden* process satisfying an Itô stochastic differential equation (SDE) and $Y_t \in \mathbb{R}$ is the *observation* process evolving according to an Itô SDE which depends on X_t:

$$dX_t = f(X_t, t)\, dt + g(X_t, t)\, dB_t^{\mathrm{X}}, \quad X_0 \sim P_0, \tag{1.1}$$

$$dY_t = h(X_t, t)\, dt + dB_t^{\mathrm{Y}}, \tag{1.2}$$

where $B_t^{\mathrm{X}}, B_t^{\mathrm{Y}} \in \mathbb{R}$ are independent Brownian motions (BMs) and $f(x,t)$, $g(x,t)$, and $h(x,t)$ are (known) functions that map $\mathbb{R} \times \mathbb{R}_{\geq 0} \mapsto \mathbb{R}$ and called the drift, diffusion, and observation function, respectively. The initial condition X_0 is independent of the BMs and has (known) distribution P_0 with finite second moment and $Y_0 = 0$ (no observation at first). Let $\mathcal{B}(\mathbb{R})$ be the Borel σ-algebra on $\mathbb{R}$. The filtering problem is to find the (regular) conditional distribution of the hidden process X_t given the history of observations $\mathcal{F}_t^Y := \sigma(Y_s : 0 \leq s \leq t)$, that is, $P_t(B) := \Pr(X_t \in B | \mathcal{F}_t^Y)$ for any Borel subset $B \in \mathcal{B}(\mathbb{R})$. This distribution is referred to as the *filtering distribution* and in the statistics literature it is also called the *posterior distribution*, while $\Pr(X_t \in B)$ is called the *prior distribution*. We let $p(x,t)$ denote the probability density function corresponding to P_t with respect to Lebesgue measure, if it exists.

Throughout the paper we assume that f, g satisfy the conditions for the well-posedness of SDEs (e.g. locally Lipschitz in x uniformly in t, see e.g. [12], Theorem

*Submitted to the editors November 26, 2020.

Funding: This research was funded by the Swiss National Science Foundation (SNSF) grants PP00P3_179060 and 31003A_175644.

†University of Bern, Department of Physiology, Bern, Switzerland (ehsan.abedi@alumni.epfl.ch, simone.surace@unibe.ch, jeanpascal.pfister@unibe.ch).

5.4). In addition, to ensure the existence of the filtering density p, we assume that $g^2 \geq \delta > 0$ is bounded below by a constant and that $f \in C_b^{1,0}$, $g^2 \in C_b^{2,0}$, and $h \in C_b^{1,0}$, where $C_b^{\kappa,j}$ denotes the space of bounded continuous functions with bounded partial derivatives up to order κ in x and j in t (see [2], Theorem 7.17, for more details).

The stochastic processes in this paper are defined on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t\geq 0})$ (satisfying the usual conditions) and are assumed to be progressively measurable (hence adapted) with respect to the filtration $(\mathcal{F}_t)_{t\geq 0}$. We further denote by $\mathbb{L}^\kappa(0,T)$ the space of processes $(F_t)_{T\geq t\geq 0}$ with $\mathbb{E}[\int_0^T |F_t|^\kappa dt] < \infty$. All SDEs are in the Itô sense and primes $(')$ denote the (partial) derivative with respect to x.

1.2. Formal Solution. It is well known that the evolution of p is described by the Kushner-Stratonovich equation (KSE) [14, 21]

$$dp(x,t) = \mathcal{L}_t^\dagger p(x,t)\, dt + p(x,t)\big(h(x,t) - \hat{h}_t\big)(dY_t - \hat{h}_t dt), \tag{1.3}$$

where $\mathcal{L}_t^\dagger \cdot = -\frac{\partial}{\partial x}(f(x,t)\cdot) + \frac{1}{2}\frac{\partial^2}{\partial x^2}(g^2(x,t)\cdot)$ is the adjoint Fokker-Planck operator and $\hat{h}_t := \mathbb{E}[h(X_t,t)|\mathcal{F}_t^Y] = \int_{\mathbb{R}} h(x,t)p(x,t)dx$. The initial condition is assumed to be $p(x,0) = p_0(x) \in C^2$. The KSE consists of two parts: the first part is associated with the prior dynamics given by the Fokker-Planck equation, while the second part can be interpreted as a correction resulting from observations, which is proportional to the so-called *innovation* term $(dY_t - \hat{h}_t dt)$. Equation (1.3) can be equivalently converted into the evolution of a given statistic using integration-by-parts. Let $\phi \in C^2$ such that $\mathbb{E}[|\phi(X_t)|] < \infty$ for all $t \geq 0$. The conditional expectation of $\phi(X_t)$ evolves as

$$d\mathbb{E}[\phi(X_t)|\mathcal{F}_t^Y] = \mathbb{E}[\mathcal{L}_t\phi(X_t)|\mathcal{F}_t^Y]\, dt + \mathbb{E}[\phi(X_t)(h(X_t,t) - \hat{h}_t)|\mathcal{F}_t^Y](dY_t - \hat{h}_t dt), \tag{1.4}$$

where $\mathcal{L}_t \cdot = f(x,t)\frac{\partial}{\partial x}\cdot + \frac{1}{2}g^2(x,t)\frac{\partial^2}{\partial x^2}\cdot$ is the generator of the process X_t.

Example (Linear-Gaussian Case). *We will use this simple case throughout the paper as an illustration of key concepts. The linear-Gaussian case is characterized by linear drift terms and additive noise, as well as a Gaussian initial distribution:*

$$f(x,t) = ax, \quad g(x,t) = b, \quad h(x,t) = cx, \qquad P_0 = \mathcal{N}(\mu_0, \rho_0), \tag{1.5}$$

where $a, b, c \in \mathbb{R}$. In this case the KSE (1.3) can be solved in closed-form (even though the coefficients f and h do not satisfy the boundedness as discussed in subsection 1.1). We have $X_t|\mathcal{F}_t^Y \sim \mathcal{N}(\hat{\mu}_t, \hat{\rho}_t)$ given by Kalman-Bucy's [11] classical result

$$d\hat{\mu}_t = a\hat{\mu}_t\, dt + c\hat{\rho}_t(dY_t - c\hat{\mu}_t\, dt), \tag{1.6}$$

$$d\hat{\rho}_t = b^2\, dt + 2a\hat{\rho}_t\, dt - c^2\hat{\rho}_t^2\, dt. \tag{1.7}$$

The coefficient of the innovation term is known as the Kalman gain $\bar{K}(t) := c\hat{\rho}_t$.

In contrast to the linear-Gaussian case, for most signal and observation models equation (1.3) does not have closed-form solutions. Likewise, equation (1.4), when it is applied to the moments of the filtering distribution, gives rise to a closure problem in which the evolution of the n-th moment $\mathbb{E}[X_t^n|\mathcal{F}_t^Y]$ generally depends on higher-order moments. Therefore, the KSE is only a "formal" solution to the filtering problem and needs to be approximated numerically in practice. Among the numerical methods, particle filters (PFs) have been widely and successfully applied because of their versatility. Below, they will be presented within a broader framework, but see e.g. [2], section 8.6, and the tutorials [7, 15] for more detailed surveys of the PFs.

1.3. Particle Filters. These methods are aimed at approximating the filtering distribution by the empirical distribution of a particle system, which in full generality is a triangular array of random variables [5] for some fixed t,

$$\{S_t^{(i,N)}, w_t^{(i,N)}\}_{i=1}^N \quad \text{with } \sum_{i=1}^N w_t^{(i,N)} = 1, \tag{1.8}$$

where $S_t^{(i,N)}$ are samples, also called *particles*, and $w_t^{(i,N)}$ are their corresponding *importance weights*, which without loss of generality are assumed to be normalized. To justify such a method, one has to study the N-particle system and show that the sequence of empirical distributions converges (at least in a weak sense) to the filtering distribution as $N \to \infty$. In this article, however, we use a mean-field-limit approach, i.e. we study abstract "particle systems" characterized by a *pair process* denoted by (S_t, W_t). The following definition is modified from [5], section 2.1.

DEFINITION 1 (Targeting Condition). *The particle system described by a pair process (S_t, W_t), representing particles and their weights respectively, is said to target the filtering distribution in the filtering problem (1.1)-(1.2) at time t if and only if*

$$\mathbb{E}[W_t \phi(S_t) | \mathcal{F}_t^Y] = \mathbb{E}[\phi(X_t) | \mathcal{F}_t^Y] \quad a.s. \tag{1.9}$$

holds for any measurable function $\phi(x) : \mathbb{R} \mapsto \mathbb{R}$ such that $\mathbb{E}[|\phi(X_t)|] < \infty$.

It should be noted that weights for a given S_t are *not unique*. For instance, if the pair (S_t, W_t) satisfies the targeting condition, then any pair $(S_t, W_t + V_t)$ is also a solution, where V_t is a stochastic process independent of S_t with zero conditional mean and finite second moment. In subsection 2.1, we further clarify this non-uniqueness.

To have the targeting condition over a period of time, i.e. a dynamic version of (1.9), we are interested in the time-evolution of (S_t, W_t). As we shall see later in subsection 2.2, this naturally leads to *McKean-Vlasov* SDEs, in which the coefficients in the dynamics of (S_t, W_t) become dependent on the targeted distribution. In numerical implementations, the left-hand side of (1.9) is approximated by a Monte-Carlo estimate $\frac{1}{M}\sum_{i=1}^N W_t^{(i,N)} \phi(S_t^{(i,N)})$ using samples $(S_t^{(i,N)}, W_t^{(i,N)})$, $i = 1, ..., N$ from the joint distribution of (S_t, W_t) given $\mathcal{F}_t^Y$, where $M := \sum_{i=1}^N W_t^{(i,N)}$, which is of order N. If the samples are chosen appropriately (e.g. i.i.d.), the Monte-Carlo estimate converges to the conditional expectation by a law of large numbers and yields asymptotic consistency of the PF. The normalized weights are then given by $w_t^{(i,N)} = \frac{1}{M} W_t^{(i,N)}$. Thus, in theory, if we combine Definition 1 with appropriate sampling, we obtain an asymptotically exact filter. In practice, however, propagation of the samples requires the McKean-Vlasov terms to be estimated based on the current sample. This estimation problem is nontrivial and introduces correlations between samples, which complicates the convergence analysis. Although interesting and worthwhile, the issues of estimation and convergence are not within the scope of the present article. Instead, we focus on the characterization of abstract particle systems.

A particle filter is said to be *unweighted* if $W_t = 1$ for all t, otherwise, it is called *weighted*. Observe that in particular, (1.9) implies (by setting $\phi = 1$) that $\mathbb{E}[W_t | \mathcal{F}_t^Y] = 1$ a.s. and due to the non-negativity of the variance of W_t, we also have $\mathbb{E}[W_t^2 | \mathcal{F}_t^Y] \geq 1$ a.s. If W_t deviates significantly from unity, this means that the Monte-Carlo variance of the weighted average is larger than it would be for an average with equal weights. This can be measured by the *effective sample size*, a commonly-used approximation of which (see e.g. [17], section 2) is given by

$$N_{\text{eff}} := \Big(\sum_{i=1}^N (w_t^{(i,N)})^2 \Big)^{-1}. \tag{1.10}$$

Since $w_t^{(i,N)} = \frac{1}{M} W_t^{(i,N)}$ as explained before, by a law of large numbers we have $M/N \to 1$ as $N \to \infty$ and therefore

$$\frac{N}{N_{\text{eff}}} = \Big(\frac{N}{M}\Big)^2 \frac{1}{N} \sum\nolimits_{i=1}^{N} \Big(W_t^{(i,N)}\Big)^2 \xrightarrow{p} \mathbb{E}[W_t^2|\mathcal{F}_t^Y] \qquad \text{as } N \to \infty, \tag{1.11}$$

provided samples become i.i.d. for large N. So in order to keep N_{eff} close to N, it is desirable that $\text{Var}[W_t|\mathcal{F}_t^Y] = \mathbb{E}[W_t^2|\mathcal{F}_t^Y] - 1$ remain close to zero. In order to assess the degeneracy of a PF algorithm, we are essentially interested in minimizing the unconditional variance of the importance weights, $\text{Var}[W_t] = \mathbb{E}[W_t^2] - 1$, taking all possible realizations of the observation process into account [6]. In subsection 2.1, we will establish the minimum-variance weight for a fixed particle distribution. We now review two well-known examples of PFs within the framework above.

• **The Bootstrap Particle Filter (BPF)**. The BPF is a weighted PF that was originally introduced by [10] and is widely used in discrete-time filtering [7]. Here we present its continuous-time formulation (see e.g. [2], chapter 9, or [15], section 6.1). The particles in this filter move with the same law as the hidden process, thereby being distributed according to the prior. The weight dynamics must consequently include observations in such a way that the weighted particles are distributed according to the posterior. The evolution of the particle system denoted by $(S_t^{\text{B}}, W_t^{\text{B}})$ reads

$$dS_t^{\text{B}} = f(S_t^{\text{B}}, t)\, dt + g(S_t^{\text{B}}, t)\, dB_t, \tag{1.12}$$

$$dW_t^{\text{B}} = W_t^{\text{B}} (h(S_t^{\text{B}}, t) - \hat{h}_t)\, (dY_t - \hat{h}_t dt), \tag{1.13}$$

with initial condition $S_0^{\text{B}} \sim P_0$ and $W_0^{\text{B}} = 1$, and B_t being a BM independent of $\{(B_t^{\text{X}})_{t\geq 0}, (B_t^{\text{Y}})_{t\geq 0}, X_0, S_0^{\text{B}}\}$ (note that in practice, particles are driven by independent BMs). The usual derivation of the BPF is based on a change of probability measure, in which W_t^{B} is the evaluation of the Radon–Nikodym derivative $d\mathbb{P}/d\mathbb{Q}$ with X_t replaced by S_t^{B}, where $\mathbb{P}$ is the (original) coupled measure of the system (X_t, Y_t) and $\mathbb{Q}$ is a new measure under which X_t and Y_t are independent, the dynamics of X_t remains unchanged, and Y_t corresponds to a BM, $dY_t = dB_t^{\text{Y}}$, ([2], chapter 9):

$$W_t^{\text{B}} = \frac{\tilde{W}_t^{\text{B}}}{\mathbb{E}_{\mathbb{Q}}[\tilde{W}_t^{\text{B}}|\mathcal{F}_t^Y]}, \quad \tilde{W}_t^{\text{B}} := \mathbb{E}_{\mathbb{Q}}\left[\frac{d\mathbb{P}}{d\mathbb{Q}}\middle|\mathcal{F}_t\right]. \tag{1.14}$$

Girsanov's theorem and Itô's formula then give rise to the equation (1.13).

Implementing the dynamics (1.12)-(1.13) in practice is straightforward. However, the BPF suffers from the weight decay, that is, most of the weights become negligibly small and only a few of them remain significant, an issue which becomes even more severe in high dimensions [22]. A common practice to overcome this issue is to periodically resample the particles, in which case equations (1.12)-(1.13) merely describe the evolution of the particle system between the resampling times. Although resampling techniques are key ingredients of weighted PFs, they are not the focus of the current work. Instead, we attempt to find ways to alleviate the weight collapse itself.

• **The Feedback Particle Filter (FPF).** The FPF is an unweighted PF that was introduced by [28], initially motivated by mean-field optimal control. In contrast to the BPF, this filter does not have weight dynamics. Instead, the particles must incorporate the observations and *interact* with each other so that they can target the filtering distribution by themselves. The key idea is to add a correction term, also called the control input, to the prior dynamics of particles,

$$dS_t^{\text{F}} = f(S_t^{\text{F}}, t)\, dt + g(S_t^{\text{F}}, t)\, dB_t + U(S_t^{\text{F}}, t)\, dt + K(S_t^{\text{F}}, t)\, dY_t, \tag{1.15}$$

$$dW_t^{\text{F}} = 0, \tag{1.16}$$

and find the unknown functions U and K such that the conditional density of S_t given $\mathcal{F}_t^Y$ solves the KSE (1.3). Here again the initial condition is $S_0^{\mathrm{F}} \sim P_0$ and $W_0^{\mathrm{F}} = 1$ and B_t is a BM independent of $\{(B_t^{\mathrm{X}})_{t\geq 0}, (B_t^{\mathrm{Y}})_{t\geq 0}, X_0, S_0^{\mathrm{F}}\}$. The paper [28] shows that $U(x,t)$ and $K(x,t)$ under certain technical assumptions must satisfy

$$\frac{\partial}{\partial x}(Kp) + (h - \hat{h}_t)p = 0, \tag{1.17}$$

$$U = -\frac{1}{2}K\big(h + \hat{h}_t - \frac{\partial}{\partial x}K\big). \tag{1.18}$$

The main challenge in this method is to find K, also called the *gain function*, for which we need to solve (1.17), which in turn depends on p. A standard solution for K takes a gradient-form, where (1.17) then becomes a weighted *Poisson equation*.

Example (Linear-Gaussian Case, Continued). *In this case,* (1.17)-(1.18) *have closed form solutions and the gradient-form solution for K equals the Kalman gain.*

Notice that the FPF does not require resampling any procedure, which is an advantage compared to the BPF. Morover, [22] shows numerically that the FPF is less prone to the curse of dimensionality (COD) than the BPF. Table 1 summarizes the comparison between the BPF and FPF.

Table 1

A comparison between the BPF and the FPF as they have been treated in the literature so far. Each has its strengths (+) *and weaknesses* (−).

Bootstrap Particle Filter [10]	**Feedback Particle Filter** [28]
a weighted particle filter	an unweighted particle filter
based on change of probability measure	motivated by optimal control
asymptotically exact (+)	asymptotically exact (+)
suffers from weight degeneracy (−)	requires gain estimation (−)
exhibits the COD (−)	potentially avoids the COD (+)

1.4. Motivation and Contribution. So far we have seen two well-known PFs, each of which was originally based on a distinct approach, yet they both satisfy the targeting condition (Definition 1). Given the comparison between these methods, it is still unknown whether combining the strengths of both types is possible. As a step towards this goal, we provide a unified treatment of a large family of weighted and unweighted PFs that is based solely on the targeting condition. Specifically,

- We first characterize the weights W_t in terms of a Radon-Nikodym derivative of the marginal distributions over S_t and find the optimal (i.e. minimum-variance) weight W_t^* for fixed particle distribution (see Theorem 2). This result also sheds some light on the non-uniqueness of importance weights in particle filtering.
- We then introduce a general dynamics for the particle system (S_t, W_t) expressed as a system of SDEs and obtain necessary conditions on its coefficients for targeting the filtering distribution (see Theorem 3, referred to as the "unifying theorem"). The necessary conditions are a system of ordinary differential equations (ODEs) in one dimension, which becomes a system of partial differential equations (PDEs) in higher dimensions. The results also hold in the unconditional case (see Corollary 4).
- As a result of the unifying theorem, we derive a class of PFs which encompasses the BPF and FPF with a smooth transition between them, thereby indicating that these methods are not different in their nature (see Proposition 5).
- The optimal importance weight W_t^* from the first theorem is studied in the context of the unifying theorem and its evolution dW_t^* is derived (see Proposition 6).

- Finally, we outline two potential applications of the unifying theorem in section 3, namely compensating for gain estimation errors with weight dynamics and providing freedom to alleviate the weight degeneracy.

For the sake of simplicity, we develop the results in the one-dimensional setting, but all results can be generalized to the multi-dimensional case.

1.5. Related Work. A unifying framework for discrete- and continuous-time filtering (as a Bayesian formulation of the data assimilation problem) from the perspective of couplings, optimal transport, and Schrödinger bridges is proposed by [19]. Similarly, in [18], three types of unweighted PFs, including the FPF, are unified in the framework of McKean-Vlasov SDEs. However, to the best of our knowledge a unification of weighted and unweighted approaches such as this one has not been attempted. This question, or more specifically how to reconcile the dynamics of the FPF with the importance weights as Radon-Nikodym derivatives of path measures, was posed in [22], section 3.2. The present paper shows that such a reconciliation is indeed possible outside of the change-of-measure framework adopted in the literature on continuous-time PFs.

This paper also touches on the notion of non-uniqueness in designing the PF, which has been discussed in the literature but mainly restricted to the linear-Gaussian case and freedom of the particle movements. For example, [1] provides a systematic exploration of the non-uniqueness within the class of linear FPF in terms of gauge transforms and [25] studies the non-uniqueness of the feedback control law in particle dynamics for different types of ensemble Kalman filters. In this work, however, we examine the non-uniqueness in the nonlinear case and in a more general setting that includes weight dynamics.

2. MAIN RESULTS.

2.1. A Characterization of Weights in Terms of the Radon-Nikodym Derivatives. The approach used to derive the importance weights in the BPF (1.14) is difficult to reconcile with filters such as the FPF, in which observation terms are included in the particle dynamics. This explicit dY term makes the measure of the particles singular with respect to the measure of the hidden process (see [22], section 3.2, for a discussion of this issue). Here we demonstrate a general relationship between the process W_t and the Radon-Nikodym derivative of the marginal distributions over S_t, which is consistent with both the BPF and FPF, and we also provide the minimum-variance choice for W_t (for a fixed particle distribution).

THEOREM 2. *Consider the filtering problem (1.1)-(1.2) with filtering distribution $P_t(B) := \Pr(X_t \in B|\mathcal{F}_t^Y)$ for any $B \in \mathcal{B}(\mathbb{R})$. Let (S_t, W_t) be a pair process characterizing an abstract particle system. Denote by Q_t the conditional distribution of S_t given $\mathcal{F}_t^Y$, i.e. $Q_t(B) := \Pr(S_t \in B|\mathcal{F}_t^Y)$. At any time t, under the condition that the particle system described by (S_t, W_t) targets P_t, according to Definition 1, we have:*

I. The distribution P_t is absolutely continuous with respect to the distribution Q_t. In particular, the Radon-Nikodym derivative dP_t/dQ_t exists, and

$$\mathbb{E}[W_t|\mathcal{F}_t^Y, S_t] = \frac{dP_t}{dQ_t}(S_t) \quad a.s. \tag{2.1}$$

II. $W_t^* := \frac{dP_t}{dQ_t}(S_t)$ mimimizes $\mathbb{E}[W_t^2]$ subject to (2.1).

The proof is given in subsection 5.1, and by reversing the arguments in the proof it is obvious that (2.1) is already sufficient to guarantee the targeting condition (1.9).

Notice also that if P_t and Q_t have densities $p(x,t)$ and $q(x,t)$ with respect to the Lebesgue measure, then equation (2.1) turns into

$$\mathbb{E}[W_t|\mathcal{F}_t^Y, S_t] = \frac{p(S_t,t)}{q(S_t,t)} \quad a.s. \tag{2.2}$$

As we observed in subsection 1.3, the importance weights are not unique. It is now also evident from equation (2.1) that there are many solutions which are not necessarily optimal. A large class of suboptimal choices for W_t is of the form $(d\tilde{P}_t/d\tilde{Q}_t)(Z_t, S_t)$, where $\tilde{P}_t$ and $\tilde{Q}_t$ are respectively the joint conditional (on $\mathcal{F}_t^Y$) distributions of (Z_t, X_t) and (Z_t, S_t) with Z_t being some arbitrary $\mathcal{F}_t$-measurable process such that the Radon-Nikodym derivative exists, and the second marginals of $\tilde{P}_t$ and $\tilde{Q}_t$ agree, respectively, with P_t and Q_t. For example, we can let $Z_t = S_0$, which takes some information from the past into account. Also the weights in the BPF (1.14) are of this form, as by the disintegration of measures and independence of $(X_t)_{t\geq 0}$ and $(Y_t)_{t\geq 0}$ under the measure $\mathbb{Q}$[1], we have $\mathbb{E}_{\mathbb{Q}}[\frac{d\mathbb{P}}{d\mathbb{Q}}|\mathcal{F}_t] = \mathbb{E}_{\mathbb{Q}}[\frac{d\mathbb{P}}{d\mathbb{Q}}|\mathcal{F}_t^Y] \times \frac{dP_t}{dQ_t}(X_t)$ and hence the conditional expectation of the (normalized) Radon-Nikodym derivative of the full measures (after replacing X_t by S_t^{B}) reduces to $\frac{dP_t}{dQ_t}(S_t^{\mathrm{B}})$.

In the FPF, we have by construction that $P_t = Q_t$ and therefore $W_t^{\mathrm{F}} = 1$, which is also the optimal choice – in fact, it is globally optimal as it achieves $\mathrm{Var}[W_t^{\mathrm{F}}] = 0$. Accordingly, this result reconciles the BPF with the FPF and all other filters for which W_t as a pathwise Radon-Nikodym derivative does not make sense. Theorem 2 says that the conditional expectation of W_t can always be interpreted as a Radon-Nikodym derivative, and is a version of (i.e. almost surely equal to) the density of the filtering distribution P_t with respect to the particle distribution Q_t. It also shows that in the BPF, W_t^{B} is different from the optimum W_t^* given that we fix the particle dynamics to the prior dynamics. This is not surprising in view of the well-known degeneracy problem of the BPF (see [22] and the references therein).

2.2. The Unifying Theorem. Considering the pair process (S_t, W_t), which represents particles and their importance weights respectively as explained in subsection 1.3, we are trying to find conditions on the dynamics of (S_t, W_t) that allow the particle system to target the filtering distribution, as defined in Definition 1. Here we make an ansatz that the stochastic dynamics of the particle system takes the following general form:

$$dS_t = u(S_t,t)\,dt + k(S_t,t)\,dY_t + v(S_t,t)\,dB_t, \tag{2.3}$$

$$dW_t = W_t\big(\gamma(S_t,t)\,dt + \varepsilon(S_t,t)\,dY_t + \zeta(S_t,t)\,dB_t\big), \tag{2.4}$$

where B_t is a BM. From this starting point, our goal is to find the conditions that the unknown functions $\{u, k, v, \gamma, \varepsilon, \zeta\}$ should satisfy. To derive the results, certain assumptions are required as listed below:

(i) Regularity conditions ensuring well-posedness of the system of SDEs (2.3)-(2.4), together with (1.1)-(1.2) (see e.g. [12], Theorem 6.30).
(ii) The initial condition is $S_0 \sim P_0$ and $W_0 = 1$.
(iii) The Brownian motion B_t is independent of $\{(B_t^{\mathrm{X}})_{t\geq 0}, (B_t^{\mathrm{Y}})_{t\geq 0}, X_0, S_0\}$.
(iv) $u \in C^{1,0}$ and $k, v \in C^{2,0}$.
(v) $\gamma, \varepsilon, \zeta \in C^{1,0}$.

The second key result of our work is the following:

[1] i.e. $\mathbb{P}|_{\mathcal{F}_t^Y \vee \sigma(X_t)} = \mathbb{P}|_{\mathcal{F}_t^Y} \otimes \mathbb{P}_{X_t|\mathcal{F}_t^Y}$ and $\mathbb{Q}|_{\mathcal{F}_t^Y \vee \sigma(X_t)} = \mathbb{Q}|_{\mathcal{F}_t^Y} \otimes \mathbb{Q}|_{\sigma(X_t)}$.

THEOREM 3 (Unifying Theorem). *Consider the filtering problem (1.1)-(1.2) with filtering density $p(x,t)$, which satisfies the KSE (1.3). Let (S_t, W_t) be a pair process, representing particles and their importance weights respectively, that evolves according to the dynamics (2.3)-(2.4) under the assumptions (i)-(v). Then if the particle system described by (S_t, W_t) targets the filtering distribution for all $0 < t < T$, according to Definition 1, the functions $\{u(x,t), k(x,t), v(x,t), \gamma(x,t), \varepsilon(x,t), \zeta(x,t)\}$ satisfy the following equations for all $0 \leq t < T$*

$$\frac{\partial}{\partial x}(kp) + (h - \hat{h}_t - \varepsilon)p = 0, \tag{2.5}$$

$$\frac{1}{2}\frac{\partial^2}{\partial x^2}\big((k^2 + v^2 - g^2)p\big) - \frac{\partial}{\partial x}\big((u + k\varepsilon + v\zeta - f)p\big) + ((h - \hat{h}_t)\hat{h}_t + \gamma)p = 0, \tag{2.6}$$

in addition, the functions $\{\varepsilon(x,t), \gamma(x,t)\}$ have zero mean under the filtering distribution for all $0 \leq t < T$, that is,

$$\int_{\mathbb{R}} \varepsilon(x,t)p(x,t)dx = 0, \qquad \int_{\mathbb{R}} \gamma(x,t)p(x,t)dx = 0. \tag{2.7}$$

The proof appears in subsection 5.2. In short, the results follow from the fact that the terms multiplying dY_t and dt on both sides of $d\mathbb{E}[W_t\phi(S_t)|\mathcal{F}_t^Y] = d\mathbb{E}\big[\phi(X_t)|\mathcal{F}_t^Y\big]$ should be equal a.s., regardless of ϕ. In particular, considering $\phi = 1$ leads to the last statement (2.7), while considering the class of compactly supported test functions $\phi \in C_k^2$ yields the system of ODEs (2.5)-(2.6), which become PDEs in higher dimensions. Note that through these equations, the coefficients in the particle system dynamics depend on the targeted distribution. This means that the system of SDEs (2.3)-(2.4) are of McKean-Vlasov type. Recall that based on formula (1.9), the distribution targeted by a particle system is obtained by evaluating the left-hand side of this formula with an indicator function as a test function ϕ.

Compared to the BPF (1.12)-(1.13) and FPF (1.15)-(1.16), our ansatz (2.3)-(2.4) is clearly more general in several aspects, for example the presence of dB_t and dY_t in both dynamics. Thus, we may refer to this approach as "hybrid particle filter". As we shall show in subsection 2.4, the presence of dB_t in the weight process actually decreases its variance. Here the particle dynamics is still supposed to not involve W_t explicitly and the coefficients in the weight dynamics are intentionally chosen to be linear in W_t. This choice has an advantage as can be seen in the proof of the theorem.

The theorem above, while providing only necessary conditions for targeting the filtering distribution, sheds light on the freedom in choosing the coefficients of the particle and weight dynamics. It is easy to verify that the setting $\{u = f,\ k = 0,\ v = g,\ \zeta = 0\}$ yields the BPF (1.12)-(1.13). We demonstrate in subsection 2.3 that the FPF also satisfies the necessary conditions. Note that the first equation (2.5) is similar to the gain equation (1.17) in the FPF except the extra term εp, which arises here due to the non-zero weight dynamics. This freedom might help us compensate for the gain estimation errors with weight dynamics as will be outlined in subsection 3.1.

We close this subsection by pointing out that our results also hold in the unconditional setting, i.e. for a particle system targeting the solution of the Fokker-Planck equation. In particular, if we set $h = 0$, then the observation process Y_t does not provide any information about X_t and the KSE (1.3) reduces to the Fokker-Planck equation. The corollary below follows immediately from Theorem 3 and shows that even in this setting, there exists intrinsic freedom in constructing the dynamics of the particle system while keeping its distribution invariant.

COROLLARY 4. *Consider the stochastic process X_t satisfying the SDE (1.1) and let $\bar{p}(x,t)$ denote the probability density function of X_t, which satisfies the Fokker-Planck equation. Let (S_t, W_t) be a pair process, representing particles and their importance weights respectively, that evolves according to the dynamics below*

$$dS_t = u(S_t,t)\,dt + v(S_t,t)\,dB_t, \qquad dW_t = W_t\big(\gamma(S_t,t)\,dt + \zeta(S_t,t)\,dB_t\big), \tag{2.8}$$

under the assumptions (i)-(v), if applicable. Then if the particle system described by (S_t, W_t) targets $\bar{p}$, i.e., $\mathbb{E}[W_t\phi(S_t)] = \mathbb{E}[\phi(X_t)]$ holds for any measurable function ϕ with $\mathbb{E}[|\phi(X_t)|] < \infty$, for all $0 < t < T$, the functions $\{u(x,t), v(x,t), \gamma(x,t), \zeta(x,t)\}$ satisfy the following equation for all $0 \le t < T$

$$\frac{1}{2}\frac{\partial^2}{\partial x^2}\big((v^2 - g^2)\bar{p}\big) - \frac{\partial}{\partial x}\big((u + v\zeta - f)\bar{p}\big) + \gamma\bar{p} = 0, \tag{2.9}$$

in addition, the function $\gamma(x,t)$ satisfies $\int_{\mathbb{R}} \gamma(x,t)\bar{p}(x,t)dx = 0$ for all $0 \le t < T$.

2.3. A Class of Particle Filters. Our goal now is to introduce a class of PFs within the result of Theorem 3 that encompasses the BPF as well as the FPF. This demonstrates how these seemingly different methods can be derived from the same context. Observe that in equation (2.5), $\varepsilon = h - \hat{h}_t$ implies $k = 0$ while $\varepsilon = 0$ yields equation (1.17). Thus, a choice for ε that linearly interpolates between $h - \hat{h}_t$ and 0 makes it possible to have a smooth transition between the BPF and FPF, though it does not simplify the gain equation. The proposition below states the result.

PROPOSITION 5. *Under the assumptions of Theorem 3, and assuming that (2.5) holds, a particular solution to equation (2.6) is given by the following class:*

$$\varepsilon = \eta(h - \hat{h}_t) + \tilde{\varepsilon}, \tag{2.10}$$

$$\gamma = -(\alpha + \eta - \alpha\eta)(h - \hat{h}_t)\hat{h}_t - (1-\alpha)\tilde{\varepsilon}\hat{h}_t, \tag{2.11}$$

$$v^2 = g^2 - \beta k^2, \tag{2.12}$$

$$u = f - \tfrac{1}{2}k\big(\vartheta_1 h + \vartheta_2\hat{h}_t + (1+\beta)\tilde{\varepsilon} - (1-\beta)\tfrac{\partial}{\partial x}k\big) - v\zeta, \tag{2.13}$$

where $\tilde{\varepsilon} \in C^{1,0}$ is an arbitrary function with zero mean under the filtering distribution such that ε meets the requirements of (i), $\{\alpha,\ \beta,\ \eta\}$ are free parameters in $\mathbb{R}$, either constant or continuously time-varying, and ϑ_1, ϑ_2 are defined as follows:

$$\vartheta_1 := 1 - \beta + \eta + \beta\eta, \qquad \vartheta_2 := 1 + \beta - \eta - \beta\eta - 2\alpha. \tag{2.14}$$

The proof is simple and it is given in subsection 5.3. In particular, if in the class above, we fix $\{\tilde{\varepsilon} = 0,\ \zeta = 0,\ \beta = 0\}$ and let the others $\{\alpha,\ \eta\}$ be free, we obtain a subclass that interpolates the BPF (when $\eta = 1$) and the FPF (when $\eta = 0,\ \alpha = 0$). The parameter η can be interpreted as "observation parameter", which determines how much the observation process is incorporated into the particle or weight dynamics. We refer to α as "drift parameter", which only appears in drift functions u, γ. Lastly, we call β the "diffusion parameter" since it controls the magnitude of the diffusion coefficient v. Among these, η is the most relevant one as far as filtering is concerned.

It should be noted that not all parameter choices for $\{\alpha,\ \beta,\ \eta\}$ give rise to a "practical" particle filter. In practice, one has to check other criteria, for example the non-degeneracy of the particle distribution and the stability of the system. This can be seen explicitly in the linear-Gaussian case below.

EXAMPLE (Linear-Gaussian Case, Continued). *Consider the three-parameter $\{\alpha, \beta, \eta\}$ subclass of particle filters given by Proposition 5 after setting $\tilde{\varepsilon} = 0$ and $\zeta = 0$ for the linear-Gaussian setting (1.5) with the filtering distribution (1.6)-(1.7). Then the (unweighted) particle distribution reads $S_t|\mathcal{F}_t^Y \sim \mathcal{N}(\mu_t, \rho_t)$, where*

$$d\mu_t = a\mu_t\,dt + c\hat{\rho}_t(1-\eta)\big(dY_t - c[\tfrac{1}{2}\vartheta_1\mu_t + \tfrac{1}{2}\vartheta_2\hat{\mu}_t]\,dt\big), \tag{2.15}$$

$$d\rho_t = b^2\,dt + 2a\rho_t\,dt - c^2\hat{\rho}_t(1-\eta)[\vartheta_1\rho_t + (\beta-\beta\eta)\hat{\rho}_t]\,dt. \tag{2.16}$$

The derivation is briefly explained in subsection 5.4. Observe how particle distribution interpolates between the prior distribution (when $\eta = 1$) and the posterior distribution (when $\eta = 0, \alpha = 0$). It is easy to confirm that in the latter case, where β remains as the only free parameter, the PF resulting from Proposition 5 corresponds to the unweighted linear PF stated in [1], equation (17), with one BM, if it is indeed rewritten in terms of v. Notably, $v = 0$ recovers the deterministic linear FPF introduced in [24].

As the right-hand side of (2.12) must be non-negative, we deduce that $\beta(1-\eta)^2 \leq b^2/(c^2\hat{\rho}_t^2)$. Moreover, in the case where $\hat{\rho}_\infty := \lim_{t\to\infty}\hat{\rho}_t < \infty$, if we want to avoid that the variance grows exponentially, we need to restrict η such that $\eta^2 < 1 - 2a/(c^2\hat{\rho}_\infty)$. Constraints like these become important in numerical implementations.

2.4. Stochastic Differential of the Optimal Weight. We saw in subsection 2.1 that weights for a fixed particle distribution Q_t are not unique and indeed $W_t^* := \frac{dP_t}{dQ_t}(S_t)$ is the one that minimizes the variance. In the context of the unifying theorem, which assumes additional constraints regarding the time-evolution of the particle system, this non-uniqueness means that if we fix the particle dynamics, there are many possibilities for the weight dynamics. In other words, if we fix the functions $\{u, k, v\}$, we are then left with three unknowns $\{\gamma, \varepsilon, \zeta\}$ but only two equations (2.5)-(2.6) to constrain them. Here we demonstrate that for each choice of the dynamics of S_t according to (2.3), the dynamics of the optimal weight W_t^* also takes the form of (2.4) whose coefficients denoted by $\{\gamma^*, \varepsilon^*, \zeta^*\}$ are given by the proposition below.

PROPOSITION 6. *Fix the particle dynamics (2.3) with $S_0 \sim P_0$ under the assumptions (iii)-(iv). Then the optimal weight W_t^* from Theorem 2 satisfies the SDE*

$$dW_t^* = W_t^*\big(\gamma^*(S_t,t)\,dt + \varepsilon^*(S_t,t)\,dY_t + \zeta^*(S_t,t)\,dB_t\big), \tag{2.17}$$

where $\varepsilon^*(x,t)$, $\gamma^*(x,t)$ are the unique solutions to equations (2.5)-(2.6) after setting

$$\zeta(x,t) = \zeta^*(x,t) := v(x,t)\tfrac{\partial}{\partial x}\big(\log\tfrac{p(x,t)}{q(x,t)}\big). \tag{2.18}$$

In addition, $\varepsilon^*(x,t)$, $\gamma^*(x,t)$ satisfy the condition (2.7).

The proof, which appears in subsection 5.5, has a straightforward idea, but requires lengthy calculations to find the stochastic differential of $W_t^* := \frac{p(S_t,t)}{q(S_t,t)}$ based on the Kunita–Itô-Wentzell formula ([13], Theorem 1.1), as the functions $p(x,t)$ and $q(x,t)$ satisfy SPDEs and their ratio is evaluated at S_t, which solves an SDE.

The presence of the noise term $\zeta^*(S_t,t)\,dB_t$ in equation (2.17) might appear counterintuitive in view of the goal to minimize the variance of W_t. Indeed, adding an independent BM term would only increase the variance. However, the BM appearing in (2.17) is the same as that driving the process S_t, introducing correlation between W_t^* and S_t, which consequently helps W_t^* to achieve the minimal variance possible given the dynamics of S_t. In particular, if we restrict the particle evolution to the prior dynamics (i.e. $u = f$, $k = 0$, $v = g$), then Proposition 6 results in

$$dW_t^* = W_t^*(h(S_t,t) - \hat{h}_t)\,(dY_t - \hat{h}_t dt) + W_t^*(\lambda^*(S_t,t)\,dt + \zeta^*(S_t,t)\,dB_t), \tag{2.19}$$

where ζ^* is given by (2.18) with $v = g$ and λ^* is the remaining drift coefficient

$$\lambda^*(x,t) := \tfrac{1}{p(x,t)} \tfrac{\partial}{\partial x}\big(g^2(x,t)p(x,t)\tfrac{\partial}{\partial x}(\log \tfrac{p(x,t)}{q(x,t)})\big). \tag{2.20}$$

Notice that the presence of an additional term in (2.19) compared to the weight process in the BPF (1.13) indicates that the importance sampling used in the BPF is not optimal. Nonetheless, this correction term is not easy to evaluate in practice, since the functions λ^*, ζ^*, which involve the densities p, q explicitly, must be estimated from samples. In the linear-Gaussian case, however, the analytical computation is possible:

EXAMPLE (Linear-Gaussian Case, Continued). *Let the particle process evolve as the prior dynamics and denote by $\mathcal{N}(\mu_t, \rho_t)$ its distribution. Then equations (2.19)-(2.20) can be computed analytically and yield the following PF:*

$$dS_t = aS_t\,dt + b\,dB_t, \tag{2.21}$$

$$\begin{aligned} dW_t^* = {} & W_t^*(cS_t - c\hat{\mu}_t)(dY_t - c\hat{\mu}_t dt) \\ & + W_t^* b^2\big(\tfrac{1}{\rho_t} - \tfrac{1}{\hat{\rho}_t} + (\tfrac{S_t-\hat{\mu}_t}{\hat{\rho}_t})^2 - \tfrac{(S_t-\mu_t)(S_t-\hat{\mu}_t)}{\rho_t\hat{\rho}_t}\big)\,dt \\ & + W_t^* b\big(\tfrac{S_t-\mu_t}{\rho_t} - \tfrac{S_t-\hat{\mu}_t}{\hat{\rho}_t}\big)\,dB_t. \end{aligned} \tag{2.22}$$

3. APPLICATIONS.

3.1. Compensating for Gain Approximation with Weight Dynamics. The gain function K in the FPF is a solution to equation (1.17), which is fixed, i.e. it only depends on the model. In the unifying theorem, however, the function k, which has a similar role as K, is a solution to the equation (2.5), which now has a term ε which can be chosen freely. Observe that $\frac{\partial}{\partial x}\big((k-K)p\big) = \varepsilon p$, which means that any deviation of k from K in the particle dynamics is compensated by ε from the weight dynamics such that the particle system can eventually target the filtering distribution. Recall from condition (2.7) that the mean of ε is zero under the filtering distribution. The variance of ε under the filtering distribution can serve as an "error", measuring the difference between k and K. Moreover, smaller values of ε are preferable because to keep the weights close to unity, the coefficients in the weight dynamics, including ε, should remain as close to zero as possible.

There are now two approaches to exploiting this freedom in (2.5): 1) to first set ε and then solve the equation for k, 2) to set k and find ε afterwards. In 1), the presence of ε allows us to modify the equation to some extent which might help us to use a simpler gain estimation method and compensate for it by an appropriate weight dynamics. It will be for future research to explore this possibility. In 2), which we explore in more detail, instead of setting k explicitly, we may restrict ourselves to a specific class of functions denoted by $\mathcal{K}$ and pose the following variational problem:

$$\begin{aligned} \underset{k}{\text{minimize}} \quad & \mathcal{I}[k] := \int_{\mathbb{R}} \varepsilon^2(x,t)p(x,t)dx, \\ \text{subject to} \quad & k \in \mathcal{K}, \quad \varepsilon p = (h - \hat{h}_t)p + \tfrac{\partial}{\partial x}(kp), \end{aligned} \tag{3.1}$$

where the second constraint comes from equation (2.5) and k must also satisfy the regularity assumptions required for Theorem 3. The objective functional $\mathcal{I}$ can be interpreted as the square of the Fisher-Rao norm $\|\dot{p}\|_{\mathrm{FR}}^2$ of the fictional change in the distribution corresponding to the continuity equation $\dot{p} + \frac{\partial}{\partial x}\big((k-K)p\big) = 0$ (here we omit the time dependence of k, K and the dot relates to a fictional time associated with the flow of the fixed vector field $k - K$). This contrasts with the use

of the infinitesimal Wasserstein-2 or Otto's norm $\|\dot{p}\|_{W_2}^2 = \int (k(x) - K(x))^2 p(x) dx$ (see [16]). As an example, we consider the case when $\mathcal{K}$ is the set of constant functions. In this case, the minimization of $\|\dot{p}\|_{W_2}^2$ over $\mathcal{K}$ gives rise to the standard constant gain approximation in the FPF ([26], Example 2 and Remark 5), and analogously, the problem (3.1) admits a simple analytical solution, which we present in the proposition below and we call Fisher-optimal constant gain for the aforementioned reasons.

PROPOSITION 7 (Fisher-Optimal Constant Gain Approximation). *Consider the equation (2.5) in Theorem 3 and let $\mathcal{K}$ be the set of functions independent of x with elements $\bar{k}(t)$, i.e. $\bar{k}(t) \in \mathcal{K} = \{k(x,t) : k'(x,t) = 0\}$. Then the solution $\bar{k}^*(t)$ to the variational problem (3.1) at time t is*

$$\bar{k}^*(t) = \frac{\mathbb{E}[h'(X_t,t)|\mathcal{F}_t^Y]}{\mathbb{E}[\psi^2(X_t,t)|\mathcal{F}_t^Y]}, \tag{3.2}$$

and the corresponding ε is $\varepsilon^*(x,t) := h(x,t) - \hat{h}_t + \bar{k}^*(t)\psi(x,t)$, where we have defined $\psi(x,t) := p'(x,t)/p(x,t)$.

The proof is given in subsection 5.6. Note that $\mathbb{E}[\psi^2(X_t,t)|\mathcal{F}_t^Y]$ is itself a Fisher information, namely that of the 1-parameter model $p_\theta(x,t) = p(x-\theta,t)$ at $\theta = 0$.

EXAMPLE (Linear-Gaussian Case, Continued). *It is easy to verify that in the linear-Gaussian case (1.5), $\bar{k}^*(t)$ from (3.2) yields the Kalman gain, which in turn makes $\varepsilon^*(x,t)$ and thus the objective functional zero. Observe that*

$$h'(x,t) = c \quad \Longrightarrow \quad \mathbb{E}[h'(X_t,t)|\mathcal{F}_t^Y] = c, \tag{3.3}$$

$$\psi(x,t) = -(x - \hat{\mu}_t)/\hat{\rho}_t \quad \Longrightarrow \quad \mathbb{E}[\psi^2(X_t,t)|\mathcal{F}_t^Y] = 1/\hat{\rho}_t, \tag{3.4}$$

which gives the Kalman gain $\bar{k}^*(t) = c\hat{\rho}_t$.

In general (i.e. non-linear or non-Gaussian case), $\psi(x,t)$ and the Fisher information $\mathbb{E}[\psi^2(X_t,t)|\mathcal{F}_t^Y]$ need to be estimated from samples. Estimators for these quantities can be found e.g. in [3, 4, 8, 20]. The proposition above can be generalized to richer classes of functions $\mathcal{K}$, e.g. functions of the form $k(x,t) = \sum_j a_j(t)\varphi_j(x,t)$, where $\{\varphi_j\}_j$ is an appropriate set of basis functions and $\{a_j\}_j$ are some coefficients.

3.2. Providing Freedom to Alleviate Weight Degeneracy. Weight decay is a major issue among the weighted PFs. As discussed in subsection 1.3, in order to have less weight degeneracy, $\mathbb{E}[W_t^2]$ should remain small. In subsection 2.1 and subsection 2.4, we derived the optimal importance weight W_t^* and its stochastic dynamics dW_t^* under the constraint that particle distribution is given i.e. particle dynamics is fixed. Here we would like to derive a general formula for $d\mathbb{E}[W_t^2]$ corresponding to the weight dynamics of form (2.4) in terms of $\{\gamma, \varepsilon, \zeta\}$. By Itô's formula, we have

$$dW_t^2 = 2W_t^2\big(\gamma(S_t,t)dt + \varepsilon(S_t,t)dY_t + \zeta(S_t,t)dB_t + \tfrac{1}{2}\varepsilon^2(S_t,t)dt + \tfrac{1}{2}\zeta^2(S_t,t)dt\big). \tag{3.5}$$

Applying Fubini's theorem to the integral form of SDE above would suffice for deriving $d\mathbb{E}[W_t^2]$. However, the presence of W_t^2, which multiplies functions of S_t, as well as dY_t, which in turn depends on the hidden process X_t, make the analysis of $d\mathbb{E}[W_t^2]$ complicated. It is interesting to note that studying $d\log(W_t)$ does not have the first issue because it no longer involves W_t explicitly. Observe that by Itô's formula

$$d\log(W_t) = \gamma(S_t,t)dt + \varepsilon(S_t,t)dY_t + \zeta(S_t,t)dB_t - \tfrac{1}{2}\varepsilon^2(S_t,t)dt - \tfrac{1}{2}\zeta^2(S_t,t)dt. \tag{3.6}$$

To overcome the second issue, it turns out that if we use the fact that the innovation process is a Brownian motion, we will get a more useful formula as follows:

PROPOSITION 8. *Consider the filtering problem (1.1)-(1.2). Let (S_t, W_t) be a pair process such that W_t solves the SDE (2.4) under the assumptions (ii), (iii), and (v) while S_t is a progressively measurable process with respect to $\mathcal{F}_t^B \vee \mathcal{F}_t^Y \vee \sigma(S_0)$. Suppose that $h(X_\cdot, \cdot) \in \mathbb{L}^1(0,t)$ and $\mathbb{E}[W_t^2], \mathbb{E}[|\log W_t|] < \infty$ for all $t \geq 0$. Then*

$$d\mathbb{E}[W_t^2] = 2\mathbb{E}\big[W_t^2\big(\gamma(S_t,t) + \varepsilon(S_t,t)\hat{h}_t + \tfrac{1}{2}\varepsilon^2(S_t,t) + \tfrac{1}{2}\zeta^2(S_t,t)\big)\big]\,dt, \tag{3.7}$$

$$d\mathbb{E}[\log(W_t)] = \mathbb{E}\big[\gamma(S_t,t) + \varepsilon(S_t,t)\hat{h}_t - \tfrac{1}{2}\varepsilon^2(S_t,t) - \tfrac{1}{2}\zeta^2(S_t,t)\big]\,dt. \tag{3.8}$$

The proof appears in subsection 5.7. Notice that due to the non-negativity of $\mathrm{Var}[W_t]$, we have $\mathbb{E}[W_t^2] \geq 1$ and due to $\log(x) \leq x-1$, we have $\mathbb{E}[\log(W_t)] \leq 0$. It can be shown (by counterexample) that $\mathbb{E}[\log(W_t)]$ is not necessarily a proper measure for degeneracy. However, the formulas above are insightful for investigating the possibility of exploiting the freedom given by the terms $\gamma, \varepsilon, \zeta$ in order to find other steady-state solutions for $\mathbb{E}[W_t^2]$ besides the minimum-variance solution in Proposition 6. We defer this investigation to future research but to illustrate the usefulness of the formulas above, we apply them to the weights in the BPF (1.13):

$$d\mathbb{E}[{W_t^{\mathrm{B}}}^2] = \mathbb{E}\big[{W_t^{\mathrm{B}}}^2(h(S_t^{\mathrm{B}},t) - \hat{h}_t)^2\big]\,dt, \tag{3.9}$$

$$d\mathbb{E}[\log(W_t^{\mathrm{B}})] = -\tfrac{1}{2}\mathbb{E}\big[(h(S_t^{\mathrm{B}},t) - \hat{h}_t)^2\big]\,dt. \tag{3.10}$$

Equation (3.9) implies that in the BPF, we always have $d\mathbb{E}[{W_t^{\mathrm{B}}}^2] \geq 0$ and hence the number of effective particles will inevitably decay over time. Note also that if the observation process is m-dimensional, the drift coefficient on the right-hand side of (3.9) is given by $\mathbb{E}[{W_t^{\mathrm{B}}}^2 \|h(S_t^{\mathrm{B}},t) - \hat{h}_t\|^2]$ and therefore the time constant of the weight decay will scale with $1/m$ (compare the analysis with [22], section 3.1.1).

4. DISCUSSION. Existing particle filters fall into two distinct types. Unweighted PFs (such as the FPF) assimilate new data by moving around particles while keeping the weights associated to each particle fixed whereas weighted PFs (such as the BPF) assimilate new data by reweighing particles. In this paper, we proposed a unifying framework for these types of PFs. Our proposed hybrid filter allows particles to be moved as well as reweighed in response to new observations. This gives a lot of freedom on how to design a PF. This freedom obviously needs to be constrained in order to make sure that the empirical distribution of weighted particles effectively converges to the filtering distribution (i.e. to be asymptotically exact). The necessary conditions are summarised by equations (2.5)-(2.6) in the unifying theorem.

Even after having constrained the freedom of the particle and weight dynamics to satisfy the targeting condition, there is still substantial freedom that could be exploited. An interesting extension of the present work will be to determine whether there exists a "sweet spot" where the strengths of unweighted PFs (i.e. the absence of the weight collapse problem) could be combined with the strengths of simple weighted PFs such as the BPF (where the solution of equations (2.5)-(2.6) is trivial). Practically, this would amount to defining a cost function which combines the cost associated to the severity of the weight decay as well as the cost for computing the solutions of equations (2.5)-(2.6). We leave this question for further work.

Another extension of the present work would be to relax the strong assumptions made in the SDEs (2.3)-(2.4). Indeed, in (2.3), it is assumed that the particle dynamics does not involve W_t explicitly and (2.4) assumes that the coefficients in the weight dynamics are linear in W_t. In this construction, S_t is the main process but W_t can be thought of as an auxiliary process driven by S_t (recall (3.6)) and transforming

the distribution of S_t into the filtering distribution. These assumptions were made to simplify the expression for the necessary conditions in the unifying theorem. However, if we relax them, we will obtain even more freedom on the choice of particle and weight dynamics that could be leveraged to increase the chance of obtaining a sweet spot as described in the previous paragraph. For example, we could enforce a drift term in the weight dynamics that pulls the weight back to unity and compensate for this specific choice by appropriate functions in the particle dynamics. This could be interpreted as a smooth resampling procedure unlike classical resampling where particles and weights have the undesirable feature of changing abruptly at the resampling times.

Finally, this paper is entirely focused on necessary conditions that the hybrid particle filter needs to satisfy in order to target the filtering distribution. An obvious next question will be to determine sufficient conditions that need to be satisfied by the model such that (2.5)-(2.6) guarantee that the targeting condition is met for an extended period of time (i.e. the converse of Theorem 3). The main difficulty will be to establish conditions under which the solutions of equations (2.5)-(2.6) make the SDEs (2.3)-(2.4) well-posed. This question has been partly addressed in [18] for some special cases but remains open in the general case considered here. Furthermore, additional work is required to establish sufficient conditions under which a PF with finite number N of particles has a uniformly bounded error. Indeed, at the end of subsection 2.3, we saw that in a simple linear-Gaussian case, the variance of the unweighted particles can grow exponentially if a specific parameter is above a given threshold (which is not recognized by the necessary conditions). This feature is clearly undesirable when the number of particles is finite since it implies that the number of samples that effectively support the filtering distribution decreases with time, which is similar to what occurs in the BPF. It would therefore be desirable to derive sufficient conditions that guarantee the stability of the filter.

5. PROOFS. This section presents all the proofs of our results. In a nutshell, the key results follow directly from the targeting condition (1.9) by evaluating different types of test functions ϕ, specifically, the indicator function $\phi(x) = \mathbb{1}_B(x)$, constant function $\phi(x) = 1$, and $\phi \in C_k^2$, which lead to (2.1), (2.7), and (2.5)-(2.6), respectively.

5.1. Proof of the Radon-Nikodym Characterization of the Weight.

Proof of Theorem 2.

I. By the targeting condition (1.9), we have for all integrable functions ϕ

$$\mathbb{E}[W_t\phi(S_t)|\mathcal{F}_t^Y] = \mathbb{E}\Big[\mathbb{E}[W_t\phi(S_t)|\mathcal{F}_t^Y, S_t]\Big|\mathcal{F}_t^Y\Big] \tag{5.1}$$

$$= \mathbb{E}\Big[\phi(S_t)\mathbb{E}[W_t|\mathcal{F}_t^Y, S_t]\Big|\mathcal{F}_t^Y\Big] \tag{5.2}$$

$$= \int_{\mathbb{R}} \phi(x)\mathbb{E}[W_t|\mathcal{F}_t^Y, S_t = x]Q_t(dx) \tag{5.3}$$

$$= \int_{\mathbb{R}} \phi(x)P_t(dx) = \mathbb{E}[\phi(X_t)|\mathcal{F}_t^Y]. \tag{5.4}$$

Let $B \in \mathcal{B}(\mathbb{R})$ be a Borel subset. Take $\phi(x) = \mathbb{1}_B(x)$ the indicator function of B, then

$$\int_B \mathbb{E}[W_t|\mathcal{F}_t^Y, S_t = x]Q_t(dx) = P_t(B), \tag{5.5}$$

which shows the first claim (2.1).

II. This follows directly from Lemma 9 just below. ∎

LEMMA 9. *Let $(\Omega, \mathcal{F}, P)$ be a probability space and W^* a $\mathcal{G}$-measurable random variable defined on this space. Suppose $\mathcal{G} \subseteq \mathcal{F}$ and $\mathbb{E}[W^{*2}] < \infty$. Then W^* is the (almost surely unique) solution to the optimization problem*

$$\begin{aligned} &\underset{W}{\text{minimize}} && \mathbb{E}[W^2], \\ &\text{subject to} && \mathbb{E}[W|\mathcal{G}] = W^*. \end{aligned} \tag{5.6}$$

Proof. $\mathbb{E}[W^2]$ can be written as

$$\mathbb{E}[W^2] = \mathbb{E}[(W - W^*)^2] - \mathbb{E}[W^{*2}] + 2\mathbb{E}[WW^*] \tag{5.7}$$

$$= \mathbb{E}[(W - W^*)^2] - \mathbb{E}[W^{*2}] + 2\mathbb{E}\big[\mathbb{E}[WW^*|\mathcal{G}]\big] \tag{5.8}$$

$$= \mathbb{E}[(W - W^*)^2] - \mathbb{E}[W^{*2}] + 2\mathbb{E}\big[W^*\mathbb{E}[W|\mathcal{G}]\big] \tag{5.9}$$

$$= \mathbb{E}[(W - W^*)^2] + \mathbb{E}[W^{*2}]. \tag{5.10}$$

As $\mathbb{E}[W^{*2}]$ is fixed, to minimize $\mathbb{E}[W^2]$, we have to minimize $\mathbb{E}[(W - W^*)^2]$. This is attained by any random variable W_{opt} such that $\mathbb{E}[(W_{\text{opt}} - W^*)^2] = 0$. Then Chebyshev's inequality implies that

$$\Pr(|W_{\text{opt}} - W^*| \geq \lambda) = 0 \quad \text{for all } \lambda > 0. \tag{5.11}$$

In other words, $W_{\text{opt}} = W^*$ almost surely. ∎

5.2. Proof of the Unifying Theorem. This subsection is devoted to the proof of Theorem 3. A key ingredient for the proof is the next lemma, whose first two assertions come from Lemma 2 in [27]. It allows us to not only interchange conditional expectations with integrals but also adapt the σ-algebra accordingly, which makes it distinct and stronger from the normal conditional Fubini theorem.

LEMMA 10. *Take the system of SDEs (2.3)-(2.4) for the pair process (S_t, W_t) under the assumptions (ii)-(v). Let $F(x, w, t)$ be an $\mathbb{R}$-valued measurable function such that $F(S_\cdot, W_\cdot, \cdot) \in \mathbb{L}^2(0, t)$. Then*

$$\mathbb{E}\Big[\int_0^t F(S_s, W_s, s)\, ds \Big| \mathcal{F}_t^Y\Big] = \int_0^t \mathbb{E}\big[F(S_s, W_s, s)\big|\mathcal{F}_s^Y\big]\, ds, \tag{5.12}$$

$$\mathbb{E}\Big[\int_0^t F(S_s, W_s, s)\, dY_s \Big| \mathcal{F}_t^Y\Big] = \int_0^t \mathbb{E}\big[F(S_s, W_s, s)\big|\mathcal{F}_s^Y\big]\, dY_s, \tag{5.13}$$

$$\mathbb{E}\Big[\int_0^t F(S_s, W_s, s)\, dB_s \Big| \mathcal{F}_t^Y\Big] = 0. \tag{5.14}$$

Proof. The process $F(S_t, W_t, t)$ is $\mathcal{F}_t^Y \vee \mathcal{F}_t^B \vee \sigma(S_0)$-measurable. Statements (5.12) and (5.13) then follow directly from Lemma 2 in [27]. The last claim (5.14) is similar to the zero-mean property of the Itô integral and has a similar proof, yet it involves conditional expectation. For the Itô integral, we have

$$\int_0^t F(S_s, W_s, s) dB_s = \lim_{n\to\infty} \sum_{i=0}^{n-1} F(S_{t_i}, W_{t_i}, t_i)(B_{t_{i+1}} - B_{t_i}) \tag{5.15}$$

in probability, where $\{t_i\}_{i=0}^n$ is a partition of $[0,t]$ with $\max_i(t_{i+1}-t_i) \to 0$ as $n \to \infty$. Consequently, by the dominated convergence theorem for conditional expectations

$$\mathbb{E}\Big[\int_0^t F(S_s, W_s, s) dB_s \Big| \mathcal{F}_t^Y\Big] = \lim_{n\to\infty} \mathbb{E}\Big[\sum_{i=0}^{n-1} F(S_{t_i}, W_{t_i}, t_i)(B_{t_{i+1}} - B_{t_i})\Big|\mathcal{F}_t^Y\Big] \tag{5.16}$$

$$= \lim_{n\to\infty} \sum_{i=0}^{n-1} \mathbb{E}\big[F(S_{t_i}, W_{t_i}, t_i)\big|\mathcal{F}_t^Y\big]\mathbb{E}\big[B_{t_{i+1}} - B_{t_i}\big] = 0, \tag{5.17}$$

where we used the fact that the Brownian motion increment $B_{t_{i+1}} - B_{t_i}$ is independent of $\mathcal{F}_t^Y$ as well as the random values S_{t_i} and W_{t_i} at time t_i and its mean is zero. ∎

Given the lemma above, we are now ready to prove Theorem 3.

Proof of Theorem 3. By the targeting assumption for all $0 < t < T$ as well as the initial condition, we know that

$$\mathbb{E}[W_t\phi(S_t)|\mathcal{F}_t^Y] = \mathbb{E}[\phi(X_t)|\mathcal{F}_t^Y] \qquad (0 \le t < T) \tag{5.18}$$

holds a.s. for any measurable function ϕ with $\mathbb{E}[|\phi(X_t)|] < \infty$, in particular for any integrable $\phi \in C^2$, which allows us to write an equality for the stochastic differential of these processes in Itô sense:

$$d\mathbb{E}[W_t\phi(S_t)|\mathcal{F}_t^Y] = d\mathbb{E}[\phi(X_t)|\mathcal{F}_t^Y] \qquad (0 \le t < T). \tag{5.19}$$

The right-hand side of expression above is given by equation (1.4), which consists of two terms multiplying dY_t and dt. The plan is to compute the left-hand side in terms of the unknown functions $\{u, k, v, \gamma, \varepsilon, \zeta\}$ from SDEs (2.3)-(2.4). It turns out that the left-hand side also consists of two terms multiplying dY_t and dt, since we take conditional expectation with respect to $\mathcal{F}_t^Y$ and thus terms multiplying dB_t vanish. Two fundamental ODEs (or PDEs in higher dimensions) finally follow from the fact that the terms multiplying dY_t and dt on each side of (5.19) are equal a.s. regardless of the test function ϕ. The proof is structured in three steps:

(a) We first find the stochastic differential $d(W_t\phi(S_t))$ from SDEs (2.3)-(2.4) using Itô's formula.
(b) In order to obtain $d\mathbb{E}[W_t\phi(S_t)|\mathcal{F}_t^Y]$, we write the result of the previous step in integral form, take the conditional expectation of both sides with respect to $\mathcal{F}_t^Y$, use Lemma 10 to interchange conditional expectations with integrals, and finally turn the result back into the differential form.
(c) We use the targeting assumption (5.18) to convert the conditional expectations that involve (W_t, S_t) into posterior expectations (this becomes possible because of the special form that we take for the stochastic dynamics of the particle system). Finally, we investigate the implications of the equalities resulting from (5.19) for some class of test functions ϕ.

Some aspects of our proof are inspired by the usual proof of the Fokker–Planck equation (e.g. see [23], proof of Theorem 5.4) and the derivation of the FPF [27, 28].

Step (a): Take any test function $\phi(x) \in C^2$. To obtain the stochastic differential of $W_t\phi(S_t)$ from the SDEs (2.3)-(2.4), we apply Itô's formula

$$d\big(W_t\phi(S_t)\big) = W_t\mathcal{A}_t\phi(S_t)\,dt + W_t\mathcal{B}_t\phi(S_t)\,dY_t + W_t\mathcal{C}_t\phi(S_t)\,dB_t, \tag{5.20}$$

where the linear operators $\{\mathcal{A}_t, \mathcal{B}_t, \mathcal{C}_t\}$ have been defined as

$$\mathcal{A}_t\phi(x) := \gamma(x,t)\phi(x) + \big[u(x,t) + k(x,t)\varepsilon(x,t) + v(x,t)\zeta(x,t)\big]\phi'(x) + \tfrac{1}{2}\big[k(x,t)^2 + v(x,t)^2\big]\phi''(x), \tag{5.21}$$

$$\mathcal{B}_t\phi(x) := \varepsilon(x,t)\phi(x) + k(x,t)\phi'(x), \tag{5.22}$$

$$\mathcal{C}_t\phi(x) := \zeta(x,t)\phi(x) + v(x,t)\phi'(x). \tag{5.23}$$

Step (b): Writing the SDE (5.20) in integral form and taking the conditional expectation of both sides gives the following:

$$\mathbb{E}\big[W_t\phi(S_t)\big|\mathcal{F}_t^Y\big] = \mathbb{E}\big[W_0\phi(S_0)\big|\mathcal{F}_t^Y\big] + \mathbb{E}\Big[\int_0^t W_s\mathcal{A}_s\phi(S_s)ds\Big|\mathcal{F}_t^Y\Big] + \mathbb{E}\Big[\int_0^t W_s\mathcal{B}_s\phi(S_s)dY_s\Big|\mathcal{F}_t^Y\Big] + \mathbb{E}\Big[\int_0^t W_s\mathcal{C}_s\phi(S_s)dB_s\Big|\mathcal{F}_t^Y\Big]. \tag{5.24}$$

Now apply Lemma 10 to interchange conditional expectations with integrals in the expression above, provided that $W_\cdot\mathcal{A}_\cdot\phi(S_\cdot), W_\cdot\mathcal{B}_\cdot\phi(S_\cdot), W_\cdot\mathcal{C}_\cdot\phi(S_\cdot) \in \mathbb{L}^2(0,t)$. Moreover, the last term is zero by the third claim of the aforementioned lemma. Then

$$\mathbb{E}\big[W_t\phi(S_t)\big|\mathcal{F}_t^Y\big] = \mathbb{E}\big[W_0\phi(S_0)\big|\mathcal{F}_0^Y\big] + \int_0^t \mathbb{E}\big[W_s\mathcal{A}_s\phi(S_s)\big|\mathcal{F}_s^Y\big]ds + \int_0^t \mathbb{E}\big[W_s\mathcal{B}_s\phi(S_s)\big|\mathcal{F}_s^Y\big]dY_s, \tag{5.25}$$

which can be put in differential form as

$$d\mathbb{E}\big[W_t\phi(S_t)\big|\mathcal{F}_t^Y\big] = \mathbb{E}\big[W_t\mathcal{A}_t\phi(S_t)\big|\mathcal{F}_t^Y\big]dt + \mathbb{E}\big[W_t\mathcal{B}_t\phi(S_t)\big|\mathcal{F}_t^Y\big]dY_t. \tag{5.26}$$

Step (c): It becomes clear now why our ansatz (2.3)-(2.4) for the dynamics of the particle system is advantageous. As the particle dynamics does not involve W_t explicitly and the coefficients in the weight dynamics are linear in W_t, the arguments of the conditional expectations in (5.26) become linear in W_t. This allows us to use the targeting assumption (5.18), which holds for any (integrable) measurable function, in order to convert conditional expectations involving (S_t, W_t) to posterior expectations:

$$d\mathbb{E}[W_t\phi(S_t)|\mathcal{F}_t^Y] = \mathbb{E}[\mathcal{A}_t\phi(X_t)|\mathcal{F}_t^Y]dt + \mathbb{E}[\mathcal{B}_t\phi(X_t)|\mathcal{F}_t^Y]dY_t. \tag{5.27}$$

Now we have an expression for the left-hand side of (5.19) in terms of posterior expectations and we recall that the right-hand side is given by equation (1.4). As terms multiplying dt and dY_t on both sides match, we have

$$\mathbb{E}[\mathcal{A}_t\phi(X_t)|\mathcal{F}_t^Y] = \mathbb{E}[\mathcal{L}_t\phi(X_t)|\mathcal{F}_t^Y] - \mathbb{E}[\phi(X_t)(h(X_t,t) - \hat{h}_t)\hat{h}_t|\mathcal{F}_t^Y], \tag{5.28}$$

$$\mathbb{E}[\mathcal{B}_t\phi(X_t)|\mathcal{F}_t^Y] = \mathbb{E}[\phi(X_t)(h(X_t,t) - \hat{h}_t)|\mathcal{F}_t^Y], \tag{5.29}$$

almost surely. Equalities above hold for any $\phi \in C^2$. In particular, $\phi = 1$ implies

$$\mathbb{E}[\gamma(X_t,t)|\mathcal{F}_t^Y] = 0, \quad \mathbb{E}[\varepsilon(X_t,t)|\mathcal{F}_t^Y] = 0, \tag{5.30}$$

almost surely, which proves the last claim in the theorem.

Additionally, we can use integration-by-parts in (5.28)-(5.29) to rewrite them only in terms of ϕ (but not its derivatives). If we further restrict ourselves to compactly

supported test functions $\phi \in C_k^2$, boundary terms vanish and we formally have

$$\int_{\mathbb{R}} \phi(x)\Big[\mathcal{A}_t^\dagger p(x,t) - \mathcal{L}_t^\dagger p(x,t) + (h(x,t) - \hat{h}_t)\hat{h}_t p(x,t)\Big]\, dx = 0, \tag{5.31}$$

$$\int_{\mathbb{R}} \phi(x)\Big[\mathcal{B}_t^\dagger p(x,t) - (h(x,t) - \hat{h}_t) p(x,t)\Big]\, dx = 0, \tag{5.32}$$

where $\mathcal{A}_t^\dagger$, $\mathcal{B}_t^\dagger$, and $\mathcal{L}_t^\dagger$ are the formal adjoints of $\mathcal{A}_t$, $\mathcal{B}_t$, and $\mathcal{L}_t$, respectively, with respect to Lebesgue measure. Since these equations hold for all $\phi \in C_k^2$ and the terms inside the square brackets are continuous in x (given the assumptions (iv) and (v)), the fundamental lemma in the calculus of variations (see e.g. Lemma 1 in [9]) implies that the terms inside the square brackets are identically zero on $\mathbb{R}$, which gives two main equations of the theorem as we will see soon. The remainder of the proof will be concerned with calculating the adjoint operators $\mathcal{A}_t^\dagger$ and $\mathcal{B}_t^\dagger$.

The integrals we deal with have the general form of $\mathbb{E}[l(X_t,t)\phi'(X_t)|\mathcal{F}_t^Y]$ or $\mathbb{E}[l(X_t,t)\phi''(X_t)|\mathcal{F}_t^Y]$, where $l(x,t)$ is a $C^{1,0}$ or $C^{2,0}$ function, respectively. We have

$$\mathbb{E}[l(X_t,t)\phi'(X_t)|\mathcal{F}_t^Y] = \int_{\mathbb{R}} \phi' lp\, dx = \phi lp\Big|_{-\infty}^{\infty} - \int_{\mathbb{R}} \phi (lp)'\, dx = -\int_{\mathbb{R}} \phi (lp)'\, dx, \tag{5.33}$$

in which the boundary term is zero because ϕ has compact support. Similarly, we can use the integration-by-parts twice and write

$$\mathbb{E}[l(X_t,t)\phi''(X_t)|\mathcal{F}_t^Y] = \int_{\mathbb{R}} \phi'' lp\, dx = -\int_{\mathbb{R}} \phi' (lp)'\, dx = \int_{\mathbb{R}} \phi (lp)''\, dx. \tag{5.34}$$

Using this technique, we have

$$\mathbb{E}[\mathcal{A}_t\phi(X_t)|\mathcal{F}_t^Y] = \int_{\mathbb{R}} \phi\Big[\underbrace{\varepsilon p - (kp)'}_{\mathcal{A}_t^\dagger p(x,t)}\Big]\, dx, \tag{5.35}$$

$$\mathbb{E}[\mathcal{B}_t\phi(X_t)|\mathcal{F}_t^Y] = \int_{\mathbb{R}} \phi\Big[\underbrace{\gamma p - (up)' - (k\varepsilon p)' - (v\zeta p)' + \frac{1}{2}(k^2 p + v^2 p)''}_{\mathcal{B}_t^\dagger p(x,t)}\Big]\, dx. \tag{5.36}$$

Plugging $\mathcal{A}_t^\dagger p(x,t)$ and $\mathcal{B}_t^\dagger p(x,t)$ into (5.31)-(5.32) completes the proof. ∎

5.3. Derivation of the Class of Particle Filters.

Proof of Proposition 5. Equation (2.6) is a second-order ODE, consisting of three parts. A trivial solution to this equation can be obtained by setting each part to zero. Alternatively, one can use (2.5) to first change the form of (2.6) and then set each part to zero. Specifically, the second-order term $\frac{\partial^2}{\partial x^2}(k^2 p)$ in equation (2.6) can be converted from second-order derivative into first-order derivative using (2.5) as follows:

$$\tfrac{\partial^2}{\partial x^2}(k^2 p) = \tfrac{\partial}{\partial x}\big(kk'p + k(kp)'\big) = \tfrac{\partial}{\partial x}\big(k(k' - h + \hat{h}_t + \varepsilon)p\big). \tag{5.37}$$

Likewise, the zero-order term $(h - \hat{h}_t)\hat{h}_t p$ in equation (2.6) can yield a first-order derivative using (2.5):

$$(h - \hat{h}_t)\hat{h}_t p = -\tfrac{\partial}{\partial x}(k\hat{h}_t p) + \varepsilon \hat{h}_t p. \tag{5.38}$$

To obtain a general expression, we may simply split

$$\frac{\partial^2}{\partial x^2}(k^2 p) = \beta \frac{\partial^2}{\partial x^2}(k^2 p) + (1-\beta)\frac{\partial^2}{\partial x^2}(k^2 p), \tag{5.39}$$

$$(h-\hat{h}_t)\hat{h}_t p = \alpha(h-\hat{h}_t)\hat{h}_t p + (1-\alpha)(h-\hat{h}_t)\hat{h}_t p, \tag{5.40}$$

where $\alpha, \beta \in \mathbb{R}$ are free parameters, which can be (continuous) functions of t, but not of x, because we would like to have the possibility to take them inside the derivatives in next calculation steps. In equations above, we let the first parts remain unchanged, while for the second parts, we use (5.37) and (5.38). Equation (2.6) then becomes

$$\begin{aligned} &\frac{1}{2}\frac{\partial^2}{\partial x^2}\big((\beta k^2 + v^2 - g^2)p\big) \\ &- \frac{\partial}{\partial x}\Big(\big(u - f + \frac{1}{2}k\big((1-\beta)h + (1+\beta-2\alpha)\hat{h}_t + (1+\beta)\varepsilon - (1-\beta)k'\big) + v\zeta\big)p\Big) \\ &\qquad + \big(\alpha(h-\hat{h}_t)\hat{h}_t + (1-\alpha)\varepsilon\hat{h}_t + \gamma\big)p = 0. \end{aligned} \tag{5.41}$$

In the equation above, by separately setting each of the three terms to zero, we obtain a particular solution. Furthermore, we wish to derive the results for a ε that consists of an interpolation. As we discussed in subsection 2.3, the choice of ε that linearly interpolates between $h - \hat{h}_t$ and 0 is a relevant choice for our purpose. We also let ε consist of a free function (to not restrict ourselves to just an interpolation). So we take $\varepsilon = \eta(h - \hat{h}_t) + \tilde{\varepsilon}$, where η is a free parameter like α, β, and $\tilde{\varepsilon}$ is an arbitrary function with zero mean under the posterior distribution (to guarantee the condition (2.7)). To sum up, plugging in ε to the modified ODE (5.41) and setting each term to zero yield the results of this proposition. ∎

5.4. Derivation of the Particle Distribution. Here we first present a lemma that describes the evolution equation of $q(x,t)$ for a particle dynamics of the general form (2.3). We then apply the result to the linear-Gaussian case in particular. This lemma will also be used later for the proof of Proposition 6.

LEMMA 11 (Proposition 1 in [26]). *Consider the stochastic process S_t satisfying the SDE (2.3) under the assumptions (iii)-(iv). The probability density function of S_t given $\mathcal{F}_t^Y$, denoted by $q(x,t)$, evolves as the following SPDE*

$$dq(x,t) = \mathcal{J}_t^\dagger q(x,t)\,dt + \mathcal{K}_t^\dagger q(x,t)\,dY_t, \tag{5.42}$$

where $\mathcal{J}_t^\dagger \cdot := -\frac{\partial}{\partial x}(u(x,t)\cdot) + \frac{1}{2}\frac{\partial^2}{\partial x^2}(k^2(x,t)\cdot + v^2(x,t)\cdot)$ and $\mathcal{K}_t^\dagger \cdot := -\frac{\partial}{\partial x}(k(x,t)\cdot)$.

Proof of Statement (2.15)-(2.16). After setting $\tilde{\varepsilon} = 0$ and $\zeta = 0$ in Proposition 5 for the linear-Gaussian setting, the coefficients of the particle dynamics are given by

$$k(t) = (1-\eta)c\hat{\rho}_t, \tag{5.43}$$

$$v^2(t) = b^2 - \beta(1-\eta)^2 c^2 \hat{\rho}_t^2, \tag{5.44}$$

$$u(x,t) = ax - \tfrac{1}{2}(1-\eta)c^2\hat{\rho}_t(\vartheta_1 x + \vartheta_2 \hat{\mu}_t), \tag{5.45}$$

where (5.43) is obtained by solving the equation (2.5). Due to the linearity and initial Gaussian distribution, the particle distribution will stay Gaussian, $S_t|\mathcal{F}_t^Y \sim \mathcal{N}(\mu_t, \rho_t)$. Hence, it only remains to find the evolution of the mean $\mu_t = \mathbb{E}[S_t|\mathcal{F}_t^Y]$ and variance $\rho_t = \mathrm{Var}[S_t|\mathcal{F}_t^Y]$. Using the integration-by-parts, we can derive from (5.42) that

$$d\mathbb{E}[S_t|\mathcal{F}_t^Y] = \mathbb{E}[u(S_t,t)|\mathcal{F}_t^Y]\,dt + \mathbb{E}[k(S_t,t)|\mathcal{F}_t^Y]\,dY_t, \tag{5.46}$$

$$\begin{aligned} d\mathrm{Var}[S_t|\mathcal{F}_t^Y] =& \mathbb{E}[(S_t - \mathbb{E}[S_t|\mathcal{F}_t^Y])u(S_t,t) + v^2(S_t,t)|\mathcal{F}_t^Y]\,dt \\ &+ \mathrm{Var}[k(S_t,t)|\mathcal{F}_t^Y]\,dt + \mathbb{E}[(S_t - \mathbb{E}[S_t|\mathcal{F}_t^Y])k(S_t,t)|\mathcal{F}_t^Y]\,dY_t. \end{aligned} \tag{5.47}$$

Substituting (5.43)-(5.45) into (5.46)-(5.47) yields the result. ∎

5.5. Derivation of the Optimal Weight Dynamics.

Proof of Proposition 6. The goal is to obtain the stochastic differential $d(\frac{p(S_t,t)}{q(S_t,t)})$, for which we now know that $p(x,t)$ and $q(x,t)$ satisfy the SPDEs (1.3) and (5.42), respectively, and S_t solves the SDE (2.3). As a first step, we calculate by Itô's lemma

$$d\big(\tfrac{p(x,t)}{q(x,t)}\big) = \Lambda_1(x,t)\,dt + \Lambda_2(x,t)\,dY_t, \tag{5.48}$$

where the functions $\{\Lambda_1(x,t), \Lambda_2(x,t)\}$ are given by

$$\Lambda_1 := \tfrac{p}{q}\big(\tfrac{1}{p}\mathcal{L}_t^\dagger p - (h-\hat{h}_t)\hat{h}_t - \tfrac{1}{q}\mathcal{J}_t^\dagger q - \tfrac{1}{q}(h-\hat{h}_t)\mathcal{K}_t^\dagger q + \tfrac{1}{q^2}(\mathcal{K}_t^\dagger q)^2\big), \tag{5.49}$$

$$\Lambda_2 := \tfrac{p}{q}\big((h-\hat{h}_t) - \tfrac{1}{q}\mathcal{K}_t^\dagger q\big). \tag{5.50}$$

Next, to calculate $d(\frac{p(S_t,t)}{q(S_t,t)})$, the Kunita-Itô-Wentzell formula can be used (see [13], Theorem 1.1, and set $F_t(x) = \frac{p(x,t)}{q(x,t)}$, $M_t = S_t$ as continuous semimartingales). We get

$$d\big(\tfrac{p(S_t,t)}{q(S_t,t)}\big) = \Upsilon_1(S_t,t)\,dt + \Upsilon_2(S_t,t)\,dY_t + \Upsilon_3(S_t,t)\,dB_t, \tag{5.51}$$

where the functions $\{\Upsilon_1(x,t), \Upsilon_2(x,t), \Upsilon_3(x,t)\}$ are given by

$$\Upsilon_1 := \Lambda_1 + \mathcal{J}_t\tfrac{p}{q} + \mathcal{K}_t\Lambda_2, \quad \Upsilon_2 := \Lambda_2 + \mathcal{K}_t\tfrac{p}{q}, \quad \Upsilon_3 := \mathcal{V}_t\tfrac{p}{q}, \tag{5.52}$$

with operators $\mathcal{J}_t\cdot := u(x,t)\frac{\partial}{\partial x}\cdot + \frac{1}{2}(k^2(x,t)+v^2(x,t))\frac{\partial^2}{\partial x^2}\cdot$ and $\mathcal{K}_t\cdot := k(x,t)\frac{\partial}{\partial x}\cdot$ whose adjoints have been already introduced in Lemma 11. We have also defined the operator $\mathcal{V}_t\cdot := v(x,t)\frac{\partial}{\partial x}\cdot$. It is easy to see that $\Upsilon_3 = \frac{p}{q}(v(\log\frac{p}{q})') =: \frac{p}{q}\zeta^*$, which proves (2.18). Further, it can be verified that Υ_2 and Υ_1 correspond to $\frac{p}{q}\varepsilon^*$ and $\frac{p}{q}\gamma^*$, respectively, where ε^* and γ^* are the (unique) solutions to equations (2.5)-(2.6) after setting $\zeta = \zeta^*$. The calculation steps are, however, omitted on account of space. Finally, to show that ε^*, γ^* also satisfy (2.7), it is easy to see that the optimal weight satisfies $\mathbb{E}[W_t^*|\mathcal{F}_t^Y] = 1$ a.s., which leads to the same result as shown in (5.30). ∎

5.6. Derivation of the Fisher-Optimal Constant Gain Approximation.

Proof of Proposition 7. The objective functional $\mathcal{I}[\bar{k}]$ can be written as

$$\begin{aligned}\mathbb{E}[\varepsilon^2(X_t,t)|\mathcal{F}_t^Y] = \mathbb{E}[(h(X_t,t)-\hat{h}_t)^2|\mathcal{F}_t^Y] \\ + \bar{k}(t)^2\mathbb{E}[\psi^2(X_t,t)|\mathcal{F}_t^Y] - 2\bar{k}(t)\mathbb{E}[h'(X_t,t)|\mathcal{F}_t^Y],\end{aligned} \tag{5.53}$$

where we used integration by parts for the last term and the boundedness of $h(x,t)$. For a fixed t, we take the derivative of the objective functional with respect to $\bar{k}(t)$

$$\frac{d}{d\bar{k}}\mathbb{E}[\varepsilon^2(X_t,t)|\mathcal{F}_t^Y] = 2\bar{k}(t)\mathbb{E}[\psi^2(X_t,t)|\mathcal{F}_t^Y] - 2\mathbb{E}[h'(X_t,t)|\mathcal{F}_t^Y]. \tag{5.54}$$

Furthermore, observe that the second derivative is always non-negative:

$$\frac{d^2}{d\bar{k}^2}\mathbb{E}[\varepsilon^2(X_t,t)|\mathcal{F}_t^Y] = 2\mathbb{E}[\psi^2(X_t,t)|\mathcal{F}_t^Y] \geq 0. \tag{5.55}$$

Thus, the minimizer $\bar{k}^*(t)$ is found by setting the first derivative (5.54) to zero. ∎

5.7. Derivation of Weight Degeneracy Dynamics.

LEMMA 12 (Proposition 2.30. in [2]). *Consider the filtering problem (1.1)-(1.2). The innovation process I_t, which evolves according to the SDE $dI_t = dY_t - \hat{h}_t dt$, is a $\mathcal{F}_t^Y$-adapted Brownian motion if $h(X_\cdot, \cdot) \in \mathbb{L}^1(0,t)$ holds.*

Proof of Proposition 8. Plug $dY_t = dI_t + \hat{h}_t dt$ into SDEs (3.5)-(3.6), write them in integral form, and take the expectation of both sides. In addition to B_t, the innovation process I_t is a Brownian motion by Lemma 12. Hence, terms that involve B_t and I_t vanish due to the fact that the expectation of Itô Integral is zero. Finally, applying Fubini's theorem (see e.g. [12], Theorem 2.39) for the remaining terms and turning the result back into the differential form give the results. ∎

REFERENCES

[1] E. ABEDI AND S. C. SURACE, *Gauge Freedom within the Class of Linear Feedback Particle Filters*, in 2019 IEEE 58th Conference on Decision and Control (CDC), 2019, pp. 666–671.

[2] A. BAIN AND D. CRISAN, *Fundamentals of Stochastic Filtering*, Springer-Verlag New York, 2009, https://doi.org/10.1007/978-0-387-76896-0.

[3] P. K. BHATTACHARYA, *Estimation of a probability density function and its derivatives*, Sankhyā: The Indian Journal of Statistics, Series A, 29 (1967), pp. 373–382.

[4] W. CAO, A. DYTSO, M. FAUSS, H. V. POOR, AND G. FENG, *Nonparametric Estimation of the Fisher Information and Its Applications*, 2020, https://arxiv.org/abs/2005.03622.

[5] N. CHOPIN, *Central limit theorem for sequential Monte Carlo methods and its application to Bayesian inference*, Ann. Statist., 32 (2004), pp. 2385–2411, https://doi.org/10.1214/009053604000000698.

[6] A. DOUCET, S. GODSILL, AND C. ANDRIEU, *On Sequential Monte Carlo Sampling Methods for Bayesian Filtering*, Statistics and Computing, 10 (2000), pp. 197–208, https://doi.org/10.1023/A:1008935410038.

[7] A. DOUCET AND A. JOHANSEN, *A Tutorial on Particle Filtering and Smoothing: Fifteen Years Later*, Handbook of Nonlinear Filtering, 12 (2009).

[8] V. FABIAN, *Estimation of the Derivative of the Logarithm of a Density*, The Annals of Statistics, 1 (1973), pp. 557–561, https://doi.org/10.1214/aos/1176342424.

[9] I. M. GELFAND AND S. V. FOMIN, *Calculus of Variations*, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1963. Translated and Edited by R. A. Silverman.

[10] N. J. GORDON, D. J. SALMOND, AND A. F. M. SMITH, *Novel approach to nonlinear/non-Gaussian Bayesian state estimation*, IEE Proceedings F - Radar and Signal Processing, 140 (1993), pp. 107–113, https://doi.org/10.1049/ip-f-2.1993.0015.

[11] R. E. KALMAN AND R. S. BUCY, *New Results in Linear Filtering and Prediction Theory*, Journal of Basic Engineering, 83 (1961), pp. 95–108, https://doi.org/10.1115/1.3658902.

[12] F. C. KLEBANER, *Introduction to Stochastic Calculus With Applications*, Imperial College Press, second ed., 2005, https://doi.org/10.1142/P386.

[13] H. KUNITA, *Some extensions of Ito's formula*, Séminaire de probabilités de Strasbourg, 15 (1981), pp. 118–141, http://www.numdam.org/item/SPS_1981__15__118_0.

[14] H. J. KUSHNER, *On the Differential Equations Satisfied by Conditional Probablitity Densities of Markov Processes, with Applications*, Journal of the Society for Industrial and Applied Mathematics Series A Control, 2 (1964), pp. 106–119, https://doi.org/10.1137/0302009.

[15] A. KUTSCHIREITER, S. C. SURACE, AND J.-P. PFISTER, *The Hitchhiker's guide to nonlinear filtering*, Journal of Mathematical Psychology, 94 (2020), p. 102307, https://doi.org/10.1016/j.jmp.2019.102307.

[16] J. LOTT, *Some geometric calculations on Wasserstein space*, Communications in Mathematical Physics, 277 (2008), pp. 423–437, https://doi.org/10.1007/s00220-007-0367-3.

[17] L. MARTINO, V. ELVIRA, AND F. LOUZADA, *Effective sample size for importance sampling based on discrepancy measures*, Signal Processing, 131 (2017), pp. 386 – 401, https://doi.org/10.1016/j.sigpro.2016.08.025.

[18] S. PATHIRAJA, S. REICH, AND W. STANNAT, *McKean-Vlasov SDEs in nonlinear filtering*, 2020, https://arxiv.org/abs/2007.12658.

[19] S. REICH, *Data assimilation: The Schrödinger perspective*, Acta Numerica, 28 (2019), pp. 635–711, https://doi.org/10.1017/S0962492919000011.

[20] E. F. SCHUSTER, *Estimation of a Probability Density Function and Its Derivatives*, The An-

nals of Mathematical Statistics, 40 (1969), pp. 1187–1195, https://doi.org/10.1214/aoms/1177697495.

[21] R. L. Stratonovich, *Conditional Markov Processes*, Theory of Probability & Its Applications, 5 (1960), pp. 156–178, https://doi.org/10.1137/1105015.

[22] S. C. Surace, A. Kutschireiter, and J.-P. Pfister, *How to Avoid the Curse of Dimensionality: Scalability of Particle Filters with and without Importance Weights*, SIAM Review, 61 (2019), pp. 79–91, https://doi.org/10.1137/17M1125340.

[23] S. Särkkä and A. Solin, *Applied Stochastic Differential Equations*, Institute of Mathematical Statistics Textbooks, Cambridge University Press, 2019, https://doi.org/10.1017/9781108186735.006.

[24] A. Taghvaei and P. G. Mehta, *An optimal transport formulation of the linear feedback particle filter*, in 2016 American Control Conference (ACC), July 2016, pp. 3614–3619, https://doi.org/10.1109/ACC.2016.7525474.

[25] A. Taghvaei and P. G. Mehta, *An Optimal Transport Formulation of the Ensemble Kalman Filter*, IEEE Transactions on Automatic Control, (2020), pp. 1–1, https://doi.org/10.1109/TAC.2020.3015410.

[26] T. Yang, R. S. Laugesen, P. G. Mehta, and S. P. Meyn, *Multivariable feedback particle filter*, Automatica, 71 (2016), pp. 10 – 23, https://doi.org/10.1016/j.automatica.2016.04.019.

[27] T. Yang, P. G. Mehta, and S. P. Meyn, *Feedback particle filter with mean-field coupling*, in 2011 50th IEEE Conference on Decision and Control and European Control Conference, 2011, pp. 7909–7916.

[28] T. Yang, P. G. Mehta, and S. P. Meyn, *Feedback Particle Filter*, IEEE Transactions on Automatic Control, 58 (2013), pp. 2465–2480.