arXiv:1611.08761v1 [math.PR] 26 Nov 2016

Ergodicity and Accuracy of Optimal Particle Filters for Bayesian Data Assimilation

November 29, 2016

D. Kelly[⋆] **and A.M. Stuart**[†]

[⋆] New York University, Email: `dtbkelly@gmail.com`
[†] California Institute of Technology, Email: `astuart@caltech.edu`

Abstract

Data assimilation refers to the methodology of combining dynamical models and observed data with the objective of improving state estimation. Most data assimilation algorithms are viewed as approximations of the Bayesian posterior (filtering distribution) on the signal given the observations. Some of these approximations are controlled, such as particle filters which may be refined to produce the true filtering distribution in the large particle number limit, and some are uncontrolled, such as ensemble Kalman filter methods which do not recover the true filtering distribution in the large ensemble limit. Other data assimilation algorithms, such as cycled 3DVAR methods, may be thought of as approximating the mean of the posterior, but are also uncontrolled in general. For particle filters and ensemble Kalman filters it is of practical importance to understand how and why data assimilation methods can be effective when used with a fixed small number of particles, since for many large-scale applications it is not practical to deploy algorithms close to the large particle limit asymptotic. In this paper we address this question for particle filters and, in particular, study their accuracy (in the small noise limit) and ergodicity (for noisy signal and observation) without appealing to the large particle number limit. We first prove the accuracy and ergodicity properties for the true filtering distribution, working in the setting of conditional Gaussianity for the dynamics-observation model. We then show that these properties are inherited by optimal particle filters for any fixed number of particles. For completeness we also prove large particle number consistency results for the optimal particle filters, by writing the update equations for the underlying distributions as recursions. In addition to looking at the optimal particle filter with standard resampling, we derive all the above results for the Gaussianized optimal particle filter and show that the theoretical properties are favorable when compared to the standard optimal particle filter.

1 Introduction

1.1 Background and Literature Review

Data assimilation describes the blending of dynamical models with data, with the objective of improving state estimation and forecasts. The use of data assimilation originated in the geophysical sciences, but is now ubiquitous in engineering and the applied sciences. In numerical weather prediction, large scale ocean-atmosphere models are assimilated with massive data sets, comprising observational data from satellites, ground based weather stations and underwater sensors for example [3]. Data assimilation is prevalent in robotics; the SLAM problem seeks to use sensory data made by robots in an unknown environment to create a map of that environment and locate the robot within it [41]. It is used in modelling of traffic flow [48]. And data assimilation is being used in bio-medical applications such as glucose-insulin systems [37] and the sleep cycle [38]. These examples serve to illustrate the growth in the use of the methodology, its breadth of applicability and the very different levels of fidelity present in the models and the data in these many applications.

Although typical data assimilation problems can be understood from a Bayesian perspective, for non-linear and potentially high dimensional models it is often infeasible to make useful computations with the posterior. To

circumvent this problem, practitioners have developed assimilation methods that approximate the true posterior, but for which computations are more feasible. In the engineering communities, particle filters have been developed for this purpose, providing empirical approximations of non-Gaussian posteriors [11, 10]. In geoscience communities, methods are typically built on Kalman filtering theory, after making suitable Gaussian approximations [25]. These include variational methods like 3DVAR and 4DVAR [14, 28], the extended Kalman filter (ExKF) [12] and the ensemble Kalman filter (EnKF) [5, 13] which relies on a Gaussian ansatz rendering it, in general, invalid as an approximation of the true filtering distribution [26].

Despite their enormous success, many of these algorithms remain mysterious from a theoretical perspective. At the heart of the mystery is the fact that data assimilation methods are frequently and successfully implemented in regimes where the approximate filter is not provably valid; it is not known which features of the posterior (the true filtering distribution) are reflected in the approximation and which are not. For example, the ensemble Kalman filter is often implemented with ensemble size several orders of magnitude smaller than needed to reproduce large sample size behaviour, and is applied to problems for which the Gaussian ansatz may not be valid; it nonetheless can still exhibit skillful state estimates, with high correlations between the estimate and true trajectories [16, 29]. Indeed, the success of the methods in this non-asymptotic regime is the crux of their success; the methods would often be computationally intractable at large ensemble sizes.

This lack of theory has motivated recent efforts to better understand the properties of data assimilation methods in the practical, non-asymptotic regimes. The 3DVAR algorithm has been investigated in the context of toy models for numerical weather prediction, including the Lorenz-63 [24], Lorenz-96 [23] and $2d$ Navier-Stokes equations [22]; see also [33]. These works focus primarily on the question of accuracy - how well does the state estimate track the true underlying signal. Accuracy for the EnKF with fixed ensemble size was first investigated in [19]. Accuracy for EnKF has been further developed in [30] using linear models with random coefficients, but much more realistic (practical) assumptions on observations than [19], and moreover focussing on covariance consistency through the Mahalanobis norm. The articles [42, 43] were the first to investigate the stability of EnKF with fixed ensemble size, by formulating the filter as a Markov chain and applying coupling techniques. This line of research has been continued in [8] by framing the EnKF as a McKean-Vlasov system. The limitations of the non-practical regimes have also been investigated; in [17] the authors construct simple dissipative dynamical models for which the EnKF is shown to be highly unstable with respect to initial perturbations. This was the first theoretical insight into the frequently observed effect of catastrophic filter divergence [15].

For the nonlinear filtering distribution itself, there has been a great deal of research over the last several decades, particularly on the question of stability. Conditional ergodicity for the filtering distribution for general nonlinear hidden Markov models has been investigated in [21] and later refined in [45]. Ergodicity for nonlinear filters has been discussed in [20, 9, 7] and exponential convergence results were first obtained in [2, 4].

1.2 Our Contributions

For particle filters, much of the theoretical literature focuses on the question of consistency in the large ensemble limit, that is, does the empirical approximation converge to the true posterior as the number of particles in the ensemble N approaches infinity. In many high dimensional applications, particle filters are implemented in the non-asymptotic regime, notably in robotics [41] and ocean-atmosphere forecasting [46]. In geosciences, new filtering algorithms have been proposed to beat the curse of dimensionality and are implemented with ensemble sizes many orders of magnitude smaller than the state dimension [47]. In this article, we continue the trend of analyzing algorithms in practical regimes by focusing on the success of particle filters for fixed ensemble sizes. In this regime, we ask the question of whether particle filters inherit important long-time properties from the true posterior distributions. In particular, we address the following questions: if it is known that the true posterior distribution is accurate and conditionally ergodic, can the same be proved of the approximate filter.

We focus our attention on the optimal particle filter (OPF) [1, 49, 27]. The OPF is a sequential importance sampling procedure whereby particle updates are proposed using an optimal convex combination of the model prediction and the observational data at the next time step. For details on the OPF, including the justification for calling it optimal, see [11]. There are two main reasons that we focus our attentions on the OPF. First, the optimal particle filter is known to compare favorably to the standard particle filter, particularly from the perspective of weight degeneracy in high dimensions [39, 40]. Indeed the optimal particle filter can be considered a special case of more complicated filters that have been proposed to beat the curse of dimensionality [6, 47]. Second,

under natural assumptions on the dynamics-observation model, the optimal particle filters can be formulated as a random dynamical system, which is very similar to the 3DVAR algorithm. This means that techniques for proving accuracy for the 3DVAR filter in earlier literature [35] can be leveraged for the OPF.

Throughout the article, we make the assumption of conditional Gaussianity for the dynamics-observation model. This framework is frequently employed in practice, particularly in geoscience data assimilation problems. Under this assumption, we show that the true posterior, the filtering distribution, satisfies the long-time properties of stability and accuracy. The stability result is a type of conditional ergodicity: two copies of the posterior distribution, initialized at different points but conditioned on the same fixed data, will converge to each other in the long time limit, with exponential rate. The accuracy result states that, if sufficiently many variables are observed, the posterior will concentrate around the true trajectory in the long time limit. Related conditional ergodicity results are obtained under quite general assumptions in [44, 45]. However we believe that the approach we adopt in this article is of considerable value, not only here but also in other potential applications, due to the simplicity of the underlying coupling argument, the formulation through random dynamical systems and the explicit link to 3DVAR type algorithms.

We then show that, under the same model-observation assumptions, the OPF also exhibits the long-time properties of stability and accuracy for any fixed ensemble size. For the conditional ergodicity result, we show that the two copies of the particle ensembles, initialized differently, but updated with the same observational data, will converge to each other in the long term limit, in a distributional sense. Both results use very similar arguments to those employed for the posterior.

In addition, we also establish large ensemble consistency results for the OPF. Here we employ a technique exposed very clearly in [34], which finds a recursion that is approximately satisfied by the bootstrap particle filter, and leverages this fact to obtain an estimate on the distance between the true posterior and the empirical approximation. We show that the same idea can be applied to not only the OPF, but a very large class of sequential importance sampling procedures. We would like to comment that large particle consistency results for particle filters should not be considered practical results for high dimensional data assimilation problems, as in practice particle filters are never implemented in this regime. The consistency results are included here as they are practically informative for low dimensional data assimilation problems and moreover as the results are natural consequences of the random dynamical system formulation that has been adopted for accuracy and ergodicity results. For high dimensional data assimilation problems, it may be more practical to look at covariance consistency, as done in [30].

In addition to obtaining the results of stability, accuracy and consistency for the OPF, for which we perform resampling at the end of each assimilation cycle, we also prove the corresponding results for the so called *Gaussianized* OPF. The Gaussianized OPF was introduced in [18] and differs from the OPF only in the implementation of the resampling. Nevertheless, it was shown numerically in [18] that the GOPF compares favorably to the OPF, particularly when applied to high dimensional models. The analysis in this article lends theoretical weight to the advantages of the GOPF over the OPF. In particular we find that the upper bound on the convergence rate for conditional ergodicity for the GOPF has favourable dependence on dimension when compared with the OPF.

1.3 Structure of Article and Notation

The remainder of the article is structured as follows. At the end of this section we introduce some notation and terminology that will be useful in the sequel. In Section 2, we formulate the Bayesian problem of data assimilation, introduce the model-observation assumptions under which we work, and prove the stability and accuracy results for the true posterior. In Section 3, we introduce the bootstrap particle filter, optimal particle filter and Gaussianized optimal particle filter. In Section 4, we prove the conditional ergodicity results for the optimal particle filters. In Section 5, we prove the accuracy results for the optimal particle filters. Finally, in Section 6, we prove the consistency results for the optimal particle filters.

Throughout we let $\mathcal{X}$ denote the finite dimensional Euclidean state space and and we let $\mathcal{Y}$ denote the finite dimensional Euclidean observation space. We write $\mathcal{M}(\mathcal{X})$ for the set of probability measures on $\mathcal{X}$. We denote the Euclidean norm on $\mathcal{X}$ by $|\cdot|$ and for a symmetric positive definite matrix $A \in L(\mathcal{X}, \mathcal{X})$, we define $|\cdot|_A = |A^{-1/2} \cdot|$. We define $S^N : \mathcal{M}(\mathcal{X}) \to \mathcal{M}(\mathcal{X})$ to be the sampling operator $S^N \mu = \frac{1}{N} \sum_{n=1}^{N} \delta_{u^{(n)}}$ where

$u^{(n)} \sim \mu$ are i.i.d. random variables.

2 Bayesian Data Assimilation

We describe the set-up which encompasses all the work in this paper, and then study the conditional ergodicity and accuracy of the true filtering distribution.

2.1 Set-Up

The state model is taken to be a discrete time Markov chain $\{u_k\}_{k \geq 0}$ taking values in the state space $\mathcal{X}$. We assume that the initial condition u_0 of the chain is distributed according to μ_0, where $\mu_0 \in \mathcal{M}(\mathcal{X})$. The transition kernel for the Markov chain is given by $P(u_{k+1}|u_k)$. For each $k \geq 1$, we make an observation of the Markov chain

$$y_{k+1} = h(u_{k+1}) + \eta_{k+1} , \tag{2.1}$$

where $h : \mathcal{X} \to \mathcal{Y}$ maps the state space to the observation space, and $\eta_k \sim N(0, \Gamma)$ are centred i.i.d. random variables representing observational noise. We denote by $Y_k = (y_1, \ldots, y_k)$ the accumulated observational data up to time k. We are interested in studying and approximating the filtering distribution $\mu_k(\cdot) = \mathbb{P}(u_k \in \cdot | Y_k)$ for all $k \geq 1$. We will denote the density of μ_k by $P(u_k|Y_k)$.

The focus of this article is analysis of variants of the optimal particle filter, and in this setting we will always require the following assumptions on the dynamics-observation model:

Assumption 2.1. The dynamics-observation model is given by

$$u_{k+1} = \psi(u_k) + \xi_k, \tag{2.2a}$$

$$y_{k+1} = Hu_{k+1} + \eta_{k+1}, \tag{2.2b}$$

where $u_0 \sim \mu_0, \xi_k \sim N(0, \sigma^2 \Sigma_0)$ i.i.d. and where $\eta_k \sim N(0, \gamma^2 \Gamma_0)$ i.i.d. and $\sigma, \gamma \neq 0$. We require that Σ_0 and Γ_0 are strictly positive-definite, and that the function $\psi(\cdot)$ is bounded.

In order to facilitate comparison with the standard bootstrap filter, for which the theory we describe is sometimes derived under different conditions, we retain the general setting of a dynamics given by a general Markov chain with observations (2.1).

For most of the results in this article, we will be interested in properties of conditional distributions $P(u_k|Y_k)$ when the observational data Y_k is generated by a fixed realization of the model. For this reason, we introduce the following notation to emphasize that we are considering a fixed realization of the data, generated by a fixed trajectory of the underlying dynamical system.

Assumption 2.2. Fix $u_0^\dagger \in \mathcal{X}$ and positive semi-definite matrices Σ_* and Γ_* on $\mathcal{X}$ and $\mathcal{Y}$ respectively. Let $\{u_k^\dagger\}$ be a realization of the dynamics satisfying

$$u_{k+1}^\dagger = \psi(u_k^\dagger) + r\gamma \xi_k^\dagger$$

where $u_0^\dagger \in \mathcal{X}$ is fixed and $\xi_k^\dagger \sim N(0, \Sigma_*)$ i.i.d. and similarly define $\{y_k^\dagger\}$ by

$$y_{k+1}^\dagger = Hu_{k+1}^\dagger + \gamma \eta_{k+1}^\dagger \tag{2.3}$$

where $\eta_{k+1}^\dagger \sim N(0, \Gamma_*)$ i.i.d. We will refer to $\{u_k^\dagger\}_{k \geq 0}$ as the *true signal* and $\{y_k^\dagger\}_{k \geq 1}$ as the given *fixed data*. As above, we use the shorthand $Y_k^\dagger = \{y_i^\dagger\}_{i=1}^k$.

Note that this data is not necessarily generated from the same statistical model used to define the filtering distribution since Σ_* and Γ_* may differ from Σ and Γ, and the initial condition is fixed. Note also that we have set $\sigma = r\gamma$ and that, when studying accuracy, we will consider families of data sets and truths parameterized by $\gamma \to 0$; in this setting it is natural to think of r and the noise sequences $\{\xi_k^\dagger\}_{k \geq 0}$ and $\{\eta_k^\dagger\}_{k \geq 0}$ as fixed, whilst the truth and data sequences will depend on the value of γ.

4

Assumption 2.1 renders the conditional distribution $P(u_k|Y_k)$ tractable as it has a conditionally Gaussian structure described by the inhomogeneous Markov process with transition kernel

$$p_{k+1}(u_k, du_{k+1}) \propto \exp\left(-\frac{1}{2}|y_{k+1} - Hu_{k+1}|_\Gamma^2 - \frac{1}{2}|u_{k+1} - \psi(u_k)|_\Sigma^2 \right) du_{k+1} \tag{2.4}$$

and initialized at the measure μ_0. A simple completion of the square yields an alternative representation for the transition kernel, namely

$$p_{k+1}(u_k, du_{k+1}) \propto \exp\left(-\frac{1}{2}|y_{k+1} - H\psi(u_k)|_S^2 - \frac{1}{2}|u_{k+1} - m_{k+1}|_C^2 \right) du_{k+1} \;, \tag{2.5}$$

where

$$C^{-1} = \Sigma^{-1} + H^*\Gamma^{-1}H \quad, \quad S = H\Sigma H^* + \Gamma$$
$$m_{k+1} = C\left(\Sigma^{-1}\psi(u_k) + H^*\Gamma^{-1}y_{k+1} \right) \;. \tag{2.6}$$

The conditional mean m_{k+1} is often given in *Kalman filter* form

$$m_{k+1} = (I - KH)\psi(u_k) + Ky_{k+1} \;, \tag{2.7}$$

where K is the *Kalman gain* matrix

$$K = \Sigma H^* S^{-1}. \tag{2.8}$$

The expression (2.4) is an application of Bayes' formula in the form

$$p_{k+1}(u_k, du_{k+1}) \propto P(y_{k+1}|u_{k+1})P(u_{k+1}|u_k)du_{k+1}$$

whilst (2.5) follows from a second application of Bayes' formula to derive the identity

$$P(y_{k+1}|u_{k+1})P(u_{k+1}|u_k)du_{k+1} = P(y_{k+1}|u_k)P(u_{k+1}|u_k, y_{k+1})du_{k+1}.$$

We note that

$$P(y_{k+1}|u_k) = Z_S^{-1} \exp\left(-\frac{1}{2}|y_{k+1} - H\psi(u_k)|_S^2 \right)$$
$$P(u_{k+1}|u_k, y_{k+1}) = Z_C^{-1} \exp\left(-\frac{1}{2}|u_{k+1} - m_{k+1}|_C^2 \right) \;. \tag{2.9}$$

These formulae are prevalent in the data assimilation literature; in particular (2.7) describes the evolution of the mean state estimate in the cycled 3DVAR algorithm, setting $u_{k+1} = m_{k+1}$ [25]. We will make use of the formulae in section 3 when describing optimal particle filters as random dynamical systems.

We now state two theorems concerning the ergodicity of the filtering distribution itself. The first shows that, when initialized at two different points, the filtering distributions converge towards one another at a geometric rate. The second shows that, on average over all instances of randomness in the signal process and the observations process, the mean under the filtering distribution is, asymptotically for large k, $\mathcal{O}(\gamma)$ from the truth. We state and prove these two results in this section. The remainder of the paper is devoted to establishing analogous results for various particle filters.

2.2 Conditional Ergodicity

The ergodicity result uses a metric on probability measures to quantify convergence of differently initialized posteriors in the long time limit. To this end, we define the total variation metric on $\mathcal{M}(\mathcal{X})$ by

$$d_{TV}(\mu, \nu) = \frac{1}{2} \sup_{|h| \leq 1} |\mu(h) - \nu(h)|$$

where the supremum is taken over all bounded functions $h : \mathcal{X} \to \mathbb{R}$ with $|h| \leq 1$, and where we define $\mu(h) := \int_{\mathcal{X}} h(x)\mu(dx)$ for any probability measure $\mu \in \mathcal{M}(\mathcal{X})$ and any real-valued test function h bounded by 1 on $\mathcal{X}$.

Theorem 2.3. *Consider the filtering distributions μ_k, μ'_k, defined by Assumption 2.1, and initialized at two different Dirac masses $\mu_0 = \delta_{z_0}$ and $\mu_0 = \delta_{z'_0}$ respectively. Assume moreover that the data appearing in each filtering distribution is the same, and given by $\{y^\dagger_k\}_{k \geq 1}$ defined in Assumption 2.2. Then there exists $\mathfrak{z} \in (0, 1)$ such that, almost surely with respect to the randomness generating the data $\{y^\dagger_k\}_{k \geq 1}$,*

$$\lim_{k \to \infty} \left(d_{\text{TV}}\left(\mu_k, \mu'_k \right) \right)^{1/k} = \mathfrak{z} . \tag{2.10}$$

Proof. The proof uses a standard coupling procedure, as can be found for instance in [32]; we follow the exposition of the methodology in [31]. We divide the proof into three steps; Step A is where we recast the problem as a coupling problem, Step B is where we obtain a minorization condition and Step C is where we apply the law of large numbers to obtain a rate. Subsequent proofs will use the same structure and it will be useful to refer to specific steps in the proof later on.

Step A: We first introduce two Markov chains z_k, z'_k whose laws are given by μ_k, μ'_k respectively. To this end, define z_k to be the inhomogeneous Markov chain, with initialization z_0 co-inciding with the center of the Dirac mass defined above, and with inhomogeneous transition kernel p_k defined by (2.4). If we define recursively the kernels $p^k(z, \cdot)$ by

$$p^{k+1}(z, \cdot) = \int_{\mathcal{X}} p_{k+1}(v, \cdot) p^k(z, dv) \tag{2.11}$$

with $p^0(z, \cdot) = \delta_z(\cdot)$, then it is easy to see that $\mu_k = p^k(z_0, \cdot)$ and $\mu'_k = p^k(z'_0, \cdot)$ for all $k \geq 0$. Since $p^k(z, \cdot)$ is the law of z_k, we see that z_k is indeed the required Markov chain, and similarly for z'_k.

The main step is to prove a minorization condition for the transition kernel p_{k+1}. That is, we seek a measure $\mathbb{Q} \in \mathcal{M}(\mathcal{X})$ and a sequence of constants $\epsilon_k > 0$ satisfying

$$p_{k+1}(u, A) \geq \epsilon_k \mathbb{Q}(A) \tag{2.12}$$

for all $u \in \mathcal{X}$ and all measurable sets $A \subset \mathcal{X}$. Given a minorization condition, we obtain the result via the following coupling argument. The minorization condition allows us to define a new Markov kernel

$$\widetilde{p}_{k+1}(x, A) = (1 - \epsilon_k)^{-1} \big(p_{k+1}(x, A) - \epsilon_k \mathbb{Q}(A) \big) .$$

We then define a Markov chain $\tilde{z}_k$ by $\tilde{z}_0 = z_0$ and

$$\tilde{z}_{k+1} \sim \widetilde{p}_{k+1}(\tilde{z}_k, \cdot) \quad w.p. \, (1 - \epsilon_k) \tag{2.13a}$$
$$\tilde{z}_{k+1} \sim \mathbb{Q}(\cdot) \quad w.p. \, \epsilon_k. \tag{2.13b}$$

and similarly for $\tilde{z}'_k$. By evaluating $\mathbb{E}\varphi(\tilde{z}_k)$ for a suitable class of test functions φ, it is easy to verify that the Markov chain $\tilde{z}_k$ is equivalent in law to z_k.

We derive the minorization condition which allows this coupling below, in Step B. Here in Step A we couple the two Markov chains $\tilde{z}_k, \tilde{z}'_k$ in such a way that $\tilde{z}_k = \tilde{z}'_k$ for all $k \geq \tau$, where τ is the first coupling time, that is, the smallest n such that $\tilde{z}_n = \tilde{z}'_n$. Importantly the two chains $\tilde{z}_k, \tilde{z}'_k$ share the same random variables and live on the same probability space; in particular, once a draw from $\mathbb{Q}(\cdot)$ is made, the chains can be coupled and remain identical thereafter. Let A_k be the event that, in (2.13), the first (state dependent) Markov kernel is picked at times $j = 0, \cdots, k - 1$. Then since $p^k(z_0, \cdot) = \mu_k$ is the law of $\tilde{z}_k$, and similarly for $\tilde{z}'_k$, we have that

$$d_{\text{TV}}\big(p^k(z_0, \cdot), p^k(z'_0, \cdot) \big) = \frac{1}{2} \sup_{|f|_\infty \leq 1} |\mathbb{E}\big(f(\tilde{z}_k) - f(\tilde{z}'_k) \big)|$$

$$= \frac{1}{2} \sup_{|f|_\infty \leq 1} \left| \mathbb{E}\Big(\big(f(\tilde{z}_k) - f(\tilde{z}'_k) \big) \mathbb{I}_{A_k} + \big(f(\tilde{z}_k) - f(\tilde{z}'_k) \big) \mathbb{I}_{A_k^c} \Big) \right| .$$

Note that for this coupling the second term vanishes, as in the event A_k^c, the two chains will have coupled for $\tau < k$. It follows that

$$d_{\text{TV}}\big(p^k(z_0, \cdot), p^k(z'_0, \cdot) \big) \leq \mathbb{E}(\mathbb{I}_{A_k}) = \mathbb{P}(A_k) = \Pi_{j=1}^k (1 - \epsilon_j) .$$

6

To obtain the result (2.10), we need to understand the limiting behaviour of the constants ϵ_j appearing in the minorization condition (2.12). Hence we turn our attention toward obtaining the minorization condition.

Step B: If we recall (2.4) and let $\rho_{k,0}^\dagger = \sigma H \xi_k^\dagger + \gamma \eta_k^\dagger$, then we obtain, via $Z \leq 1$, the identity

$$p_{k+1}(u, dv) \geq \exp\left(-\frac{1}{2}|H\psi(u_k^\dagger) + \rho_{k,0}^\dagger - Hv|_\Gamma^2 - \frac{1}{2}|v - \psi(u)|_\Sigma^2\right)dv \tag{2.14}$$

$$\geq \exp\left(-2|H\psi(u_k^\dagger)|_\Gamma^2 - |\psi(u)|_\Sigma^2 - 2|\rho_{k,0}^\dagger|_\Gamma^2 - |Hv|_\Gamma^2 - |v|_\Sigma^2\right)dv \tag{2.15}$$

$$\geq \exp\left(-\lambda^2 - 2|\rho_{k,0}^\dagger|_\Gamma^2\right)\exp\left(-\frac{1}{2}|v|_D^2\right)dv \tag{2.16}$$

$$\tag{2.17}$$

where

$$\lambda^2 = \sup_{u,v}\left(2|H\psi(v)|_\Gamma^2 + |\psi(u)|_\Sigma^2\right)$$

and

$$\frac{1}{2}D^{-1} = \Sigma^{-1} + H^*\Gamma^{-1}H .$$

Thus we have a minorization condition of the form

$$p_{k+1}(u, A) \geq \epsilon_k \mathbb{Q}(A)$$

where $\mathbb{Q}(\cdot)$ is the Gaussian $N(0, D)$ and $\epsilon_k = \exp\left(-\lambda^2 - 2|\rho_{k,0}^\dagger|_\Gamma^2\right)$.

Step C: By the argument above, it follows that

$$d_{\mathrm{TV}}\left(p^k(z_0, \cdot), p^k(z_0', \cdot)\right)^{1/k} \leq \mathfrak{z}_k$$

where $\mathfrak{z}_k = \left(\Pi_{j=1}^k(1 - \epsilon_j)\right)^{1/k}$. Since the ϵ_k are i.i.d. and integrable, by the law of large numbers, almost surely with respect to the randomness generating the true signal and the data, we have

$$\ln \mathfrak{z}_k = \frac{1}{k}\sum_{j=1}^k \ln(1 - \epsilon_j) \to \mathbb{E}\ln(1 - \epsilon_1) = -\mathbb{E}\sum_{n=1}^\infty \frac{1}{n}\epsilon_1^n .$$

But $\epsilon_1 = c\exp(-2|\rho_{1,0}^\dagger|_\Gamma^2)$ for some $c \in (0, 1]$. Since $\rho_{1,0}^\dagger$ is Gaussian it follows that the nth moment of ϵ_1 scales like $n^{-\frac{1}{2}}$ so that the limit of $\ln \mathfrak{z}_k$ is negative and finite; the result follows. □

2.3 Accuracy

We now discuss accuracy of the posterior filtering distribution in the small noise limit $\gamma \ll 1$.

Assumption 2.4. Let $r = \sigma/\gamma$ and assume that there is $r_c > 0$ such that, for all $r \in [0, r_c)$, the function $(I - KH)\psi(\cdot)$, with K defined through (2.6) and (2.8), is globally Lipschitz on $\mathcal{X}$ with respect to the norm $\|\cdot\|$ and with constant $\alpha = \alpha(r) < 1$.

Theorem 2.5. *Suppose Assumption 2.4 holds for some $r_c > 0$. Then for all $r \in [0, r_c)$ we have*

$$\limsup_{k\to\infty} \mathbb{E}\|u_k - \mathbb{E}^{\mu_k} u_k\|^2 \leq c\gamma^2 ,$$

where $\mathbb{E}^{\mu_k}$ denotes expectation over the posterior μ_k defined through Assumption 2.1 and $\mathbb{E}$ denotes expectation over the dynamical model and the observational data.

Proof. This follows similarly to Corollary 4.3 in [36], using the fact that the mean of the filtering distribution is optimal in the sense that

$$\mathbb{E}\|u_k - \mathbb{E}^{\mu_k} u_k\|^2 \leq \mathbb{E}\|u_k - m_k\|^2$$

for any Y_k−measurable sequence $\{m_k\}$. We use for m_k the 3DVAR filter

$$m_{k+1} = (I - KH)\psi(m_k) + Ky_{k+1}.$$

Let $e_k = u_k - m_k$. Following closely Theorem 4.10 of [25] we obtain

$$\mathbb{E}\|e_{k+1}\|^2_{k+1} \leq \alpha^2 \mathbb{E}\|e_k\|^2 + \mathcal{O}(\gamma^2).$$

Application of the Gronwall lemma, plus use of the optimality property, gives the required bound. $\qquad\square$

3 Particle Filters With Resampling

In this section we introduce the bootstrap particle filter, and the optimal particle filters, in all three cases with resampling at every step. Assumption 2.1 ensures that the three particle filters have an elegant interpretation as a random dynamical system (RDS) which, in addition, is useful for our analyses. We thus introduce the filters in this way before giving the algorithmic definition which is more commonly found in the literature. The bootstrap particle filter will not be the focus of subsequent theory, but does serve as an important motivation for the optimal particle filters, and in particular for the consistency results in Section 6.

For each of the three particle filters, we will make frequent use of a resampling operator, which draws a sample $u_k^{(n)}$ from $\{\widehat{u}_k^{(m)}\}_{m=1}^N$ with weights $\{w_k^{(m)}\}_{m=1}^N$ which sum to one. To define this operator, we define the intervals $I_k^{(m)} = [\alpha_k^{(m)}, \alpha_k^{(m+1)})$ where $\alpha_k^{(m+1)} = \alpha_k^{(m)} + w_k^{(m)}$ and then set

$$u_k^{(n)} = \sum_{m=1}^N \mathbb{I}_{I_k^{(m)}}(r_k^{(n)})\widehat{u}_k^{(m)} \tag{3.1}$$

where $r_k^{(n)} \sim U(0,1)$ i.i.d.. Since the weights sum to one, $r_k^{(n)}$ will lie in exactly one of the intervals $I_k^{(i_*)}$ and we will have $u_k^{(n)} = \widehat{u}_k^{(i_*)}$. We also notice that

$$\sum_{m=1}^N \frac{1}{N}\delta_{u_k^{(m)}} = S^N \sum_{m=1}^N w_k^{(m)}\delta_{\widehat{u}_k^{(m)}}$$

where S^N is the sampling operator defined previously.

3.1 The Bootstrap Particle Filter

The bootstrap particle filter (BPF) approximates the filtering distribution μ_k with an empirical measure

$$\rho_k^N = \sum_{n=1}^N \frac{1}{N}\delta_{u_k^{(n)}} . \tag{3.2}$$

The particle positions $\{u_k^{(n)}\}_{n=1}^N$ are defined as follows.

$$\widehat{u}_{k+1}^{(n)} = \psi(u_k^{(n)}) + \xi_k^{(n)} \quad , \quad \xi_k^{(n)} \sim N(0, \Sigma) \text{ i.i.d.},$$

$$u_{k+1}^{(n)} = \sum_{m=1}^N \mathbb{I}_{I_{k+1}^{(m)}}(r_{k+1}^{(n)})\widehat{u}_{k+1}^{(m)} , \tag{3.3}$$

where the second equation uses the resampling operator defined in (3.1) with weights computed according to

$$w_{k+1}^{(n),*} = \exp(-\frac{1}{2}|y_{k+1} - H\widehat{u}_{k+1}^{(n)}|_\Gamma^2) \quad , \quad w_{k+1}^{(n)} = \frac{w_{k+1}^{(n),*}}{\sum_{j=1}^{N} w_{k+1}^{(j),*}} \ . \tag{3.4}$$

Thus, for each particle in the RDS, we propagate them forward using the dynamical model and then re-sample from the weighted particles to account for the observation likelihood.

As with most particle filters, the motivation for the bootstrap particle filter stems from an importance sampling scheme applied to a particular decomposition of the filtering distribution. By Bayes formula, we have

$$\begin{aligned}
P(u_{k+1}|Y_{k+1}) &= P(u_{k+1}|Y_k, y_{k+1}) \\
&= \frac{1}{P(y_{k+1}|Y_k)} P(y_{k+1}|u_{k+1}, Y_k) P(u_{k+1}|Y_k) \\
&= \int \frac{1}{P(y_{k+1}|Y_k)} P(y_{k+1}|u_{k+1}) P(u_{k+1}|u_k, Y_k) P(u_k|Y_k) du_k \\
&= \int \frac{1}{P(y_{k+1}|Y_k)} P(y_{k+1}|u_{k+1}) P(u_{k+1}|u_k) P(u_k|Y_k) du_k \\
&= \frac{1}{P(y_{k+1}|Y_k)} P(y_{k+1}|u_{k+1}) \int P(u_{k+1}|u_k) P(u_k|Y_k) du_k
\end{aligned} \tag{3.5}$$

The bootstrap particle filter approximates the posterior via a sequential application of importance sampling, using $P(u_{k+1}|Y_k) = \int P(u_{k+1}|u_k) P(u_k|Y_k) du_k$ as the proposal and re-weighting according to the likelihood $P(y_{k+1}|u_{k+1})$. Thus the method is typically described by the following algorithm for updating the particle positions. The particles are initialized with $u_0^{(n)} \sim \mu_0$ and then updated by

1. Draw $\widehat{u}_{k+1}^{(n)} \sim P(u_{k+1}|u_k^{(n)})$.

2. Define the weights $w_{k+1}^{(n)}$ for $n = 1, \ldots, N$ by

$$w_{k+1}^{(n),*} = P(y_{k+1}|\widehat{u}_{k+1}^{(n)}) \quad , \quad w_{k+1}^{(n)} = \frac{w_{k+1}^{(n),*}}{\sum_{m=1}^{N} w_{k+1}^{(m),*}} \ .$$

3. Draw $u_{k+1}^{(n)}$ from $\{\widehat{u}_{k+1}^{(n)}\}_{n=1}^{N}$ with weights $\{w_{k+1}^{(n)}\}_{n=1}^{N}$.

Under Assumption 2.1, it is clear that the sampling and re-weighting procedures are consistent with (3.3). Note that the normalization factor $P(y_{k+1}|Y_k)$ is not required in the algorithm and is instead approximated via the normalization procedure in the second step.

It is useful to define the related measure

$$\widehat{\rho}_k^N = \sum_{n=1}^{N} w_k^{(n)} \delta_{\widehat{u}_k^{(n)}} \ , \tag{3.6}$$

with $\widehat{\rho}_0^N = \mu_0$, which is related to the bootstrap particle filter by $\rho_k^N = S^N \widehat{\rho}_k^N$. As we shall see in Section 6, the advantage of $\widehat{\rho}_k^N$ is that it has a recursive definition which allows for elegant proofs of consistency results [34].

3.2 Optimal Particle Filter

The optimal particle filter with resampling can also be defined as a RDS. We once again approximate the filtering distribution μ_k with an empirical distribution

$$\mu_k^N = \sum_{n=1}^{N} \frac{1}{N} \delta_{u_k^{(n)}} \ . \tag{3.7}$$

Under Assumption 2.1 the particles in this approximation are defined as follows. The particle positions are initialized with $u_0^{(n)} \sim \mu_0$ and then evolve according to the RDS

$$\widehat{u}_{k+1}^{(n)} = (I - KH)\psi(u_k^{(n)}) + Ky_{k+1} + \zeta_k^{(n)} \quad , \quad \zeta_k^{(n)} \sim N(0, C) \text{ i.i.d.}$$

$$u_{k+1}^{(n)} = \sum_{m=1}^{N} \mathbb{I}_{I_{k+1}^{(m)}}(r_{k+1}^{(n)})\widehat{u}_{k+1}^{(m)} . \tag{3.8}$$

where C, S, K are defined in (2.6), (2.7) and as with the BPF, the second equation uses the resampling operator defined in (3.1) but now using weights computed by

$$w_{k+1}^{(n),*} = \exp(-\frac{1}{2}|y_{k+1} - H\psi(u_k^{(n)})|_S^2) \quad , \quad w_{k+1}^{(n)} = \frac{w_{k+1}^{(n),*}}{\sum_{m=1}^{N} w_{k+1}^{(m),*}} . \tag{3.9}$$

In light of the formulae given in (2.9), which are derived under Assumption 2.1, we see that the optimal particle filter is updating the particle positions by sampling from $P(u_{k+1}|u_k^{(n)}, y_{k+1})$ and then re-sampling to account for the likelihood factor $P(y_{k+1}|u_k^{(n)})$. In particular, without necessarily making Assumption 2.1, the optimal particle filter is a sequential importance sampling scheme applied to the following decomposition of the filtering distribution

$$P(u_{k+1}|Y_{k+1}) = \int_{\mathcal{X}} P(u_{k+1}, u_k|Y_{k+1})du_k$$

$$= \int_{\mathcal{X}} P(u_{k+1}|u_k, y_{k+1})P(u_k|Y_{k+1})du_k \tag{3.10}$$

$$= \int_{\mathcal{X}} \frac{P(y_{k+1}|u_k)}{P(y_{k+1}|Y_k)} P(u_{k+1}|u_k, y_{k+1})P(u_k|Y_k)du_k .$$

In the algorithmic setting, the filter is initialized with $u_0^{(n)} \sim \mu_0$, then for $k \geq 0$

1. Draw $\widehat{u}_{k+1}^{(n)}$ from $P(u_{k+1}|u_k^{(n)}, y_{k+1})$

2. Define the weights $w_{k+1}^{(n)}$ for $n = 1, \ldots, N$ by

$$w_{k+1}^{(n),*} = P(y_{k+1}|\widehat{u}_k^{(n)}) \quad , \quad w_{k+1}^{(n)} = \frac{w_{k+1}^{(n),*}}{\sum_{m=1}^{N} w_{k+1}^{(m),*}} .$$

3. Draw $u_{k+1}^{(n)}$ from $\{\widehat{u}_{k+1}^{(m)}\}_{m=1}^{N}$ with weights $\{w_{k+1}^{(m)}\}_{m=1}^{N}$.

It is important to note that, although the OPF is well defined in this general setting for any choice of dynamics-obsevation model, it is only implementable under stringent assumptions on the forward and observation model, such as those given in Assumption 2.1; under this assumption the step 1 may be implemented using the formulae given in (2.9) and exploited in the derivation of (3.8). However we emphasize that models satisfying Assumption 2.1 do arise frequently in practice.

As with the BPF, it is beneficial to consider the related particle filter given by

$$\widehat{\mu}_k^N = \sum_{n=1}^{N} w_k^{(n)} \delta_{\widehat{u}_k^{(n)}} \tag{3.11}$$

for $k \geq 1$ and with $\widehat{\mu}_0^N = \mu_0$. Clearly, we have that $\mu_k^N = S^N \widehat{\mu}_k^N$.

3.3 Gaussianized Optimal Particle Filter

In [18], an alternative implementation of the OPF is investigated and found to have superior performance on a range of test problems, particularly with respect to the curse of dimensionality. We refer to this filter as the Gaussianized optimal particle filter (GOPF). Once again, we approximate the filtering distribution with an empirical measure

$$\nu_k^N = \sum_{n=1}^{N} \frac{1}{N} \delta_{v_k^{(n)}} \ . \tag{3.12}$$

As in the previous subsection, we first describe the filter under Assumption 2.1. The filter is initialized with $v_0^{(n)} \sim \mu_0$, with subsequent iterates generated by the RDS

$$\tilde{v}_k^{(n)} = \sum_{m=1}^{N} \mathbb{I}_{I_{k+1}^{(m)}}(r_{k+1}^{(n)}) v_k^{(m)} \ ,$$

$$v_{k+1}^{(n)} = (I - KH)\psi(\tilde{v}_k^{(n)}) + K y_{k+1} + \zeta_k^{(n)} \quad , \quad \zeta_k^{(n)} \sim N(0,C) \text{ i.i.d.} \tag{3.13}$$

and the weights appearing in the resampling operator are given by

$$w_{k+1}^{(n),*} = \exp(-\frac{1}{2}|y_{k+1} - H\psi(v_k^{(n)})|_S^2) \quad , \quad w_{k+1}^{(n)} = \frac{w_{k+1}^{(n),*}}{\sum_{m=1}^{N} w_{k+1}^{(m),*}} \ . \tag{3.14}$$

Thus, the update procedure for GOPF is *weight-resample-propagate*, as opposed to *propagate-weight-resample* for the OPF. Hence the only difference between the OPF and GOPF is the implementation of the resampling step.

In our analysis it is sometimes useful to consider the equivalent RDS

$$\widehat{v}_{k+1}^{(m,n)} = (I - KH)\psi(v_k^{(m)}) + K y_{k+1} + \zeta_k^{(m,n)} \quad \zeta_k^{(m,n)} \sim N(0,C) \text{ i.i.d.}$$

$$v_{k+1}^{(n)} = \sum_{m=1}^{N} \mathbb{I}_{I_{k+1}^{(m)}}(r_{k+1}^{(n)}) \widehat{v}_{k+1}^{(m,n)} \ . \tag{3.15}$$

The sequences $v_k^{(n)}$ defined in (3.13) and (3.15) agree because for every n there is exactly one $m = m^*(n)$ such that $\widehat{v}_{k+1}^{(m^*(n),n)}$ survives the resampling step. Writing it this way allows certain parts of our subsequent analysis to be performed very similarly for both the OPF and GOPF.

For a general dynamics-observation model, the GOPF is described by the algorithm

1. Define the weights $w_{k+1}^{(n)}$ for $n = 1, \ldots, N$ by

$$w_{k+1}^{(n),*} = P(y_{k+1}|v_k^{(n)}) \quad , \quad w_{k+1}^{(n)} = \frac{w_{k+1}^{(n),*}}{\sum_{m=1}^{N} w_{k+1}^{(m),*}} \ .$$

2. Draw $\tilde{v}_k^{(n)}$ from $\{v_k^{(m)}\}_{m=1}^{N}$ with weights $\{w_{k+1}^{(m)}\}_{m=1}^{N}$.

3. Draw $v_{k+1}^{(n)}$ from $P(u_{k+1}|\tilde{v}_k^{(n)}, y_{k+1})$.

Unlike for the previous filters, there is no need to define an associated 'hatted' measure, as the GOPF can be shown to satisfy a very natural recursion. This will be discussed in Section 6.

4 Ergodicity for Optimal Particle Filters

In this section we study the conditional ergodicity of the two optimal particle filters. The proofs are structurally very similar to that for the filtering distribution itself.

4.1 Optimal Filter

Before stating the conditional ergodicity result, we first need some notation. Define $u_k = (u_k^{(1)}, \ldots, u_k^{(N)})$ to be particle positions defined by the RDS (3.8) with $\mu_0 = \delta_{z_0}$ and similarly with $u_k' = (u_k^{(1)'}, \ldots, u_k^{(N)'})$ with $\mu_0 = \delta_{z_0'}$. Then u_k is a Markov chain taking values on $\mathcal{X}^N$, whose Markov kernel we denote by $q_k(z, \cdot)$. As in the proof of Theorem 2.3, the law of u_k is given by $q^k(z_0, \cdot)$, defined recursively as in (2.11) and similarly the law of u_k' is given by $q^k(z_0', \cdot)$. The conditional ergodicity result states that if the two filters u_k, u_k' are driven by the same observational data, then the law of u_k will converge to the law of u_k' exponentially as $k \to \infty$. [1]

Theorem 4.1. *Consider the OPF particles u_k, u_k' defined above. Assume moreover that the observational data used to define each filter is the same, and given by $\{y_k^\dagger\}_{k \geq 1}$ from Assumption 2.2. Then there exists $\mathfrak{z}_N \in (0, 1)$ such that, almost surely with respect to the randomness generating $\{y_k^\dagger\}_{k \geq 1}$,*

$$\lim_{k \to \infty} \left(d_{\mathrm{TV}}\left(q^k(z_0, \cdot), q^k(z_0', \cdot) \right) \right)^{1/k} = \mathfrak{z}_N \, . \tag{4.1}$$

Proof. We will follow the proof of Theorem 2.3 closely, via Steps A, B and C. Step A proceeds exactly as in the proof of Theorem 2.3. In particular, if we assume a minorization condition

$$q_{k+1}(u, \cdot) \geq \epsilon_k \mathbb{Q}(\cdot) \, , \tag{4.2}$$

and repeat the coupling argument from Step A, with u_k, u_k' in place of z_k, z_k', the we obtain the bound

$$d_{\mathrm{TV}}\left(q^k(z_0, \cdot), q^k(z_0', \cdot) \right) \leq \Pi_{j=1}^k (1 - \epsilon_j) \, .$$

This is indeed how the proof proceeds, with the caveat that the minorization constants ϵ_k depend on the number of particles N. In Step B, we obtain the minorization condition (4.2). Let $\widehat{u}_k = (\widehat{u}_k^{(1)}, \ldots, \widehat{u}_k^{(N)})$ as defined in the RDS (3.8). Before proceeding, we introduce some preliminaries. Using the fact that

$$y_{k+1}^\dagger = H\psi(u_k^\dagger) + \gamma(rH\xi_k^\dagger + \eta_{k+1}^\dagger)$$

and defining

$$a_k = \left((I - KH)\psi(u_k^{(n)}) + KH\psi(u_k^\dagger) \right)_{n=1}^N \quad \zeta_k = \left(\zeta_k^{(n)} \right)_{n=1}^N$$

$$\rho_{k,0}^\dagger = \gamma(rH\xi_k^\dagger + \eta_{k+1}^\dagger), \quad \rho_k^\dagger = \left(\gamma K(rH\xi_k^\dagger + \eta_{k+1}^\dagger) \right)_{n=1}^N$$

we see that

$$\widehat{u}_{k+1} = a_k + \rho_k^\dagger + \zeta_k \, .$$

The next element of the sequence, u_{k+1}, is then defined by the second identity in (3.8). We are interested in the conditional ergodicity of $\{u_k\}_{k=1}^\infty$ with the sequence $\{\rho_k^\dagger\}_{k=1}^\infty$ fixed. By Assumption 2.1, a_k is bounded uniformly in k. We define the covariance operator $\mathfrak{C} \in L(\mathcal{X}^N, \mathcal{X}^N)$ to be a block diagonal covariance with each diagonal entry equal to C and then

$$R = \sup_{(u,v)} \left(|(I - KH)\psi(u) + KH\psi(v)|_C^2 \right) \, ,$$

which is finite by Assumption 2.1.

Now, let E_0 be the event that, upon resampling, every particle survives the resampling. There are $N!$ such permutations. We will do the calculation in the case of a trivial permutation, that is, where each particle is mapped to itself under the resampling. However the bounds which follow work for any permutation of this

---

because we do not use any information about location of the mean; we simply use bounds on the drift ψ. If each particle is mapped to itself, then $u_{k+1}^{(n)} = \widehat{u}_{k+1}^{(n)}$ for all $n = 1, \ldots, N$. It follows that

$$
\begin{aligned}
q_{k+1}(u, A) &= \mathbb{P}(u_{k+1} \in A | u_k = u) \\
&\geq \mathbb{P}(u_{k+1} \in A | u_k = u, E_0) \mathbb{P}(E_0) \\
&= \mathbb{P}(\widehat{u}_{k+1} \in A | u_k = u) \mathbb{P}(E_0) .
\end{aligned}
$$

We first note that

$$
\begin{aligned}
\mathbb{P}(\widehat{u}_{k+1} \in A | u_k = u) &= \frac{1}{\sqrt{(2\pi)^{dN} \det \mathfrak{C}}} \int_A \exp\left(-\frac{1}{2}|x - a_k - \rho_k^\dagger|_{\mathfrak{C}}^2\right) dx \\
&\geq \frac{\exp\left(-|a_k + \rho_k^\dagger|_{\mathfrak{C}}^2\right)}{\sqrt{(2\pi)^{dN} \det \mathfrak{C}}} \int_A \exp\left(-|x|_{\mathfrak{C}}^2\right) dx \\
&\geq 2^{-dN/2} \exp(-2|a_k|_{\mathfrak{C}}^2) \exp(-2|\rho_k^\dagger|_{\mathfrak{C}}^2) \mathbb{Q}_{\mathfrak{C}}(A) \\
&\geq 2^{-dN/2} \exp(-2NR^2) \exp(-2|\rho_k^\dagger|_{\mathfrak{C}}^2) \mathbb{Q}_{\mathfrak{C}}(A)
\end{aligned}
$$

where $\mathbb{Q}_{\mathfrak{C}}(A)$ is the Gaussian measure $N(0, \frac{1}{2}\mathfrak{C})$. Thus we have shown that

$$
\mathbb{P}(\widehat{u}_{k+1} \in A | u_k = u) \geq \delta_k \mathbb{Q}_{\mathfrak{C}}(A) \tag{4.3}
$$

where

$$
\delta_k = 2^{-dN/2} \exp(-2NR^2) \exp(-2|\rho_k^\dagger|_{\mathfrak{C}}^2) .
$$

Moreover, we have that

$$
\mathbb{P}(E_0) = N! \Pi_{n=1}^N w_{k+1}^{(n)}.
$$

Note that we have the bound $w_{k+1}^{(n)} \geq w_{k+1}^{(n),*}/N$ for each $n = 1, \ldots, N$ because each $w_{k+1}^{(m),*}$ is bounded by 1. But we have

$$
\begin{aligned}
w_{k+1}^{(n),*} &= \exp\left(-\frac{1}{2}|y_{k+1} - H\psi(u_k^{(n)})|_S^2\right) \\
&= \exp\left(-\frac{1}{2}|H\psi(u_k^\dagger) - H\psi(u_k^{(n)}) + \rho_{k,0}^\dagger|_S^2\right) \\
&\geq \exp\left(-r^2 - |\rho_{k,0}^\dagger|_S^2\right)
\end{aligned}
$$

where

$$
r^2 = \sup_{u,v} |H\psi(u) - H\psi(v)|_S^2
$$

which is finite by Assumption 2.1. From this we see that

$$
\mathbb{P}(E_0) \geq N! \frac{1}{N^N} \exp\left(-Nr^2 - N|\rho_{k,0}^\dagger|_S^2\right).
$$

Thus we obtain the minorization conditon (4.2) where

$$
\epsilon_k = N! \frac{1}{N^N} \exp\left(-Nr^2 - N|\rho_{k,0}^\dagger|_S^2\right) \delta_k \quad , \quad \mathbb{Q} = \mathbb{Q}_{\mathfrak{C}} .
$$

Finally, Step C follows identically to the proof of Theorem 2.3 and the proof is complete.

$\square$

4.2 Gaussianized Optimal Filter

As in the last section, we define $v_k = (v_k^{(1)}, \ldots, v_k^{(N)})$ and similarly for v_k' using the RDS but now for the GOPF (3.13) (or alternatively (3.15)) with distinct initializations $\mu_0 = \delta_{z_0}$ and $\mu_0 = \delta_{z_0'}$. As with Theorem 4.1, we let $q^k(z_0, \cdot)$ denote the law of v_k.

Theorem 4.2. *Consider the GOPF particles v_k, v_k' defined above. Assume moreover that the observational data used to define each filter is the same, and given by $\{y_k^\dagger\}_{k \geq 1}$ from Assumption 2.2. Then there exists $\mathfrak{z}_N \in (0, 1)$ such that, almost surely with respect to the randomness generating $\{y_k^\dagger\}_{k \geq 1}$,*

$$\lim_{k \to \infty} \left(d_{\mathrm{TV}} \left(q^k(z_0, \cdot), q^k(z_0', \cdot) \right) \right)^{1/k} = \mathfrak{z}_N . \tag{4.4}$$

Proof. The proof follows similarly to that of Theorem 4.1, in particular it suffices to obtain a minorization condition for $q_{k+1}(v, \cdot)$. We will use the RDS representation (3.15), which we now recall

$$\begin{aligned} \widehat{v}_{k+1}^{(m,n)} &= (I - KH)\psi(v_k^{(m)}) + Ky_{k+1} + \zeta_k^{(m,n)} \quad \zeta_k^{(m,n)} \sim N(0, C) \text{ i.i.d.} \\ v_{k+1}^{(n)} &= \sum_{m=1}^{N} \mathbb{I}_{I_{k+1}^{(m)}}(r_{k+1}^{(n)}) \widehat{v}_{k+1}^{(m,n)} . \end{aligned} \tag{4.5}$$

In this formulation, note that for each n there is one and only one $m = m^*(n)$ such that $\mathbb{I}_{I_k^{(m)}}(r_k^{(n)}) = 1$. We see that

$$v_k := (v_k^{(n)})_{n=1}^N = (\widehat{v}_k^{(m^*(n),n)})_{n=1}^N .$$

Using the fact that

$$y_{k+1}^\dagger = H\psi(u_k^\dagger) + \gamma(rH\xi_k^\dagger + \eta_{k+1}^\dagger)$$

and defining

$$\begin{aligned} a_k &= \left((I - KH)\psi(v_k^{(m^*(n))}) + KH\psi(u_k^\dagger) \right)_{n=1}^N \quad \zeta_k = \left(\zeta_k^{(m^*(n),n)} \right)_{n=1}^N \\ \rho_{k,0}^\dagger &= \gamma(rH\xi_k^\dagger + \eta_{k+1}^\dagger), \quad \rho_k^\dagger = \left(\gamma K(rH\xi_k^\dagger + \eta_{k+1}^\dagger) \right)_{n=1}^N \end{aligned}$$

we see that

$$v_{k+1} = a_k + \rho_k^\dagger + \zeta_k .$$

Now notice that

$$\begin{aligned} q_{k+1}(v, A) = \mathbb{P}(v_{k+1} \in A | v_k = v) &= \mathbb{P}\left((\widehat{v}_{k+1}^{(m^*(n),n)})_{n=1}^N \in A | v_k = v \right) \\ &= \frac{1}{\sqrt{(2\pi)^{dN} \det \mathfrak{C}}} \int_A \exp\left(-\frac{1}{2} |x - a_k - \rho_k^\dagger|_{\mathfrak{C}}^2 \right) dx \\ &\geq \frac{\exp\left(-|a_k + \rho_k^\dagger|_{\mathfrak{C}}^2 \right)}{\sqrt{(2\pi)^d \det \mathfrak{C}}} \int_A \exp\left(-|x|_{\mathfrak{C}}^2 \right) dx \\ &\geq 2^{-dN/2} \exp(-2|a_k|_{\mathfrak{C}}^2) \exp(-2|\rho_k^\dagger|_{\mathfrak{C}}^2) \mathbb{Q}_{\mathfrak{C}}(A) \\ &\geq 2^{-dN/2} \exp(-2NR^2) \exp(-2|\rho_k^\dagger|_{\mathfrak{C}}^2) \mathbb{Q}_{\mathfrak{C}}(A) \end{aligned}$$

where $\mathbb{Q}_{\mathfrak{C}}$ is the Gaussian measure $N(0, \frac{1}{2}\mathfrak{C})$. Thus we have shown that

$$q_{k+1}(v, A) \geq \delta_k \mathbb{Q}_{\mathfrak{C}}(A) \tag{4.6}$$

where

$$\delta_k = 2^{-dN/2} \exp(-2NR^2) \exp(-2|\rho_k^\dagger|_{\mathfrak{C}}^2) .$$

The remainder of the proof (step C) follows identically to Theorem 2.3, Theorem 4.1. $\qquad \square$

Remark 4.3. We can compare our (upper bounds on the) rates of convergence for the true filtering distribution, and the two optimal filters, using the minorization constants. From the proof of Theorem 2.3, we see that the minorization constants determine the rate of convergence in the statement of conditional ergodicity. In particular for k large we have that

$$d_{TV}(\mu_k, \mu_k') \lesssim \mathfrak{z}^k .$$

where

$$\log \mathfrak{z} = \mathbb{E} \log(1 - \epsilon_1) .$$

The corresponding statements for the law of the particles in each optimal particle filter hold similarly, but ϵ_1 depends on the number of particles. For the OPF we have

$$\epsilon_1 = N! \frac{1}{N^N} \exp\left(-Nr^2 - N|\rho_{1,0}^\dagger|_S^2\right) \delta_1 ,$$

where

$$\delta_1 = 2^{-dN/2} \exp(-2NR^2) \exp(-2|\rho_1^\dagger|_{\mathfrak{c}}^2) ;$$

for the GOPF we simply have $\epsilon_1 = \delta_1$. The extra N dependence in the OPF clearly leads to a slower (upper bound on the) rate of convergence for the OPF. Thus, by this simple argument, we obtain a better convergence rate for the GOPF than for the OPF. This suggests that the GOPF may have a better rate of convergence for fixed ensemble sizes; further analysis or experimental study of this point would be of interest.

5   Accuracy for Optimal Particle Filters

In this section we study the accuracy of the optimal particle filters, in the small noise limit $\gamma \to 0$. The expectation appearing in the theorem statements is with respect to the noise generating the data, and with respect to the randomness within the particle filter itself. Note that this situation differs from that in the accruacy result for the filter itself which uses data generated by the statistical model itself. Assumption 2.2 relaxes this assumption.

5.1   Optimal Particle Filter

Theorem 5.1. *Let Assumption 2.4 hold and consider the OPF with particles $\{u_k^{(n)}\}_{n=1}^N$ defined by (3.8) with data $\{y_k^\dagger\}$ from Assumption 2.2. It follows that there is constant c such that*

$$\limsup_{k \to \infty} \mathbb{E}\left(\max_n \|u_k^{(n)} - u_k^\dagger\|^2\right) \le c\gamma^2.$$

Proof. First recall the notation $\Sigma = \sigma\Sigma_0$, $\Gamma = \gamma\Gamma_0$ and $r = \sigma/\gamma$. Now define

$$S_0 = r^2 H\Sigma_0 H^* + \Gamma_0,$$
$$C_0 = r^2(I - KH)\Sigma_0$$

and note that

$$S = \gamma^2 S_0, C = \gamma^2 C_0, K = r^2 \Sigma_0 H^* S_0^{-1} .$$

We will use the RDS representation

$$\widehat{u}_{k+1}^{(n)} = (I - KH)\psi(u_k^{(n)}) + Ky_{k+1}^\dagger + \gamma\zeta_{0,k}^{(n)} \quad , \quad \zeta_k^{(n)} \sim N(0, C) \text{ i.i.d.}$$

$$u_{k+1}^{(n)} = \sum_{m=1}^N \mathbb{I}_{I_{k+1}^{(m)}}(r_{k+1}^{(n)})\widehat{u}_{k+1}^{(n)} .$$

(5.1)

where $\zeta_{0,k}^{(n)} \sim N(0, C_0)$ i.i.d.. Hence we have

$$u_{k+1}^\dagger = (I - KH)\psi(u_k^\dagger) + KH\psi(u_k^\dagger) + r\gamma\xi_k^\dagger,$$

15

$$\widehat{u}_{k+1}^{(n)} = (I - KH)\psi(u_k^{(n)}) + K(H\psi(u_k^\dagger) + \gamma\eta_{k+1}^\dagger) + \gamma\zeta_{0,k}^{(n)}.$$

where $\zeta_{0,k}^{(n)} \sim N(0, C_0)$ i.i.d. Subtracting, we obtain

$$\widehat{u}_{k+1}^{(n)} - u_{k+1}^\dagger = (I - KH)\big(\psi(u_k^{(n)}) - \psi(u_k^\dagger)\big) + \gamma\iota_k^{(n)} \tag{5.2}$$

where $\iota_k^{(n)} := (K\eta_{k+1}^\dagger + \zeta_{0,k}^{(n)} - r\xi_k^\dagger)$. Moreover we have the identity

$$u_{k+1}^\dagger = \sum_{m=1}^{N} \mathbb{I}_{I_{k+1}^{(m)}}(r_{k+1}^{(n)})u_{k+1}^\dagger \ . \tag{5.3}$$

Thus, defining

$$e_k^{(n)} = u_k^{(n)} - u_k^\dagger, \quad \widehat{e}_k^{(n)} = \widehat{u}_k^{(n)} - u_k^\dagger$$

we have from (5.1) and (5.3)

$$e_{k+1}^{(n)} = \sum_{m=1}^{N} \mathbb{I}_{I_{k+1}^{(m)}}(r_{k+1}^{(n)})\widehat{e}_{k+1}^{(m)} \ .$$

Thus

$$\max_n \|e_{k+1}^{(n)}\|^2 \leq \max_m \|\widehat{e}_{k+1}^{(m)}\|^2 \ ,$$

where the norm is the one in which we have a contraction. Using (5.2), the Lipschitz property of $(I - KH)\psi(\cdot)$, taking expectations and using independence, yields

$$\mathbb{E}\Big(\max_n \|u_{k+1}^{(n)} - u_{k+1}^\dagger\|^2\Big) \leq \alpha^2 \mathbb{E}\Big(\max_n \|u_k^{(n)} - u_k^\dagger\|^2\Big) + \mathcal{O}(\gamma^2)$$

and the result follows by Gronwall. $\qquad\square$

5.2 Gaussianized Optimal Filter

Theorem 5.2. *Let Assumption 2.4 hold and consider the GOPF with particles $\{v_k^{(n)}\}_{n=1}^N$ defined by* (3.13) *(or* (3.15)*) with data $\{y_k^\dagger\}$ from Assumption 2.2. It follows that there is constant c such that*

$$\limsup_{k\to\infty} \mathbb{E}\Big(\max_n \|v_k^{(n)} - u_k^\dagger\|^2\Big) \leq c\gamma^2.$$

Proof. Recall the notation defined at the beginning of the proof of Theorem 5.1. Recall also the RDS representation of the GOPF (3.15)

$$\widehat{v}_{k+1}^{(m,n)} = (I - KH)\psi(v_k^{(m)}) + Ky_{k+1} + \gamma\zeta_{0,k}^{(m,n)}$$
$$v_{k+1}^{(n)} = \sum_{m=1}^{N} \mathbb{I}_{I_{k+1}^{(m)}}(r_{k+1}^{(n)})\widehat{v}_{k+1}^{(m,n)} \ . \tag{5.4}$$

where we now have $\zeta_{0,k}^{(m,n)} \sim N(0, C_0)$ i.i.d. , recalling that $C = \gamma^2 C_0$. We also have the identity

$$u_{k+1}^\dagger = (I - KH)\psi(u_k^\dagger) + KH\psi(u_k^\dagger) + r\gamma\xi_k^\dagger \ , \tag{5.5}$$

where $\zeta_{0,k}^{(n)} \sim N(0, C_0)$ i.i.d. Subtracting, we obtain

$$\widehat{v}_{k+1}^{(m,n)} - u_{k+1}^\dagger = (I - KH)\big(\psi(v_k^{(m)}) - \psi(u_k^\dagger)\big) + \gamma\iota_k^{(n)} \tag{5.6}$$

where $\iota_k^{(n)} := (K\eta_{k+1}^\dagger + \zeta_{0,k}^{(n)} - r\xi_k^\dagger)$. Note that

$$u_{k+1}^\dagger = \sum_{m=1}^{N} \mathbb{I}_{I_{k+1}^{(m)}}(r_{k+1}^{(n)})u_{k+1}^\dagger \tag{5.7}$$

so that, defining

$$e_k^{(n)} = v_k^{(n)} - u_k^\dagger, \quad \widehat{e}_k^{(m,n)} = \widehat{v}_k^{(m,n)} - u_k^\dagger$$

we have from (5.4), (5.6) and (5.7)

$$e_{k+1}^{(n)} = \sum_{m=1}^{N} \mathbb{I}_{I_{k+1}^{(m)}}(r_{k+1}^{(n)})\widehat{e}_{k+1}^{(m,n)}.$$

Thus

$$\max_n \|e_{k+1}^{(n)}\|^2 \leq \max_m \|\widehat{e}_{k+1}^{(m,n)}\|^2 \ ,$$

where the norm is the one in which we have a contraction. Using (5.6), using the Lipschitz property of $(I - KH)\psi(\cdot)$, taking expectations and using independence, gives

$$\mathbb{E}\Big(\max_n \|v_{k+1}^{(n)} - u_{k+1}^\dagger\|^2\Big) \leq \alpha^2 \mathbb{E}\Big(\max_n \|v_k^{(n)} - u_k^\dagger\|^2\Big) + \mathcal{O}(\gamma^2) \ .$$

The result follows by Gronwall. □

6 Consistency in the Large Particle Limit

In this section we state and prove consistency results for the BPF, OPF and GOPF introduced in Section 3. For the BPF the result is well known but we reproduce it here as it serves as an ideological template for the more complicated proofs to follow; furthermore we present the clean proof given in [34] (see also [25, Chapter 4]) as this particular approach to the result generalizes naturally to the OPF and GOPF.

6.1 Bootstrap Particle Filter

In the following, we let $f_{k+1} : \mathcal{X} \to \mathbb{R}$ be any function with $f_{k+1}(u_{k+1}) \propto P(y_{k+1}|u_{k+1})$; any proportionality constant will suffice, but the normalization constant is of course natural. As in previous sections, we let μ_k denote the filtering distribution. The following theorem is stated and then proved through a sequence of lemmas in the remainder of the subsection.

Theorem 6.1. *Let $\widehat{\rho}_k^N, \rho_k^N$ be the BPFs defined by* (3.6), (3.2) *respectively, and suppose that there exists a constant $\kappa \in (0,1]$ such that*

$$\kappa \leq f_{k+1}(u_{k+1}) \leq \kappa^{-1} \tag{6.1}$$

for all $u_{k+1} \in \mathcal{X}$, $y_{k+1} \in \mathcal{Y}$ and $k \in \{0,\ldots,K-1\}$. Then we have

$$d(\widehat{\rho}_K^N, \mu_K) \leq \sum_{k=1}^{K}(2\kappa^{-2})^k N^{-1/2} \tag{6.2}$$

and

$$d(\rho_K^N, \mu_K) \leq \sum_{k=0}^{K}(2\kappa^{-2})^k N^{-1/2} \tag{6.3}$$

for all $K, N \geq 1$.

Remark 6.2. Note that the constant κ^{-2} appearing in the estimates above arises as the ratio of the upper and lower bounds in (6.1). In particular, we cannot optimize κ by choosing a different proportionality constant for f_{k+1}.

It is straightforward to check that the filtering distribution satisfies the recursion

$$\mu_{k+1} = L_{k+1}P\mu_k \tag{6.4}$$

as this is nothing more than a statement of the Bayes formula calculation given in (3.5). Here $P : \mathcal{M}(\mathcal{X}) \to \mathcal{M}(\mathcal{X})$ is the Markov semigroup

$$P\nu(A) = \int_A P(u_{k+1}|u_k)\nu(du_k),$$

which gives the prior, and $L_{k+1} : \mathcal{M}(\mathcal{X}) \to \mathcal{M}(\mathcal{X})$ is the likelihood operator

$$L_{k+1}\nu(A) = \int_A Z^{-1}P(y_{k+1}|u_{k+1})\nu(du_{k+1})$$

for each $A \subset \mathcal{X}$ measurable and Z is a normalization constant.

In terms of understanding the approximation properties of the BPF, the key observation is that the measures $\{\widehat{\rho}_k^N\}_{k\geq 0}$ satisfy the recursion

$$\widehat{\rho}_{k+1}^N = L_{k+1}S^N P\widehat{\rho}_k^N \qquad \widehat{\rho}_0^N = \mu_0. \tag{6.5}$$

where $P : \mathcal{M}(\mathcal{X}) \to \mathcal{M}(\mathcal{X})$ is the Markov semigroup and, as defined in Section 1.3, $S^N : \mathcal{M}(\mathcal{X}) \to \mathcal{M}(\mathcal{X})$ is the sampling operator. The convergence of the measures is quantified by the metric on random elements of $\mathcal{M}(\mathcal{X})$ defined by

$$d(\mu, \nu) = \sup_{|f|_\infty \leq 1} \sqrt{\mathbb{E}^\omega |\mu(f) - \nu(f)|^2}\,,$$

where, in our setting, $\mathbb{E}^\omega$ will always denote expectation with respect to the randomness in the sampling operator S^N; this metric reduces to twice the total variation metric, used in studying ergodicity, when the measures are not random. The main ingredients for the proof are the following three estimates for the operators P, S^N and L_{k+1} with respect to the metric d.

Lemma 6.3. *We have the following*

1. $\sup_{\nu \in \mathcal{M}(\mathcal{X})} d(S^N\nu, \nu) \leq N^{-1/2}$.

2. $d(P\mu, P\nu) \leq d(\mu, \nu)$ *for all* $\mu, \nu \in \mathcal{M}(\mathcal{X})$.

Proof. See [25, Lemma 4.7, Lemma 4.8]. $\qquad\square$

We state the following Lemma in a slightly more general form than necessary for the BPF, as it will be applied in different contexts for the optimal particle filters.

Lemma 6.4. *Let $\mathcal{Z}$ be a finite dimensional Euclidean space. Suppose that $g_{k+1} : \mathcal{Z} \to [0, \infty)$ is bounded and that there exists $\kappa \in (0, 1]$ such that*

$$\kappa \leq g_{k+1}(u) \leq \kappa^{-1} \tag{6.6}$$

for all $u \in \mathcal{Z}$ and define $G_{k+1} : \mathcal{M}(\mathcal{Z}) \to \mathcal{M}(\mathcal{Z})$ by $G_{k+1}(\nu)(\varphi) = \nu(g_{k+1}\varphi)/\nu(g_{k+1})$. Then

$$d(G_{k+1}\mu, G_{k+1}\nu) \leq (2\kappa^{-2})d(\mu, \nu)$$

for all $\mu, \nu \in \mathcal{M}(\mathcal{Z})$.

Proof. See [25, Lemma 4.9]. $\qquad\square$

We can now prove the consistency result.

Proof of Theorem 6.1. First note that, taking $\mathcal{Z} = \mathcal{X}$ and $g_{k+1} = f_{k+1}$ in Lemma 6.4, we obtain $G_{k+1}\nu = L_{k+1}\nu$. Thus, by (6.1), it follows that $d(L_{k+1}\mu, L_{k+1}\nu) \leq (2\kappa^{-2})d(\mu, \nu)$ for all $\mu, \nu \in \mathcal{M}(\mathcal{X})$. Combining this fact with the recursions given in (6.5), (6.4) and the estimates given in Lemmas 6.3 we have

$$d(\widehat{\rho}_{k+1}^N, \mu_{k+1}) = d(L_{k+1}S^N P\widehat{\rho}_k^N, L_{k+1}P\mu_k)$$

$$\leq 2\kappa^{-2}d(S^N P\widehat{\rho}_k^N, P\mu_k)$$
$$\leq 2\kappa^{-2}\left(d(S^N P\widehat{\rho}_k^N, P\widehat{\rho}_k^N) + d(P\widehat{\rho}_k^N, P\mu_k)\right)$$
$$\leq 2\kappa^{-2}N^{-1/2} + 2\kappa^{-2}d(\widehat{\rho}_k^N, \mu_k).$$

And since $\widehat{\rho}_0^N = \mu_0$, we obtain (6.2) by induction. Moreover, since $\rho_k = S^N \widehat{\rho}_k^N$

$$d(\rho_k, \mu_k) = d(S^N\widehat{\rho}_k^N, \mu_k) \leq d(S^N\widehat{\rho}_k^N, \widehat{\rho}_k^N) + d(\widehat{\rho}_k^N, \mu_k)$$

and (6.3) follows. $\qquad\qquad\square$

6.2 Sequential Importance Resampler

In this section we will apply the above strategy to prove the corresponding consistency result for the OPF. Instead of restricting to the OPF, we will obtain results for the sequential importance resampler (SIR), for which the OPF is a special case. See [11, sections II, III] for background in sequential importance sampling, and on the use of resampling. As with the OPF, the SIR is an empirical measure

$$\mu_k^N = \sum_{n=1}^{N} \frac{1}{N}\delta_{u_k^{(n)}}. \tag{6.7}$$

We will abuse notation slightly by keeping the same notation for the OPF and the SIR. The particle positions are drawn from a proposal distribution $\pi(u_{k+1}|u_k, y_{k+1})$ and re-weighted accordingly. As usual, the positions are initialized with $u_0^{(n)} \sim \mu_0$ and updated by

1. Draw $\widehat{u}_{k+1}^{(n)}$ from $\pi(u_{k+1}|u_k^{(n)}, y_{k+1})$

2. Define the weights $w_{k+1}^{(n)}$ for $n = 1, \ldots, N$ by

$$w_{k+1}^{(n),*} = \frac{P(y_{k+1}|\widehat{u}_{k+1}^{(n)})P(\widehat{u}_{k+1}^{(n)}|u_k^{(n)})}{\pi(\widehat{u}_{k+1}^{(n)}|u_k^{(n)}, y_{k+1})} \quad , \quad w_{k+1}^{(n)} = \frac{w_{k+1}^{(n),*}}{\sum_{m=1}^{N} w_{k+1}^{(m),*}}.$$

3. Draw $u_{k+1}^{(n)}$ from $\{\widehat{u}_{k+1}^{(m)}\}_{m=1}^{N}$ with weights $\{w_{k+1}^{(m)}\}_{m=1}^{N}$.

Thus, if we take the proposal to be $\pi(u_{k+1}|u_k, y_{k+1}) = P(u_{k+1}|u_k, y_{k+1})$ then we obtain the OPF (3.7). Without being more specific about the proposal π, it is not possible to represent the SIR as a random dynamical system in general.

Precisely as with the OPF, for the SIR we define the related filter

$$\widehat{\mu}_k^N = \sum_{n=1}^{N} w_k^{(n)}\delta_{\widehat{u}_k^{(n)}} \tag{6.8}$$

with $\widehat{\mu}_0^N = \mu_0$ and note the important identity $\mu_k^N = S^N\widehat{\mu}_k^N$. The following theorem, and corollary, are proved in the remainder of the subsection, through a sequence of lemmas.

Theorem 6.5. *Let $\widehat{\mu}^N, \mu^N$ be the SIR filters defined by (6.8), (6.7) respectively, with proposal distribution π. Suppose that there exists $f_{k+1} : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ with*

$$f_{k+1}(u_{k+1}, u_k) \propto \frac{P(y_{k+1}|u_{k+1})P(u_{k+1}|u_k)}{\pi(u_{k+1}|u_k, y_{k+1})} \tag{6.9}$$

and satisfying

$$\kappa \leq f_{k+1}(u_{k+1}, u_k) \leq \kappa^{-1} \tag{6.10}$$

for all $u_{k+1}, u_k \in \mathcal{X}$, $k \in \{0, \ldots, K-1\}$ and some $\kappa \in (0, 1]$. Then we have

$$d(\widehat{\mu}_K^N, \mu_K) \leq \sum_{k=1}^{K} (2\kappa^{-2})^k N^{-1/2} \tag{6.11}$$

and

$$d(\mu_K^N, \mu_K) \leq \sum_{k=0}^{K} (2\kappa^{-2})^k N^{-1/2} \tag{6.12}$$

for all $K, N \geq 1$.

Remark 6.6. As for the boostrap particle filter, the appearance of κ^{-2} reflects the ratio of the upper and lower bounds in (6.9); hence there is nothing to be gained from optimizing over the constant of proportionality. If we let

$$f_{k+1}(u_{k+1}, u_k) = \frac{P(y_{k+1}|u_{k+1})P(u_{k+1}|u_k)}{\pi(u_{k+1}|u_k, y_{k+1})}$$

then the estimate (6.10) is equivalent to

$$\kappa\pi(u_{k+1}|u_k, y_{k+1}) \leq P(y_{k+1}|u_{k+1})P(u_{k+1}|u_k) \leq \kappa^{-1}\pi(u_{k+1}|u_k, y_{k+1}) .$$

This can thus be interpreted as a quantification of equivalence between measures π and the optimal proposal $P(y_{k+1}|u_{k+1})P(u_{k+1}|u_k)$.

Remark 6.7. It is important to note that Assumption 2.1 on the dynamics-observation model is not required by Theorem 6.5. For the consistency results, Assumption 2.1 is only used to ensure that (6.10) holds. This observation leads to the following corollary.

Corollary 6.8. *Let $\widehat{\mu}^N, \mu^N$ be the OPFs defined in* (3.11), (3.7) *respectively and satisfying Assumption 2.1. Then there is $\kappa = \kappa(Y_K)$ such that we have*

$$d(\widehat{\mu}_K^N, \mu_K) \leq \sum_{k=1}^{K} (2\kappa^{-2})^k N^{-1/2} \tag{6.13}$$

and

$$d(\mu_K^N, \mu_K) \leq \sum_{k=0}^{K} (2\kappa^{-2})^k N^{-1/2} \tag{6.14}$$

for all $K, N \geq 1$ and where $\kappa^{-1} = \exp\left(\max_{0 \leq j \leq K-1} |y_{j+1}|^2 + \sup_v |H\psi(v)|_S^2\right)$.

Although similar to the argument for the BPF, the recursion argument for the SIR is necessarily more complicated than that for the BPF, as the weights $w_{k+1}^{(n)}$ can potentially depend on both $u_{k+1}^{(n)}$ and $u_k^{(n)}$. This suggests that we must build a recursion which updates measures on a joint space $(u_{k+1}, u_k) \in \mathcal{X} \times \mathcal{X}$. This would also be necessary if we restricted our attention to the OPF, as the weights are defined using $u_k^{(n)}$ and not the particle positions $\widehat{u}_{k+1}^{(n)}$ after the proposal.

The recursion is defined using the following three operators.

1. First P_{k+1}^π maps probability measures on $\mathcal{X}$ to probability measures on $\mathcal{X} \times \mathcal{X}$ by

$$P_{k+1}^\pi \mu(A) = \iint_A \pi(u_{k+1}|u_k, y_{k+1})\mu(du_k)du_{k+1}$$

where A is a measurable subset of $\mathcal{X} \times \mathcal{X}$.

2. The reweighting operator L_{k+1}^π maps probability measures on $\mathcal{X} \times \mathcal{X}$ to probability measures on $\mathcal{X} \times \mathcal{X}$ and is defined by

$$L_{k+1}^\pi Q(A) = Z^{-1} \iint_A w_{k+1}(u_{k+1}, u_k) Q(du_{k+1}, du_k)$$

where Z is the normalization constant of the resulting measure. The weight function is given by

$$w_{k+1}(u_{k+1}, u_k) = \frac{P(y_{k+1}|u_{k+1})P(u_{k+1}|u_k)}{\pi(u_{k+1}|u_k, y_{k+1})} \ .$$

3. Finally, M maps probability measures on $\mathcal{X} \times \mathcal{X}$ into probability measures on $\mathcal{X}$ via marginalization onto the first component:

$$MQ(B) = \int \int_{B \times \mathcal{X}} Q(du_{k+1}, du_k)$$

.

It is easy to see that the posterior μ_k satisfies a natural recursion in terms of these operators.

Lemma 6.9. $\mu_{k+1} = ML_{k+1}^\pi P_{k+1}^\pi \mu_k$

Proof. Let $P(u_k|Y_k)$ denote the density of μ_k, then $P_{k+1}^\pi \mu_k$ is a measure on $\mathcal{X} \times \mathcal{X}$ with density

$$\pi(u_{k+1}|u_k, y_{k+1})P(u_k|Y_k) \ .$$

And $L_{k+1}^\pi P_{k+1}^\pi \mu_k$ is a measure on $\mathcal{X} \times \mathcal{X}$ with density

$$Z^{-1} w_{k+1}(u_{k+1}, u_k)\pi(u_{k+1}|u_k, y_{k+1})P(u_k|Y_k) = Z^{-1} P(y_{k+1}|u_{k+1})P(u_{k+1}|u_k)P(u_k|Y_k) \ .$$

Finally, $ML_{k+1}^\pi P_{k+1}^\pi \mu_k$ is a measure on $\mathcal{X}$ with density

$$\int_{\mathcal{X}} Z^{-1} P(y_{k+1}|u_{k+1})P(u_{k+1}|u_k)P(u_k|Y_k)du_k = Z^{-1}P(y_{k+1}|u_{k+1})P(u_{k+1}|Y_k) \ . \qquad (6.15)$$

Similarly, for the normalization factor, we have

$$Z = \int \int_{\mathcal{X} \times \mathcal{X}} P(y_{k+1}|u_{k+1})P(u_{k+1}|u_k)P(u_k|Y_k)du_k du_{k+1} = \int_{\mathcal{X}} P(y_{k+1}|u_{k+1})P(u_{k+1}|Y_k)du_k$$

and thus by Bayes' formula (6.15) is equal to $P(u_{k+1}|Y_{k+1})$ as required. $\qquad \square$

We now show that an associated recursion is satisfied by the SIR filter $\widehat{\mu}^N$.

Lemma 6.10. *Let $\widehat{\mu}^N$ be the SIR filter given by (6.8), then*

$$\widehat{\mu}_{k+1}^N = ML_{k+1}^\pi S^N P_{k+1}^\pi \widehat{\mu}_k^N$$

for all $k \geq 0$ and $N \geq 1$, where S^N denotes the sampling operator acting on $\mathcal{M}(\mathcal{X} \times \mathcal{X})$.

Proof. By definition, $\widehat{\mu}_k^N = \sum_{n=1}^N w_k^{(n)} \delta_{\widehat{u}_k^{(n)}}$ so that $P_{k+1}^\pi \widehat{\mu}_k^N \in \mathcal{M}(\mathcal{X} \times \mathcal{X})$ with density

$$\sum_{n=1}^N w_k^{(n)} \pi(u_{k+1}|\widehat{u}_k^{(n)}, y_{k+1})\delta(u_k - \widehat{u}_k^{(n)}) \ .$$

Note that a sample $U \sim P_{k+1}^\pi \widehat{\mu}_k^N$ is a pair $(\widehat{u}_{k+1}^{(n)}, u_k^{(n)})$ obtained as follows: first draw a sample $u_k^{(n)}$ from $\{\widehat{u}_k^{(n)}\}_{n=1}^N$ with weights $\{w_k^{(n)}\}_{n=1}^N$ and then draw sample $\widehat{u}_{k+1}^{(n)}$ from $\pi(u_{k+1}|u_k^{(n)}, y_{k+1})$. Thus, by definition of the $\widehat{u}_{k+1}^{(n)}$ sequence we see that $S^N P_{k+1}^\pi \widehat{\mu}_k^N$ has density

$$\frac{1}{N} \sum_{n=1}^N \delta(u_{k+1} - \widehat{u}_{k+1}^{(n)})\delta(u_k - u_k^{(n)}) \ .$$

It follows that $L_{k+1}^{\pi} S^N P_{k+1}^{\pi} \mu_k^N$ has density

$$\sum_{n=1}^{N} Z^{-1} w_{k+1}(\widehat{u}_{k+1}^{(n)}, u_k^{(n)}) \delta(u_{k+1} - \widehat{u}_{k+1}^{(n)}) \delta(u_k - u_k^{(n)})$$

and $M L_{k+1}^{\pi} S^N P_{k+1}^{\pi} \mu_k^N$ has density

$$\sum_{n=1}^{N} Z^{-1} w_{k+1}(\widehat{u}_{k+1}^{(n)}, u_k^{(n)}) \delta(u_{k+1} - \widehat{u}_{k+1}^{(n)}) .$$

Lastly, the normalization factor is given by

$$Z = \int \int_{\mathcal{X} \times \mathcal{X}} \sum_{n=1}^{N} w_{k+1}(\widehat{u}_{k+1}^{(n)}, u_k^{(n)}) \delta(u_{k+1} - \widehat{u}_{k+1}^{(n)}) \delta(u_k - u_k^{(n)}) du_k du_{k+1} = \sum_{n=1}^{N} w_{k+1}(\widehat{u}_{k+1}^{(n)}, u_k^{(n)}) ,$$

so that $Z^{-1} w_{k+1}(\widehat{u}_{k+1}^{(n)}, u_k^{(n)}) = w_{k+1}^{(n)}$ and we obtain the result. $\qquad \square$

In the final step before proving Theorem 6.5, we state some simple properties for the operators appearing in the recursions. Note that these are similar but not (all) immediately implied by the corresponding results for the BPF, Lemma 6.3.

Lemma 6.11. *We have the following simple estimates:*

1. $d(M\nu, M\mu) \leq d(\nu, \mu)$.

2. $d(P_{k+1}^{\pi}\nu, P_{k+1}^{\pi}\mu) \leq d(\nu, \mu)$

3. $\sup_{\nu \in \mathcal{M}(\mathcal{X} \times \mathcal{X})} d(S^N \nu, \nu) \leq N^{-1/2}$

Proof. Let $\tilde{f}(x, y) = f(x)$ and let $g(x, y)$ denote an arbitrary function. Then

$$M\nu(f) - M\mu(f) = \nu(\tilde{f}) - \mu(\tilde{f})$$

The first inequality follows immediately from taking supremum over all $|f| \leq 1$, which is necessarily smaller than the supremum of $\nu(g) - \mu(g)$ over all $|g| \leq 1$.

We also have

$$P_{k+1}^{\pi}\nu(g) - P_{k+1}^{\pi}\mu(g) = \nu(g^{\pi}) - \mu(g^{\pi})$$

where $g^{\pi}(u_k) = \int g(u_{k+1}, u_k) \pi(u_{k+1}|u_k, Y_{k+1}) du_{k+1}$. And since $|g^{\pi}|_{\infty} \leq 1$, the second inequality follows. The third inequality is proven in [25, Lemma 4.7], simply replacing $\mathcal{X}$ with $\mathcal{X} \times \mathcal{X}$. $\qquad \square$

We can now proceed with the main result.

Proof of Theorem 6.5. In the context of Lemma 6.4, take $\mathcal{Z} = \mathcal{X} \times \mathcal{X}$ and $g_{k+1} = f_{k+1}$, it follows that $G_{k+1}\nu = L_{k+1}^{\pi}\nu$. Indeed, for any $\varphi : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ and with $g_{k+1} = Z^{-1} w_{k+1}$ we have

$$G_{k+1}\nu(\varphi) = \frac{\nu(g_{k+1}\varphi)}{\nu(g_{k+1})} = \frac{\nu(w_{k+1}\varphi)}{\nu(w_{k+1})} = L_{k+1}^{\pi}\nu(\varphi) .$$

By Assumption 6.10, we therefore obtain from Lemma 6.4 that

$$d(L_{k+1}^{\pi}\mu, L_{k+1}^{\pi}\nu) \leq (2\kappa^{-2}) d(\mu, \nu)$$

for all $\mu, \nu \in \mathcal{M}(\mathcal{X} \times \mathcal{X})$.

Thus, using the recursions given in Lemmas 6.9, 6.10 and the estimates given in Lemma 6.11, we obtain

$$
\begin{aligned}
d(\widehat{\mu}_{k+1}^N, \mu_{k+1}) &= d(ML_{k+1}^\pi S^N P_{k+1}^\pi \widehat{\mu}_k^N, ML_{k+1}^\pi P_{k+1}^\pi \mu_k) \\
&\leq d(L_{k+1}^\pi S^N P_{k+1}^\pi \widehat{\mu}_k^N, L_{k+1}^\pi P_{k+1}^\pi \mu_k) \\
&\leq 2\kappa^{-2} d(S^N P_{k+1}^\pi \widehat{\mu}_k^N, P_{k+1}^\pi \mu_k) \\
&\leq 2\kappa^{-2} \left(d(S^N P_{k+1}^\pi \widehat{\mu}_k^N, P_{k+1}^\pi \widehat{\mu}_k^N) + d(P_{k+1}^\pi \widehat{\mu}_k^N, P_{k+1}^\pi \mu_k) \right) \\
&\leq 2\kappa^{-2} N^{-1/2} + 2\kappa^{-2} d(\widehat{\mu}_k^N, \mu_k) .
\end{aligned}
$$

and since $\widehat{\mu}_0^N = \mu_0$, we obtain (6.13) by induction. Moreover, since $\mu_k^N = S^N \widehat{\mu}_k^N$

$$
d(\mu_k^N, \mu_k) = d(S^N \widehat{\mu}_k^N, \mu_k) \leq d(S^N \widehat{\mu}_k^N, \widehat{\mu}_k^N) + d(\widehat{\mu}_k^N, \mu_k)
$$

and (6.14) follows. □

The corollary follows immediately.

Proof of Corollary 6.8. For the OPF we have

$$
\pi(u_{k+1}|u_k, y_{k+1}) = P(u_{k+1}|u_k, y_{k+1}) = \frac{P(y_{k+1}|u_{k+1})P(u_{k+1}|u_k)}{P(y_{k+1}|u_k)}
$$

where we have applied Bayes formula in the final equality. But under Assumption 2.1 we have that

$$
P(y_{k+1}|u_k) = Z_S^{-1} \exp\left(-\frac{1}{2}|y_{k+1} - H\psi(u_k)|_S^2 \right) .
$$

Thus we define f_{k+1} by

$$
f_{k+1}(u_{k+1}, u_k) = Z_S \frac{P(y_{k+1}|u_{k+1})P(u_{k+1}|u_k)}{\pi(u_{k+1}|u_k, y_{k+1})} = \exp\left(-\frac{1}{2}|y_{k+1} - H\psi(u_k)|_S^2 \right)
$$

and hence (6.10) holds with $\kappa^{-1} = \exp\left(\max_{0 \leq j \leq K-1} |y_{j+1}|^2 + \sup_v |H\psi(v)|_S^2 \right)$, which, for each Y_k, is finite by Assumption 2.1. The result follows from Theorem 6.5. □

6.3 Gaussianized Optimal Particle Filter

In this section we derive the consistency result for the GOPF.

Theorem 6.12. *Let ν^N be the GOPF defined by (3.12) and let Assumption 2.1 hold. Then there is $\kappa = \kappa(Y_K)$ such that*

$$
d(\nu_K^N, \mu_K) \leq \sum_{k=0}^K (2\kappa^{-2})^k N^{-1/2}
$$

for all $K, N \geq 1$, where $\kappa^{-1} = \exp\left(\max_{0 \leq j \leq K-1} |y_{j+1}|^2 + \sup_v |H\psi(v)|_S^2 \right)$.

For the GOPF, the consistency proof uses the same strategy, but turns out to be much more straightforward. First note that the decomposition of the filtering distribution given in (3.10) gives the recursion formula

$$
\mu_{k+1} = Q_{k+1} K_{k+1} \mu_k \tag{6.16}
$$

where $K_{k+1} : \mathcal{M}(\mathcal{X}) \to \mathcal{M}(\mathcal{X})$ is defined by

$$
K_{k+1}\mu(A) = Z^{-1} \int_A P(y_{k+1}|u_k)\mu(du_k)
$$

for all measurable $A \subset \mathcal{X}$ where Z is the normalization constant, and $Q_{k+1} : \mathcal{M}(\mathcal{X}) \to \mathcal{M}(\mathcal{X})$ is the Markov semigroup with kernel $P(u_{k+1}|u_k, y_{k+1})$.

Moreover, we have the following recursion for the GOPF.

Lemma 6.13. *The GOPF ν_k^N satsfies the recursion*

$$\nu_{k+1}^N = S^N Q_{k+1} K_{k+1} \nu_k^N \tag{6.17}$$

with $\nu_0^N = S^N \mu_0$.

Proof. Let $\nu_k^N = \frac{1}{N} \sum_{n=1}^N \delta_{v_k^{(n)}}$, then $K_{k+1} \in \mathcal{M}(\mathcal{X})$ with density

$$\sum_{n=1}^N Z^{-1} P(y_{k+1}|v_k^{(n)}) \delta(v_k - v_k^{(n)}) \; .$$

The normalization constant is given by

$$Z = \int_{\mathcal{X}} \sum_{n=1}^N P(y_{k+1}|v_k^{(n)}) \delta(v_k - v_k^{(n)}) dv_k = \sum_{n=1}^N P(y_{k+1}|v_k^{(n)})$$

and thus $Z^{-1} P(y_{k+1}|v_k^{(n)}) = w_{k+1}^{(n)}$. We then have $Q_{k+1} L_{k+1} \nu_k^N \in \mathcal{M}(\mathcal{X})$ with density

$$\sum_{n=1}^N w_{k+1}^{(n)} P(v_{k+1}|v_k^{(n)}, y_{k+1}) \; .$$

To draw a sample $v_{k+1}^{(n)}$ from this mixture model, we draw $\tilde{v}_k^{(n)}$ from $\{v_k^{(m)}\}_{m=1}^N$ with weights $\{w_{k+1}^{(m)}\}_{m=1}^N$ and then draw $v_{k+1}^{(n)}$ from $P(v_{k+1}|\tilde{v}_k^{(n)}, y_{k+1})$. It follows that $S^N Q_{k+1} K_{k+1} \nu_k^N = \nu_{k+1}^N$. $\qquad \square$

Proof. If we let

$$g_{k+1}(v_k) := Z_S P(y_{k+1}|v_k) = \exp\left(-\frac{1}{2}|y_{k+1} - H\psi(v_k)|_S^2\right) \; ,$$

then g_{k+1} satisfies the assumptions of Lemma 6.4 with

$$\kappa^{-1} = \exp\left(\max_{0 \le j \le K-1} |y_{j+1}|^2 + \sup_v |H\psi(v)|_S^2\right) .$$

In particular, since $G_{k+1}\nu = K_{k+1}\nu$, it follows from Lemma 6.4 that

$$d(K_{k+1}\mu, K_{k+1}\nu) \le (2\kappa^{-2}) d(\mu, \nu)$$

for all $\mu, \nu \in \mathcal{M}(\mathcal{X})$.

Using the recursions (6.16),(6.17) and the estimates from Lemma 6.3, we see that

$$\begin{aligned}
d(\nu_{k+1}^N, \mu_{k+1}) &= d(S^N Q_{k+1} K_{k+1} \nu_k^N, Q_{k+1} K_{k+1} \mu_k) \\
&\le d(S^N Q_{k+1} K_{k+1} \nu_k^N, Q_{k+1} K_{k+1} \nu_k^N) + d(Q_{k+1} K_{k+1} \nu_k^N, Q_{k+1} K_{k+1} \mu_k) \\
&\le N^{-1/2} + d(K_{k+1} \nu_k^N, K_{k+1} \mu_k) \\
&\le N^{-1/2} + 2\kappa^{-2} d(\nu_k^N, \mu_k^N)
\end{aligned}$$

by induction, we obtain

$$d(\nu_{k+1}^N, \mu_{k+1}^N) \le \sum_{j=0}^k (2\kappa^{-2})^j N^{-1/2} + (2\kappa^{-2})^{k+1} d(\nu_0^N, \mu_0) \; .$$

And the result follows from the fact $d(\nu_0^N, \mu_0) = d(S^N \mu_0, \mu_0) \le N^{-1/2}$. $\qquad \square$

References

[1] Hajime Akashi and Hiromitsu Kumamoto. Random sampling approach to state estimation in switching environments. *Automatica*, 13(4):429–434, 1977.

[2] Rami Atar and Ofer Zeitouni. Exponential stability for nonlinear filtering. In *Annales de l'IHP Probabilités et statistiques*, volume 33, pages 697–725, 1997.

[3] Peter Bauer, Alan Thorpe, and Gilbert Brunet. The quiet revolution of numerical weather prediction. *Nature*, 525(7567):47–55, 2015.

[4] Amarjit Budhiraja and D Ocone. Exponential stability of discrete-time filters for bounded observation noise. *Systems & Control Letters*, 30(4):185–193, 1997.

[5] Gerrit Burgers, Peter Jan van Leeuwen, and Geir Evensen. Analysis scheme in the ensemble Kalman filter. *Monthly weather review*, 126(6):1719–1724, 1998.

[6] Alexandre J Chorin and Xuemin Tu. Implicit sampling for particle filters. *Proceedings of the National Academy of Sciences*, 106(41):17249–17254, 2009.

[7] Dan Crisan and Kari Heine. Stability of the discrete time filter in terms of the tails of noise distributions. *Journal of the London Mathematical Society*, 78(2):441–458, 2008.

[8] Pierre Del Moral, Aline Kurtzmann, and Julian Tugaut. On the stability and the uniform propagation of chaos of extended ensemble Kalman-Bucy filters. *arXiv preprint arXiv:1606.08256*, 2016.

[9] Randal Douc, Gersende Fort, Eric Moulines, and Pierre Priouret. Forgetting the initial distribution for hidden markov models. *Stochastic processes and their applications*, 119(4):1235–1256, 2009.

[10] Arnaud Doucet, Nando De Freitas, and Neil Gordon. An introduction to sequential monte carlo methods. In *Sequential Monte Carlo methods in practice*, pages 3–14. Springer, 2001.

[11] Arnaud Doucet, Simon Godsill, and Christophe Andrieu. On sequential monte carlo sampling methods for Bayesian filtering. *Statistics and computing*, 10(3):197–208, 2000.

[12] Geir Evensen. Using the extended Kalman filter with a multilayer quasi-geostrophic ocean model. *Journal of Geophysical Research: Oceans*, 97(C11):17905–17924, 1992.

[13] Geir Evensen. The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean dynamics*, 53(4):343–367, 2003.

[14] Michael Ghil and Paola Malanotte-Rizzoli. Data assimilation in meteorology and oceanography. *Advances in geophysics*, 33:141–266, 1991.

[15] John Harlim, Andrew J Majda, et al. Catastrophic filter divergence in filtering nonlinear dissipative systems. *Communications in Mathematical Sciences*, 8(1):27–43, 2010.

[16] Peter L Houtekamer and Herschel L Mitchell. Ensemble kalman filtering. *Quarterly Journal of the Royal Meteorological Society*, 131(613):3269–3289, 2005.

[17] David Kelly, Andrew J Majda, and Xin T Tong. Concrete ensemble Kalman filters with rigorous catastrophic filter divergence. *Proceedings of the National Academy of Sciences*, 112(34):10589–10594, 2015.

[18] David Kelly, Eric Vanden-Eijnden, and Jonathan Weare. Implicit resampling in the optimal particle filter. 2016.

[19] DTB Kelly, KJH Law, and Andrew M Stuart. Well-posedness and accuracy of the ensemble Kalman filter in discrete and continuous time. *Nonlinearity*, 27(10):2579, 2014.

[20] ML Kleptsyna and A Yu Veretennikov. On discrete time ergodic filters with wrong initial data. *Probability Theory and Related Fields*, 141(3-4):411–444, 2008.

[21] Hiroshi Kunita. Asymptotic behavior of the nonlinear filtering errors of markov processes. *Journal of Multivariate Analysis*, 1(4):365–393, 1971.

[22] K Law, AM Stuart, KC Zygalakis, et al. Accuracy and stability of the continuous-time 3dvar filter for the Navier-Stokes equation. *Nonlinearity*, 26(8):2193, 2013.

[23] KJH Law, D Sanz-Alonso, A Shukla, and AM Stuart. Filter accuracy for the lorenz 96 model: fixed versus adaptive observation operators. *Physica D: Nonlinear Phenomena*, 325:1–13, 2016.

[24] KJH Law, Abhishek Shukla, and AM Stuart. Analysis of the 3dvar filter for the partially observed lorenz '63 model. *Discrete and Continuous Dynamical Systems A*, 34:1061–1078, 2014.

[25] Kody Law, Andrew Stuart, and Kostas Zygalakis. *Data Assimilation*. Springer, 2015.

[26] Kody JH Law, Hamidou Tembine, and Raul Tempone. Deterministic mean-field ensemble kalman filtering. *arXiv preprint arXiv:1409.0628*, 2014.

[27] Jun S Liu and Rong Chen. Blind deconvolution via sequential imputations. *Journal of the american statistical association*, 90(430):567–576, 1995.

[28] Andrew C Lorenc. Analysis methods for numerical weather prediction. *Quarterly Journal of the Royal Meteorological Society*, 112(474):1177–1194, 1986.

[29] Andrew J Majda and John Harlim. *Filtering complex turbulent systems*. Cambridge University Press, 2012.

[30] Andrew J Majda and Xin T Tong. Robustness and accuracy of finite ensemble kalman filters in large dimensions. *arXiv*, arXiv:1606.09321, 2016.

[31] Jonathan C Mattingly, Andrew M Stuart, and Desmond J Higham. Ergodicity for sdes and approximations: locally lipschitz vector fields and degenerate noise. *Stochastic processes and their applications*, 101(2):185–232, 2002.

[32] Sean P Meyn and Richard L Tweedie. *Markov Chains and Stochastic Stability*. Springer Science & Business Media, 2012.

[33] Alexander JF Moodey, Amos S Lawless, Roland WE Potthast, and Peter Jan Van Leeuwen. Nonlinear error dynamics for cycled data assimilation methods. *Inverse Problems*, 29(2):025002, 2013.

[34] Patrick Rebeschini, Ramon Van Handel, et al. Can local particle filters beat the curse of dimensionality? *The Annals of Applied Probability*, 25(5):2809–2866, 2015.

[35] Daniel Sanz-Alonso and Andrew M Stuart. Long-time asymptotics of the filtering distribution for partially observed chaotic dynamical systems. *SIAM/ASA Journal on Uncertainty Quantification*, 3(1):1200–1220, 2015.

[36] Daniel Sanz-Alonso and Andrew M Stuart. Long-time asymptotics of the filtering distribution for partially observed chaotic dynamical systems. *SIAM/ASA Journal on Uncertainty Quantification*, 3(1):1200–1220, 2015.

[37] Madineh Sedigh-Sarvestani, David J Albers, and Bruce J Gluckman. Data assimilation of glucose dynamics for use in the intensive care unit. In *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 5437–5440. IEEE, 2012.

[38] Madineh Sedigh-Sarvestani, Steven J Schiff, and Bruce J Gluckman. Reconstructing mammalian sleep dynamics with data assimilation. *PLoS Comput Biol*, 8(11):e1002788, 2012.

[39] Chris Snyder. Particle filters, the "optimal" proposal and high-dimensional systems. In *Proceedings of the ECMWF Seminar on Data Assimilation for atmosphere and ocean*, 2011.

[40] Chris Snyder, Thomas Bengtsson, and Mathias Morzfeld. Performance bounds for particle filters using the optimal proposal. *Monthly Weather Review*, 143(11):4750–4761, 2015.

[41] Sebastian Thrun. Particle filters in robotics. In *Proceedings of the Eighteenth conference on Uncertainty in artificial intelligence*, pages 511–518. Morgan Kaufmann Publishers Inc., 2002.

[42] Xin T Tong, Andrew J Majda, and David Kelly. Nonlinear stability and ergodicity of ensemble based Kalman filters. *Nonlinearity*, 29(2):657, 2016.

[43] Xin T Tong, Andrew J Majda, and David Kelly. Nonlinear stability of the ensemble Kalman filter with adaptive covariance inflation. *Communications in Mathematical Sciences*, 14(5), 2016.

[44] Xin Thomson Tong and Ramon Van Handel. Ergodicity and stability of the conditional distributions of nondegenerate Markov chains. *The Annals of Applied Probability*, pages 1495–1540, 2012.

[45] Ramon Van Handel. The stability of conditional Markov processes and Markov chains in random environments. *The Annals of Probability*, pages 1876–1925, 2009.

[46] Peter Jan Van Leeuwen. Particle filtering in geophysical systems. *Monthly Weather Review*, 137(12):4089–4114, 2009.

[47] Peter Jan van Leeuwen. Nonlinear data assimilation in geosciences: an extremely efficient particle filter. *Quarterly Journal of the Royal Meteorological Society*, 136(653):1991–1999, 2010.

[48] Daniel B Work, Sébastien Blandin, Olli-Pekka Tossavainen, Benedetto Piccoli, and Alexandre M Bayen. A traffic model for velocity data assimilation. *Applied Mathematics Research eXpress*, 2010(1):1–35, 2010.

[49] VS Zaritskii, VB Svetnik, and LI Šimelevič. Monte-carlo technique in problems of optimal information processing. *Avtomatika i Telemekhanika*, (12):95–103, 1975.