SIAM J. NUMER. ANAL.
Vol. 54, No. 3, pp. 1813–1839

MULTILEVEL ENSEMBLE KALMAN FILTERING*

HÅKON HOEL†, KODY J. H. LAW‡, AND RAUL TEMPONE§

Abstract. This work embeds a multilevel Monte Carlo sampling strategy into the Monte Carlo step of the ensemble Kalman filter (EnKF) in the setting of finite dimensional signal evolution and noisy discrete-time observations. The signal dynamics is assumed to be governed by a stochastic differential equation (SDE), and a hierarchy of time grids is introduced for multilevel numerical integration of that SDE. The resulting multilevel EnKF is proved to asymptotically outperform EnKF in terms of computational cost versus approximation accuracy. The theoretical results are illustrated numerically.

Key words. Monte Carlo, multilevel, filtering, Kalman filter, ensemble Kalman filter

AMS subject classifications. 65C30, 65Y20, 65M20, 65c35

DOI. 10.1137/15M100955X

1. Introduction.

Filtering refers to the sequential estimation of the state u and/or parameters p of a system through sequential incorporation of online data y. The most complete estimation of the state u_n at time n is given by its probability density conditional on the observations up to the given time $\pi(u_n|y_1,\dots,y_n)$ [28, 3]. For linear Gaussian systems the analytical solution may be given in closed form, via an update formula for the mean and covariance known as the Kalman filter [29]. However, in general there is no closed form solution. One must therefore resort to either algorithms which approximate the probabilistic solution by leveraging ideas from control theory [30, 28], or Monte Carlo methods to approximate the filtering distribution itself [3, 13, 12]. The ensemble Kalman filter (EnKF) [7, 15] combines elements of both approaches. In the linear Gaussian case, it converges to the Kalman filter solution [38], and even in the nonlinear case, under suitable assumptions it converges [35, 34] to what one may argue is the optimal filter among those which incorporate the data linearly [34, 37, 40]. In the case of spatial models approximated on a numerical grid, the state space itself may become very high-dimensional and even the linear solves may become intractable. Therefore, one may be inclined to use the EnKF filter even for linear Gaussian problems in which the solution is intractable despite being given in closed form on paper by the Kalman filter.

For problems which admit hierarchies of approximations with cost inversely proportional to accuracy, it is natural to leverage solutions to less expensive and less accurate approximations in order to accelerate the convergence of the more expensive and more accurate approximations. This idea originated in the iterative solution of numerical approximation of elliptic PDE as early as [16]. In the context of an iterative

*Received by the editors February 20, 2015; accepted for publication (in revised form) April 18, 2016; published electronically June 14, 2016. This research was supported by King Abdullah University of Science and Technology (KAUST). The authors were members of the SRI Center for Uncertainty Quantification at KAUST for much of the research reported.

http://www.siam.org/journals/sinum/54-3/M100955.html

†Department of Mathematics, University of Oslo, Oslo, Norway, and Applied Mathematics and Computational Sciences, KAUST, Thuwal, Saudi Arabia (haakonah@math.uio.no).

‡Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831 (lawkj@ornl.gov). The research of this author was supported by an ORNL LDRD Strategic Hire grant.

§Applied Mathematics and Computational Sciences, KAUST, Thuwal, Saudi Arabia (raul.tempone@kaust.edu.sa).

solution of PDE, the methodology, which may be used both for solution as well as preconditioner, has become known as multigrid—see [23] for a general reference. The same idea may be applied in the context of Monte Carlo approximation of random fields as proposed in [24] and later studied in detail in the context of the SDE in [18] and the PDE in [10]. There has been an explosion of recent activity since [18], including, for example, [19, 11, 26, 6, 39]. Even more recently, work is beginning to emerge extending the same multilevel framework beyond "vanilla" Monte Carlo to its manifestation in the context of Bayesian inference, anywhere that one has a discretization error inversely proportional to cost in the computation of a single sample and the Monte Carlo $\mathcal{O}(M^{-1/2})$ rate of convergence. Examples in the context of Markov chain Monte Carlo appeared in [31, 25]. To the knowledge of the authors there has yet to be an extension of the methodology to the filtering context. As a first step, this work explores the extension of the EnKF to its multilevel implementation, which is naturally referred to as multilevel EnKF (MLEnKF). In the case of linear Gaussian problems, the limiting distribution is the gold-standard Bayesian posterior distribution, while in the non-Gaussian case it is something else (see, e.g., [34, 37, 40] and references therein for further discussion).

The rest of the paper is organized as follows. In section 2 the filtering problem will be introduced, the Kalman filter and EnKF will be reviewed, and the new MLEnKF is introduced for the first time in subsection 2.4. In section 3 it is proven that indeed the MLEnKF inherits almost the same favorable asymptotic "cost-to-ε" as the standard multilevel Monte Carlo (MLMC) for a finite time horizon, and its mean-field limiting distribution is the filtering distribution in the linear and Gaussian case. In section 4 the theory is illustrated with numerical examples of the Ornstein–Uhlenbeck stochastic differential equation and the geometric Brownian motion (GBM). Finally, conclusions and future directions are presented in section 5.

2. Kalman filtering. Here the filtering problem will be introduced in section 2.1, and then the Kalman filter will be introduced for linear Gaussian state-space models in section 2.2. In section 2.3, its Monte Carlo implementation of the Kalman filter will be introduced, that is, the EnKF. In section 2.4, the multilevel implementation is introduced for the first time.

2.1. General setup. Let $(\Omega, \mathcal{E}, \mathbb{P})$ be a complete probability space, where Ω is the set of all possible outcomes, $\mathcal{E}$ is the sigma algebra generated by Ω and $\mathbb{P}$ is the associated probability measure. Consider the general stochastic signal evolution for the random variables $u_n : \Omega \to \mathbb{R}^d$, with $d < \infty$,

$$u_{n+1} = \Psi(u_n), \tag{2.1}$$

for $n = 0, 1, \ldots, N-1$. Given the history of a noisy signal observation

$$y_n = Hu_n + \eta_n,$$

where $H \in \mathbb{R}^{m\times d}$ and η_n are independent and identically distributed (i.i.d.) with $\eta_1 \sim N(0, \Gamma), \Gamma \in \mathbb{R}^{m\times m}$ symmetric positive definite, the objective is to track the signal u_n given observations $Y_n = Y_n^{\text{obs}}$ with $Y_n = (y_1, y_2, \ldots, y_n)$ and $Y_n^{\text{obs}} = (y_1^{\text{obs}}, y_2^{\text{obs}}, \ldots, y_n^{\text{obs}})$ the given observed realization of Y_n. In other words, the aim is to approximate the random variable $u_n | (Y_n = Y_n^{\text{obs}})$. Notice that under the given assumptions this is a hidden Markov model and the density of the random variable

we seek to approximate admits the following sequential structure:

$$\pi_{u_n}(u|Y_n = Y_n^{\rm obs}) = \frac{\pi_{y_n}(y_n^{obs}|u_n = u)\pi_{u_n}(u|Y_{n-1} = Y_{n-1}^{\rm obs})}{\pi_{y_n}(y_n^{\rm obs}|Y_{n-1} = Y_{n-1}^{\rm obs})}, \tag{2.2}$$

$$\pi_{u_n}(u|Y_{n-1} = Y_{n-1}^{\rm obs}) = \int_{\mathbb{R}^d} \pi_{u_n}(u|u_{n-1} = v)\pi_{u_{n-1}}(v|Y_{n-1} = Y_{n-1}^{\rm obs})\,{\rm d}v,$$

$$\pi_{y_n}(y_n^{obs}|Y_{n-1} = Y_{n-1}^{\rm obs}) = \int_{\mathbb{R}^d} \pi_{y_n}(y_n^{\rm obs}|u_n = u)\pi_{u_n}(u|Y_{n-1} = Y_{n-1}^{\rm obs})\,{\rm d}u.$$

Here, $\pi_{X_1}(x_1|X_2 = x_2)$ denotes the marginal probability density of X_1 conditional that $X_2 = x_2$.

It will be assumed that $\Psi(\cdot)$ cannot be evaluated exactly, but rather only approximately, and that there exists a hierarchy of accuracies at which it can be evaluated each with its associated cost. The explicit dependence on ω will be suppressed where confusion is not possible. In particular, we will be concerned herein with the case in which $u_{n+1} = \Psi(u_n) := u(1; u(0) = u_n)$ is given by the evolution at $t = 1$ of the SDE

$$du(t) = a(u(t))dt + b(u(t))dW(t+n), \qquad t \in (0,1), \tag{2.3}$$

given the prescribed initial condition $u(0) = u_n$, where $a : \mathbb{R}^d \to \mathbb{R}^d$, $b : \mathbb{R}^d \to \mathbb{R}^{d\times r}$, and $W : \Omega \times [0,\infty) \to \mathbb{R}^r$ is an r-dimensional Wiener process. Furthermore, a and b will satisfy the following conditions:

$$|a(u) - a(v)| + |b(u) - b(v)| \le c|u - v| \ \forall u, v \in \mathbb{R}^d \text{ and some } c > 0. \tag{2.4}$$

This fits into the framework of (2.1). Notice that the randomness comes from the initial condition and the Wiener process W only and not the coefficients; however, it could in principle come also from the latter. Indeed if the analytical solution is unknown, then the system above must be approximated, leading to the hierarchy of approximations to $\Psi(\cdot)$. In particular, denote by $\{\Psi^\ell\}_{\ell=0}^\infty$ a hierarchy of approximations to the solution $\Psi := \Psi^\infty$ of (2.3). First some assumptions must be made.

Assumption 1. For every $p \ge 2$, the solution operators $\{\Psi^\ell\}_{\ell=0}^\infty$ satisfy the following conditions for some $0 < c_\Psi < \infty$ depending on Ψ:

(i) $\|\Psi^\ell(u) - \Psi^\ell(v)\|_p < c_\Psi \|u - v\|_p$,

(ii) $\|\Psi^\ell(u)\|_p^p \le c_\Psi(1 + \|u\|_p^p)$,

where the following notation is introduced $\|u\|_p := \mathbb{E}[|u|^p]^{1/p}$.

For many numerical solvers the assumptions can be verified by application of Gronwall's inequality; cf. [20, 9]. For notational simplicity, we consider the particular case in which (2.3) is *autonomous*, such that the coefficients on the right-hand side do not depend explicitly on time. Note that the results easily extend to the nonautonomous case, provided the given assumptions on Ψ are uniform with respect to $\{\Psi_n\}_{n=1}^N$. The specialization is merely for notational convenience.

2.2. Linear Gaussian case. Consider the linear instance of (2.1), in which

$$\Psi(u_n) = Au_n + \xi_n, \tag{2.5}$$

where $A \in \mathbb{R}^{d\times d}$, and ξ_n are i.i.d. normal random variables with $\xi_1 \sim N(0, \Sigma), \Sigma \in \mathbb{R}^{d\times d}$ symmetric positive definite. This case arises when the coefficients of (2.3) are given by

$$a(u) = a_1 u + a_0 \quad \text{and} \quad b(u) = b_0. \tag{2.6}$$

Again we suppress the possible time-dependence of the random maps Ψ and matrices A and Σ just to simplify notation. For this class of problems, given a Gaussian initial condition, the filtering distribution (2.2) is known to be Gaussian and is therefore defined uniquely by its mean and covariance. Kalman filtering provides a two-step iterative procedure for computing the mean and covariance of $\widehat{v}_{n+1} := u_{n+1}|(Y_{n+1} = Y_{n+1}^{\text{obs}})$ given $\widehat{v}_n := u_n|(Y_n = Y_n^{\text{obs}}) \sim N(\widehat{m}_n, \widehat{C}_n)$, where

$$\widehat{m}_n := \mathbb{E}\left[u_n|Y_n = Y_n^{\text{obs}}\right] \quad \text{and} \quad \widehat{C}_n := \mathbb{E}\left[(u_n - \widehat{m}_n)(u_n - \widehat{m}_n)^{\mathsf{T}}|Y_n = Y_n^{\text{obs}}\right].$$

The classical Kalman filter consists of a two-step formula which maps the distribution of $\widehat{v}_n$ to the distribution of $\widehat{v}_{n+1}$. In particular, the map from $(\widehat{m}_n, \widehat{C}_n)$ to $(\widehat{m}_{n+1}, \widehat{C}_{n+1})$ is described by the following two prediction equations and two update equations:

$$\begin{aligned} m_{n+1} &= A\widehat{m}_n, & C_{n+1} &= A\widehat{C}_n A^{\mathsf{T}} + \Sigma, \\ \widehat{m}_{n+1} &= (I - K_{n+1}H)m_{n+1} + K_{n+1}y_{n+1}^{\text{obs}}, & \widehat{C}_{n+1} &= (I - K_{n+1}H)C_{n+1}, \end{aligned}$$

where the *Kalman gain* K_{n+1} is given by

$$K_{n+1} = C_{n+1}H^{\mathsf{T}}S_{n+1}^{-1}, \qquad S_{n+1} = \Gamma + HC_{n+1}H^{\mathsf{T}}.$$

Remark 1. When observations are in a lower dimension than the true signal, the iterations are more efficiently computed by introducing $d_{n+1} = y_{n+1}^{\text{obs}} - Hm_{n+1}$ and updating the mean as follows:

$$\widehat{m}_{n+1} = m_{n+1} + K_{n+1}d_{n+1}.$$

It is nonetheless instructive to observe the alternative form presented above, as it is the properties of the operators $I - K_{n+1}H$ which are responsible for stability of the algorithm [33].

2.3. EnKF. EnKF uses an ensemble of particles to estimate means and covariance matrices appearing in the Kalman filter; however, the framework can be generalized to non-Gaussian models. Let $v_{n,i}$ and $\widehat{v}_{n,i}$, respectively, denote the prediction and update of the ith particle at simulation time $t_n = n$. One EnKF two-step transition consists not of the propagation of a mean and covariance as in the original Kalman filter but instead the propagation of an ensemble $\{\widehat{v}_{n,i}\}_{i=1}^M \mapsto \{\widehat{v}_{n+1,i}\}_{i=1}^M$.[1] This procedure consists nonetheless in the prediction and update steps. In the prediction step, M particle paths are computed over one interval, i.e.,

$$v_{n+1}(\omega_i) = \Psi(\widehat{v}_n(\omega_i), \omega_i) \tag{2.7}$$

for $i = 1, \ldots, M$, where $v_n(\omega_i) := v_{n,i}$ denotes a realization corresponding to the event sample ω_i of the random variable $v_n : \Omega \to \mathbb{R}^d$, and $\Psi(\cdot, \omega_i)$ signifies the corresponding realization of the map for a given initial condition. Indeed the notation for random variable realizations, e.g., $\xi_{n,i}$ and $\xi_n(\omega_i)$, will be used interchangeably where confusion is not possible. The impetus for introduction of the latter notation

[1] Due to the implicit linear and Gaussian assumptions underlying the formulation, one may determine that it is reasonable to summarize the ensemble in its sample mean and covariance and indeed this is often done. In this case, one may construct a Gaussian from the empirical statistics and resample from that.

will become apparent in the next section. For this presentation it suffices to assume a single infinite precision map; however, there indeed may also be numerical approximation errors, i.e., Ψ^L may be used in place of Ψ for some satisfactory resolution L. The prediction step is completed by using the particle paths to compute sample mean and covariance,

$$m_{n+1}^{\mathrm{MC}} = E_M[v_{n+1}],$$
$$C_{n+1}^{\mathrm{MC}} = \mathrm{Cov}_M[v_{n+1}],$$

where the following notation is introduced:

$$E_M[v] := \frac{1}{M}\sum_{i=1}^{M} v(\omega_i), \tag{2.8}$$

$$\mathrm{Cov}_M[u,v] := E_M[uv^{\mathsf{T}}] - E_M[u]\big(E_M[v]\big)^{\mathsf{T}}, \tag{2.9}$$

as well as the shorthand $\mathrm{Cov}_M[u] := \mathrm{Cov}_M[u,u]$.

The update step consists of computing (1) auxillary matrices

$$S_{n+1}^{\mathrm{MC}} = HC_{n+1}^{\mathrm{MC}}H^{\mathsf{T}} + \Gamma \text{ and } K_{n+1}^{\mathrm{MC}} = C_{n+1}^{\mathrm{MC}}H^{\mathsf{T}}(S_{n+1}^{\mathrm{MC}})^{-1}$$

and (2) measurement corrected particle paths for $i = 1, 2, \ldots, M$,

$$\tilde{y}_{n+1,i} = y_{n+1}^{\mathrm{obs}} + \eta_{n+1,i},$$
$$\widehat{v}_{n+1,i} = (I - K_{n+1}^{\mathrm{MC}}H)v_{n+1,i} + K_{n+1}^{\mathrm{MC}}\tilde{y}_{n+1,i},$$

where $\{\eta_{n+1,i}\}_{i=1}^{M}$ are i.i.d. with $\eta_{n+1,1} \sim N(0,\Gamma)$. This last procedure may appear somewhat ad hoc. Indeed it was originally introduced in [7] to correct the statistical error induced in its absence in implementations following the original formulation of the in [14]. It has become known as the perturbed observation implementation. Due to the form of the update, all ensemble members are correlated to one another after the first update. So, the ensemble is no longer Gaussian after the first update. The measurement corrected sample mean and covariance, which need not be computed, would be given by

$$\widehat{m}_{n+1}^{\mathrm{MC}} = E_M[\widehat{v}_{n+1}],$$
$$\widehat{C}_{n+1}^{\mathrm{MC}} = \mathrm{Cov}_M[\widehat{v}_{n+1}].$$

The sample empirical distribution is defined by

$$\mu_n^{\mathrm{MC}} = \frac{1}{M}\sum_{i=1}^{M}\delta_{v_n(\omega_i)}, \tag{2.10}$$

and, for $\varphi : \mathbb{R}^d \to \mathbb{R}$, the following shorthand notation is introduced: $\mu_n^{\mathrm{MC}}(\varphi) = \int \varphi d\mu_n^{\mathrm{MC}} = E_M[v_n]$. It was shown in [38, 35] that if Ψ is of the form (2.5) and $\mathbb{E}|v_0|^p < \infty$ for all $p \geq 2$, then for all Lipschitz $\varphi : \mathbb{R}^d \to \mathbb{R}$ and all $p \geq 2$,

$$\big(\mathbb{E}|\mu_n^{\mathrm{MC}}(\varphi) - \mu_n(\varphi)|^p\big)^{1/p} \lesssim M^{-1/2}, \tag{2.11}$$

where μ_n is the filtering distribution. The notation $f(M) \lesssim g(M)$ here is used to denote $f(M) = \mathcal{O}(g(M))$.

2.4. MEnKF. MLEnKF computes particle paths on a hierarchy of accuracy levels, in this case given by increasing refinement of the temporal discretization. Let v_n^ℓ, $\widehat{v}_n^\ell$, respectively, denote the prediction and update of a particle on solution level ℓ at simulation time t_n. A solution on level ℓ is computed by the numerical integrator $v_{n+1}^\ell = \Psi^\ell(\widehat{v}_n^\ell)$. Furthermore, let the difference operator for level ℓ be given by

$$\Delta_\ell v_n(\omega) := \begin{cases} v_n^0(\omega) & \text{if } \ell = 0, \\ v_n^\ell(\omega) - v_n^{\ell-1}(\omega) & \text{else if } \ell > 0. \end{cases} \tag{2.12}$$

Then the transition from approximation of the distribution of $\widehat{v}_n$ to the distribution of $\widehat{v}_{n+1}$ in the MLEnKF framework consists of the predict/update step of generating *pairwise coupled* particle realizations on a set of levels $\ell = 0, 1, \ldots, L$. However, it is important to note that here one has correlation between pairs and also between levels due to the update, unlike the standard MLMC in which one has i.i.d. pairs. This point will be very important, and we return to it in the following section.

Similarly to the standard EnKF, the MLEnKF transition is between *multilevel* ensembles $\{[\widehat{v}_n^\ell(\omega_{\ell,i}), \widehat{v}_n^{\ell-1}(\omega_{\ell,i})]_{i=1}^{M_\ell}\}_{\ell=0}^L \mapsto \{[\widehat{v}_{n+1}^\ell(\omega_{\ell,i}), \widehat{v}_{n+1}^{\ell-1}(\omega_{\ell,i})]_{i=1}^{M_\ell}\}_{\ell=0}^L$ with the convention that $\widehat{v}_k^{-1} := 0$ for all k for ease of notation. This consists, as for EnKF, of the predict and update steps. In the predict step, particle paths are first computed on a hierarchy of levels. That is, the particle paths are computed one step forward by

$$\begin{aligned} v_{n+1}^{\ell-1}(\omega_{\ell,i}) &= \Psi^{\ell-1}(\widehat{v}_n^{\ell-1}(\omega_{\ell,i}), \omega_{\ell,i}), \\ v_{n+1}^{\ell}(\omega_{\ell,i}) &= \Psi^{\ell}(\widehat{v}_n^{\ell}(\omega_{\ell,i}), \omega_{\ell,i}) \end{aligned} \tag{2.13}$$

for the levels $\ell = 0, 1, \ldots, L$ and level particles $i = 1, 2, \ldots, M_\ell$ (where for convenience we introduce the convention that $v^{-1} := 0$). Here the introduction of noise in the second argument of the Ψ^ℓ is correlated only within pairs and is otherwise independent. Thereafter, sample mean and covariance matrices are computed as a sum of sample moments over all levels:

$$\begin{aligned} m_{n+1}^{\mathrm{ML}} &= \sum_{\ell=0}^{L} E_{M_\ell}[\Delta_\ell v_{n+1}(\omega_{\ell,\cdot})], \\ C_{n+1}^{\mathrm{ML}} &= \sum_{\ell=0}^{L} \mathrm{Cov}_{M_\ell}[v_{n+1}^\ell(\omega_{\ell,\cdot})] - \mathrm{Cov}_{M_\ell}[v_{n+1}^{\ell-1}(\omega_{\ell,\cdot})], \end{aligned}$$

where we recall the sample moment notation (2.8) and (2.9).

It is necessary for stability of the algorithm that the sample covariance appearing in the denominator of the gain is positive semidefinite, a condition which is *not* guaranteed for multilevel estimators. This will therefore be *imposed* in the algorithm. It would be of independent interest to devise multilevel estimators which preserve positivity without such an imposition. Let

$$C_n^{\mathrm{ML}} = \sum_{k=1}^{d} \lambda_k q_k q_k^{\mathsf{T}}$$

denote the eigenvalue decomposition of the symmetric multilevel covariance. Notice that the condition $\min_k(\lambda_k) \geq 0$ may not hold. Define

$$\tilde{C}_n^{\mathrm{ML}} = \sum_{k=1;\lambda_k>0}^{d} \lambda_k q_k q_k^{\mathsf{T}}. \tag{2.14}$$

It is worth noting that this is not the only way to do this, and it may be possible to use a less invasive and/or less expensive method to guarantee nonnegativity of the covariance. For example, banding [5], shrinkage [36], thresholding [4], or localization [1] are some prospective alternatives. In particular, it will be necessary to consider such alternatives as the dimension grows and the cost of factorizing C_n^{ML} becomes a dominant consideration, but this is outside the scope of the present work. In the update step the multilevel Kalman gain is defined as follows:

$$K_{n+1}^{\mathrm{ML}} = C_{n+1}^{\mathrm{ML}} H^{\mathsf{T}} (S_{n+1}^{\mathrm{ML}})^{-1}, \text{ where } S_{n+1}^{\mathrm{ML}} = H \tilde{C}_{n+1}^{\mathrm{ML}} H^{\mathsf{T}} + \Gamma. \tag{2.15}$$

Next, all particle paths are corrected according to measurements and perturbed observations are added:

$$\begin{aligned} \tilde{y}_{n+1,i}^{\ell} &= y_{n+1}^{\mathrm{obs}} + \eta_{n+1,i}^{\ell}, \\ \widehat{v}_{n+1}^{\ell-1}(\omega_{i,\ell}) &= (I - K_{n+1}^{\mathrm{ML}} H) v_{n+1}^{\ell-1}(\omega_{i,\ell}) + K_{n+1}^{\mathrm{ML}} \tilde{y}_{n+1,i}^{\ell}, \\ \widehat{v}_{n+1}^{\ell}(\omega_{i,\ell}) &= (I - K_{n+1}^{\mathrm{ML}} H) v_{n+1}^{\ell}(\omega_{i,\ell}) + K_{n+1}^{\mathrm{ML}} \tilde{y}_{n+1,i}^{\ell}, \end{aligned} \tag{2.16}$$

where $\{\eta_{n+1,i}^{\ell}\}_{i=1}^{M_\ell}$ are i.i.d. with $\eta_{n+1,1}^{\{0\}} \sim N(0,\Gamma)$. It is in this step precisely that the pairs all become correlated with one another and the situation becomes significantly more complex than the i.i.d. case. After the first update, this correlation propagates forward through (2.13) to the next observation time via this ensemble. This is the conclusion of the update step of the MLEnKF, and this multilevel ensemble is subsequently propagated forward to the next prediction time via (2.13).

The *multilevel* sample mean and covariance (in the case that (2.14) has not modified the covariance, i.e., it has all nonnegative eigenvalues without truncation) of this multilevel ensemble are given by

$$\begin{aligned} \widehat{m}_{n+1}^{\mathrm{ML}} &= \sum_{\ell=0}^{L} E_{M_\ell}[\Delta_\ell \widehat{v}_{n+1}(\omega_{\ell,\cdot})] \\ &= (I - K_{n+1}^{\mathrm{ML}} H) m_{n+1}^{\mathrm{ML}} + K_{n+1}^{\mathrm{ML}} \left[E_{M_0}[\tilde{y}_{n+1,\cdot}^{0} - y_{n+1}^{\mathrm{obs}}] + y_{n+1}^{\mathrm{obs}} \right], \end{aligned} \tag{2.17}$$

$$\begin{aligned} \widehat{C}_{n+1}^{\mathrm{ML}} &= (I - K_{n+1}^{\mathrm{ML}} H) C_{n+1}^{\mathrm{ML}} (I - K_{n+1}^{\mathrm{ML}} H)^{\mathsf{T}} + K_{n+1}^{\mathrm{ML}} \mathrm{Cov}_{M_0}[\tilde{y}_{n+1,\cdot}^{0} - y_{n+1}^{\mathrm{obs}}] {K_{n+1}^{\mathrm{ML}}}^{\mathsf{T}} \\ &= (I - K_{n+1}^{\mathrm{ML}} H) C_{n+1}^{\mathrm{ML}} + K_{n+1}^{\mathrm{ML}} \left[\mathrm{Cov}_{M_0}[\tilde{y}_{n+1,\cdot}^{0} - y_{n+1}^{\mathrm{obs}}] - \Gamma \right] {K_{n+1}^{\mathrm{ML}}}^{\mathsf{T}}. \end{aligned} \tag{2.18}$$

The second term appearing in each case is unbiased. For computing general quantities of interest, it is instructive to introduce the empirical measure of the multilevel ensemble $\{[\widehat{v}_n^{\ell}(\omega_{\ell,i}), \widehat{v}_n^{\ell-1}(\omega_{\ell,i})]_{i=1}^{M_\ell}\}_{\ell=0}^{L}$, i.e.,[2]

$$\mu_n^{\mathrm{ML}} = \frac{1}{M_0} \sum_{i=1}^{M_0} \delta_{\widehat{v}_n^0(\omega_{0,i})} + \sum_{\ell=1}^{L} \frac{1}{M_\ell} \sum_{i=1}^{M_\ell} (\delta_{\widehat{v}_n^{\ell}(\omega_{\ell,i})} - \delta_{\widehat{v}_n^{\ell-1}(\omega_{\ell,i})}). \tag{2.19}$$

Then, the following shorthand notation for multilevel sample averages can be introduced. For any $\varphi : \mathbb{R}^d \to \mathbb{R}$, let

$$\mu_n^{\mathrm{ML}}(\varphi) := \int \varphi d\mu_n^{\mathrm{ML}} = \sum_{\ell=0}^{L} \frac{1}{M_\ell} \sum_{i=1}^{M_\ell} \varphi(\widehat{v}_n^{\ell}(\omega_{\ell,i})) - \varphi(\widehat{v}_n^{\ell-1}(\omega_{\ell,i})).$$

[2]This may be done similary for the predicting distributions, but the updated distributions will be our primary interest.

2.5. Nonlinear Kalman filtering. It will be useful to introduce the limiting process, in the case of the nonlinear non-Gaussian forward model (2.1), i.e., nonlinear (2.3). The following nonlinear Markov process defines the mean-field EnKF [34]:

$$\text{Prediction} \begin{cases} v_{n+1} &= \Psi(\widehat{v}_n), \\ m_{n+1} &= \mathbb{E}[v_{n+1}], \\ C_{n+1} &= \mathbb{E}[(v_{n+1} - m_{n+1}) \otimes (v_{n+1} - m_{n+1})], \end{cases} \tag{2.20}$$

$$\text{Update} \begin{cases} S_{n+1} &= HC_{n+1}H^{\mathsf{T}} + \Gamma, \\ K_{n+1} &= C_{n+1}H^{\mathsf{T}}S_{n+1}^{-1}, \\ \tilde{y}_{n+1} &= y_{n+1}^{\text{obs}} + \tilde{\eta}_{n+1}, \\ \widehat{v}_{n+1} &= (I - K_{n+1}H)v_{n+1} + K_{n+1}\tilde{y}_{n+1}. \end{cases} \tag{2.21}$$

Here $\{\tilde{\eta}_n\}_{n=1}^N$ are i.i.d. draws from $N(0,\Gamma)$. The expectations appearing above in (2.20) are with respect to the random variable v_{n+1}, which depends upon the randomness from the initial condition $v_0 = u_0$, the maps Ψ, and $\tilde{\eta}_0, \ldots, \tilde{\eta}_n$. The observed value y^{obs} is considered fixed and is not averaged over. It is easy to verify that in the linear Gaussian case of section 2.2, the mean and variance of the above process correspond to the mean and variance of the filtering distribution. Furthermore, it was shown in [38, 35] that the single-level EnKF converges to the Kalman filtering distribution with the standard rate $\mathcal{O}(M^{-1/2})$ in this case, as stated formally in (2.11). It was furthermore shown in [35] and [34] that for nonlinear Gaussian state-space models and fully non-Gaussian models (2.1), respectively, the *same convergence property holds*, with the measure corresponding to v_n in (2.20) and (2.21) replacing μ_n in (2.11), as long as the model satisfies a Lipschitz criterion as in Assumption 1. In this work, the aim is to show that the MLEnKF converges as well, and with a cost-to-ε which is strictly smaller than its single-level EnKF counterpart. The true filtering distribution of $u_n|(Y_n = Y_n^{\text{obs}})$ will not appear in the remainder of this work, and the variable v_n will correspond to the solution of the above system (noting that the two are equivalent in the linear Gaussian case).

3. Theoretical results. The approximation error and computational cost of approximating the true filtering distribution by MLEnKF when given a sequence of observations $y_1, y_2, \ldots, y_n$ will be studied in this section. The notation $|\cdot|$ will be used for standard Euclidean norm (and the induced matrix norm) and the covariance matrix of random variables $Z, X \in \mathbb{R}^d$ will be denoted

$$\text{Cov}[Z,X] := \mathbb{E}\left[(Z - \mathbb{E}[Z])(X - \mathbb{E}[X])^{\mathsf{T}}\right]$$

with the shorthand $\text{Cov}[Z] = \text{Cov}[Z,Z]$. Before stating the main approximation theorem, it will be useful to present the basic assumptions that will be used throughout and the corresponding standard MLMC approximation results for i.i.d. samples, as well as a slight variant which will be useful in what follows.

Assumption 2. Consider the d-dimensional SDE (2.3) with initial data $u_0 \in \cup_{p\in\mathbb{N}} L^p(\Omega)$. For the hierarchy of solution operators defined in section 2, let Ψ^ℓ denote a numerical solver using a uniform time step $\Delta t^\ell = 1/N_\ell$ with $N_\ell/N_{\ell-1} \geq \widehat{N} > 1$ for $\ell = 0, 1, \ldots$. Let $\mathcal{F}$ denote the set of functions $\varphi : \mathbb{R}^d \to \mathbb{R}$ which, for all $\ell \geq 0$ and all $u, v \in \cup_{p\in\mathbb{N}} L^p(\Omega)$, and a given set of constants $\alpha, \beta, \gamma > 0$ with $\alpha \geq \min(\beta,\gamma)/2$, fulfill

(i) $\left|\mathbb{E}\left[\varphi(\Psi^\ell(u)) - \varphi(\Psi(v))\right]\right| \lesssim N_\ell^{-\alpha}$ and $|\mathbb{E}[\varphi(u) - \varphi(v)]| \lesssim N_\ell^{-\alpha}$, provided $|\mathbb{E}[u - v]| \lesssim N_\ell^{-\alpha}$;

(ii) $\|\varphi(\Psi^\ell(v)) - \varphi(\Psi^{\ell-1}(v))\|_p \lesssim N_\ell^{-\beta/2}$ for all $p \geq 2$;
(iii) $\mathcal{C}_\ell := \text{Cost}\left(\Psi^\ell(v)\right) \lesssim N_\ell^\gamma$,

where, as stated above, the notation $f(M) \lesssim g(M)$ is used to denote $f(M) = \mathcal{O}(g(M))$. Assume further that all monomials of degree less than or equal to 2 are contained in $\mathcal{F}$.

Remark 2. An implication of the above condition (ii) is that condition (i) holds with $\alpha = \beta/2$. However, for many numerical schemes, there are settings where it is possible to achieve rates $\alpha > \beta/2$ (implementationally, this may yield savings in the computational cost). The literature [32, Theorem 14.5.2] and [20, Chapter 7] provide sufficient regularity conditions on the SDE problem and φ for the Euler–Maruyama method to achieve the rate exponents $\alpha = 1$ and $\beta = 1$ and for the Milstein method to achieve $\alpha = 1$ and $\beta = 2$.

We will now state the main theorem of this paper. It gives an upper bound for the computational cost of achieving a sought accuracy in L^p-norm when using the MLEnKF method to approximate the expectation of an observable. The theorem may be considered an extension to the data assimilation setting of earlier "one-step" cost versus error results in MLMC; cf. [18, Theorem 3.1] and [10, Theorem 1]. To reduce the number of repetitions in the below proofs we notice once and for all that the process itself is in L^p by Assumption 1, hence the realization giving rise to the observations u_n and the observations themselves y_n are as well, for $n = 1, 2, \ldots, N$. It follows from this and the finite norm of K_n [35, 34, 38] that the elements $\widehat{\bar{v}}_n$ and $\bar{v}_n$ given by (2.20) and (2.21) are also in L^p for $n = 1, 2, \ldots, N$. It will be assumed that the update comes at a marginal cost with respect to the prediction. This may be the case for a complicated forward solution with small error tolerance, large ensemble, and comparably modest dimension d.

DEFINITION 3.1. *A function $\varphi : \mathbb{R}^d \to \mathbb{R}$ is said to be locally Lipschitz continuous with at most polynomial growth at infinity provided that there exist positive scalars $\nu, C_\varphi < \infty$ such that*

$$|\varphi(x) - \varphi(y)| \leq C_\varphi |x - y|(1 + |x|^\nu + |y|^\nu) \qquad \forall x, y \in \mathbb{R}^d. \tag{3.1}$$

The notation $f(M) \eqsim g(M)$ will be used to indicate that there exist constants $\tilde{c}_1, \tilde{c}_2 > 0$ such that $\tilde{c}_1 g(M) \leq f(M) \leq \tilde{c}_2 g(M)$.

THEOREM 3.2 (MLEnKF accuracy versus cost). *Suppose Assumptions* 1 *and* 2 *hold. For a given $\varepsilon > 0$, let L and $\{M_\ell\}_{\ell=0}^L$ be defined under the constraints $L \eqsim \log(\varepsilon^{-1})$ and*

$$M_\ell \eqsim \left\lceil N_\ell^{-\frac{\beta+2\gamma}{3}} \begin{cases} N_L^{2\alpha} & \text{if} \quad \beta > \gamma \\ L^2 N_L^{2\alpha} & \text{if} \quad \beta = \gamma \\ N_L^{2\alpha + \frac{2}{3}(\gamma-\beta)} & \text{if} \quad \beta < \gamma \end{cases} \right\rceil . \tag{3.2}$$

Then for all functions $\varphi \in \mathcal{F}$ that are locally Lipschitz continuous with at most polynomial growth at infinity (cf. Definition 3.1*), we have that*

$$\|\mu_n^{\text{ML}}(\varphi) - \mu_n(\varphi)\|_p \lesssim |\log(\varepsilon)|^n \varepsilon, \tag{3.3}$$

where μ_n^{ML} is the multilevel empirical measure defined in (2.19)*, where the samples are given by the multilevel predict* (2.13) *and update* (2.16) *formulae, approximating the time $t_n = n$ mean-field EnKF distribution μ_n (the filtering distribution*

$\mu_n = N(m_n, C_n)$ in the linear Gaussian case). And the computational cost of the MLEnKF estimator over the time sequence satisfies

$$\text{Cost (MLEnKF)} \lesssim \begin{cases} \varepsilon^{-2} & \text{if} \quad \beta > \gamma \\ \varepsilon^{-2} \left|\log(\varepsilon)\right|^3 & \text{if} \quad \beta = \gamma \\ \varepsilon^{-\left(2+\frac{\gamma-\beta}{\alpha}\right)} & \text{if} \quad \beta < \gamma. \end{cases} \tag{3.4}$$

Remark 3. The growth in error factor $|\log(\varepsilon)|^n$ in (3.3) is due to a propagation of perturbed observation errors of the MLEnKF estimator that has been conservatively bounded by the triangle inequality in (3.28). In our numerical tests, however, we do *not* observe the error growth factor, and therefore we believe it might be possible to eliminate this factor by sharper theoretical bounds.

The proof of Theorem 3.2 follows roughly along the same lines as that of [35], however with more notation and longer calculations due to the multilevel aspect. The proof also has connections to the work [6], in which an MLMC method is developed for estimation of higher order central moments.

It will be convenient to introduce the mean-field limiting multilevel ensemble $\{[\overline{v}_n^\ell(\omega_{\ell,i}), \overline{v}_n^{\ell-1}(\omega_{\ell,i})]_{i=1}^{M_\ell}\}_{\ell=0}^{L}$ [35, 34, 38], which evolves according to the same equations *with the same realizations of noise* except the covariance C_n, hence the Kalman gain K_n, is given by limiting formulae in (2.20) and (2.21). That is, the intralevel pairs of ensemble members $(\bar{v}_n^\ell(\omega_{\ell,i}), \bar{v}_n^{\ell-1}(\omega_{\ell,i}))$ are i.i.d. over index i, and they are independent between levels. An ensemble member $(\hat{\bar{v}}_n^\ell(\omega_{\ell,i}), \hat{\bar{v}}_n^{\ell-1}(\omega_{\ell,i}))$ maps to $(\bar{v}_{n+1}^\ell(\omega_{\ell,i}), \bar{v}_{n+1}^{\ell-1}(\omega_{\ell,i}))$ as in (2.13). Then $(\hat{\bar{v}}_{n+1}^\ell(\omega_{\ell,i}), \hat{\bar{v}}_{n+1}^{\ell-1}(\omega_{\ell,i}))$ is obtained as in (2.16), except with K_{n+1} from (2.21) replacing K_{n+1}^{ML} in (2.16). The noise realizations $\{\omega_{\ell,i}\}$ are assumed to be the same as the EnKF ensemble member $(v_n^\ell(\omega_{\ell,i}), v_n^{\ell-1}(\omega_{\ell,i}))$. The sole difference is that the limiting ensemble is independent between levels and the pairs within a level are i.i.d. *This is because the covariance and gain come from the infinite limiting system* (2.20) *and* (2.21). The only correlations are between $\bar{v}_n^\ell(\omega_{\ell,i})$ and $\bar{v}_n^{\ell-1}(\omega_{\ell,i})$, due to the $\omega_{\ell,i}$. Hence there is no multiplicative propagation of correlations within a level or between levels. This crucial fact allows us to (a) on the one hand extend standard multilevel theory for i.i.d. draws over multiple updates and (b), on the other hand, establish the required proximity of the two multilevel ensembles particlewise, based on convergence of the random gains K_n^{ML} to the deterministic ones K_n. The latter will require the greatest effort and will dominate the proof by means of technical lemmas. Note that $\|\bar{v}_n^\ell\|_p$, $\|\widehat{\bar{v}}_n^\ell\|_p$, $|K_n| < \infty$, following from Assumption 1.

The first step is to bound the multilevel predicting covariance in terms of its constituents, the ensemble members. The gain is then bounded in terms of the covariance, and ultimately the updated ensemble in terms of the predicting ensemble and the covariance. The rate appears only by virtue of the convergence of the i.i.d. ensemble covariance, and it is propagated forward by induction. Only the *predicting* covariance will be considered and hats will be omitted to avoid unnecessary notation.

Recall the multilevel Kalman gain is defined as follows:

$$K_n^{\mathrm{ML}} = C_n^{\mathrm{ML}} H^{\mathsf{T}} (H \tilde{C}_n^{\mathrm{ML}} H^{\mathsf{T}} + \Gamma)^{-1},$$

where

$$\tilde{C}_n^{\mathrm{ML}} = \sum_{k=1;\lambda_k>0}^{d} \lambda_k q_k q_k^{\mathsf{T}}, \tag{3.5}$$

for eigenpairs $\{\lambda_k, q_k\}$ of C_n^{ML}. The following micro-lemma will be necessary to control the error in the gain.

LEMMA 3.3 (multilevel covariance approximation error). *Let $\tilde{C}_n^{\mathrm{ML}}$ be given by* (3.5). *Then the following holds:*

$$|\tilde{C}_n^{\mathrm{ML}} - C_n^{\mathrm{ML}}| \leq |C_n^{\mathrm{ML}} - C_n|, \tag{3.6}$$

where $|\cdot|$ denotes the induced 2-norm for matrices.

Proof. Notice that

$$|\tilde{C}_n^{\mathrm{ML}} - C_n^{\mathrm{ML}}| = \max_{\{j;\lambda_j<0\}}\{|\lambda_j|\}. \tag{3.7}$$

Denote the associated eigenvector by $u_{\max}$ (normalized to $|u_{\max}| = 1$). Notice that for any $A = A^{\mathsf{T}}$,

$$|A| = \sup_u \frac{|u^{\mathsf{T}} A u|}{|u|^2} = \max_k |\lambda_k|,$$

where λ_k are the eigenvalues of A. Since $C_n \geq 0$, one has that

$$|u_{\max}^{\mathsf{T}}(C_n^{\mathrm{ML}} - C_n)u_{\max}| = u_{\max}^{\mathsf{T}} C_n u_{\max} - u_{\max}^{\mathsf{T}} C_n^{\mathrm{ML}} u_{\max} \geq |\tilde{C}_n^{\mathrm{ML}} - C_n^{\mathrm{ML}}|. \qquad \square$$

The next step is to bound the gain error, which is done in the following lemma.

LEMMA 3.4 (continuity of the gain in the covariance). *There is a constant $c_n < \infty$, depending on $|H|, \gamma_{\min}$, and $|K_n H|$ such that*

$$|K_n^{\mathrm{ML}} - K_n| \leq c_n |C_n^{\mathrm{ML}} - C_n|, \tag{3.8}$$

where $\gamma_{\min} > 0$ is the smallest eigenvalue of Γ.

Proof. Recall that

$$\begin{aligned} K_n - K_n^{\mathrm{ML}} = {} & C_n H^{\mathsf{T}} \left((H C_n H^{\mathsf{T}} + \Gamma)^{-1} - (H \tilde{C}_n^{\mathrm{ML}} H^{\mathsf{T}} + \Gamma)^{-1} \right) \\ & + (C_n - C_n^{\mathrm{ML}}) H^{\mathsf{T}} (H \tilde{C}_n^{\mathrm{ML}} H^{\mathsf{T}} + \Gamma)^{-1}, \end{aligned} \tag{3.9}$$

where $\tilde{C}_n^{\mathrm{ML}} \geq 0$ is defined in (2.14), and notice that

$$\begin{aligned} & (H C_n H^{\mathsf{T}} + \Gamma)^{-1} - (H \tilde{C}_n^{\mathrm{ML}} H^{\mathsf{T}} + \Gamma)^{-1} \\ & \quad = (H C_n H^{\mathsf{T}} + \Gamma)^{-1} H (\tilde{C}_n^{\mathrm{ML}} - C_n) H^{\mathsf{T}} (H \tilde{C}_n^{\mathrm{ML}} H^{\mathsf{T}} + \Gamma)^{-1}. \end{aligned} \tag{3.10}$$

So

$$\begin{aligned} K_n - K_n^{\mathrm{ML}} = {} & K_n H (\tilde{C}_n^{\mathrm{ML}} - C_n) H^{\mathsf{T}} (H \tilde{C}_n^{\mathrm{ML}} H^{\mathsf{T}} + \Gamma)^{-1} \\ & + (C_n - C_n^{\mathrm{ML}}) H^{\mathsf{T}} (H \tilde{C}_n^{\mathrm{ML}} H^{\mathsf{T}} + \Gamma)^{-1}. \end{aligned}$$

Note that $x^{\mathsf{T}}(\Gamma + B)x \geq x^{\mathsf{T}} \Gamma x \geq \gamma_{\min}$ for all $x \in \mathbb{R}^d$ whenever $B = B^{\mathsf{T}} \geq 0$, and this implies that $|(H \tilde{C}_n^{\mathrm{ML}} H^{\mathsf{T}} + \Gamma)^{-1}| \leq 1/\gamma_{\min}$. It follows by (3.6) that

$$|K_n - K_n^{\mathrm{ML}}| \leq \frac{|H|}{\gamma_{\min}} (1 + 2|K_n H|) |C_n - C_n^{\mathrm{ML}}|. \qquad \square$$

It is worth noting that the multilevel gain error is bounded by the *unmodified* multilevel sample covariance error, following from Lemma 3.3, so modification in (3.5) will not affect the ultimate approximation error.

THEOREM 3.5. *Suppose Assumptions* 1 *and* 2 *hold. For any* $\varepsilon > 0$*, let* L *and* $\{M_\ell\}_{\ell=0}^L$ *be defined as in Theorem* 3.2*. Then the following asymptotic inequality holds:*

$$\|C_n^{\mathrm{ML}} - C_n\|_p \lesssim \varepsilon + \|C_n^{\mathrm{ML}} - \bar{C}_n^{\mathrm{ML}}\|_p \tag{3.11}$$

with a cost which satisfies (3.4)*.*

Proof. Let C_n^L denote the predicting covariance of the final Lth level limiting system at time n, in the sense that the forward map above is replaced by Ψ^L, but the gain comes from the continuum mean-field limiting system. Furthermore, let $\bar{C}_n^{\mathrm{ML}}$ denote the covariance associated to the multilevel ensemble $\{(\bar{v}_{n,i}^\ell)_{i=1}^{M_\ell}\}_{\ell=1}^L$. The triangle inequality is used to split

$$|C_n^{\mathrm{ML}} - C_n| \le |C_n^L - C_n| + |\bar{C}_n^{\mathrm{ML}} - C_n^L| + |C_n^{\mathrm{ML}} - \bar{C}_n^{\mathrm{ML}}|, \tag{3.12}$$

and each term will be dealt with in turn, in the following three lemmas. The proof of the theorem is done after establishing Lemmas 3.6, 3.7, and 3.8, which provide the asymptotic bound on the first two terms. ∎

LEMMA 3.6. *Suppose Assumptions* 1 *and* 2 *hold. For any* $\varepsilon > 0$*, let* L *and* $\{M_\ell\}_{\ell=0}^L$ *be defined as in Theorem* 3.2*. Then, for any finite* n*,*

$$\max\left(\left|\mathbb{E}\left[\bar{v}_n^L - v_n\right]\right|, \left|\mathbb{E}\left[\bar{\widehat{v}}_n^L - \widehat{v}_n\right]\right|\right) \lesssim \varepsilon, \tag{3.13}$$

and for $\ell = 0, 1, \ldots, L$ *and all* $p \ge 2$*,*

$$\max\left(\left\|\bar{v}_n^\ell - \bar{v}_n^{\ell-1}\right\|_p, \left\|\bar{\widehat{v}}_n^\ell - \bar{\widehat{v}}_n^{\ell-1}\right\|_p\right) \lesssim N_\ell^{-\beta/2}. \tag{3.14}$$

Proof. Since the initial data $\bar{\widehat{v}}_0^L = \bar{v}_0^L = \widehat{v}_0 = v_0$ is the same,

$$\left|\mathbb{E}\left[\bar{\widehat{v}}_0^L - \widehat{v}_0\right]\right| = 0.$$

Now assume that the following holds:

$$\max\left(\left|\left|\mathbb{E}\left[\bar{\widehat{v}}_{n-1}^L - \widehat{v}_{n-1}\right]\right|\right|, \left|\mathbb{E}\left[\bar{v}_{n-1}^L - v_{n-1}\right]\right|\right) \lesssim \varepsilon.$$

Assumption 2(i) directly implies

$$\left|\mathbb{E}\left[\bar{v}_n^L - v_n\right]\right| \lesssim \varepsilon.$$

Futhermore, since $|K_n H| < \infty$ for any finite n,

$$\left|\mathbb{E}\left[\bar{\widehat{v}}_n^L - \widehat{v}_n\right]\right| \le |I - K_n H| \left|\mathbb{E}\left[\bar{v}_n^L - v_n\right]\right| \lesssim \varepsilon,$$

and inequality (3.13) follows by induction.

To prove inequality (3.14), recall that due to the matching initial data, the inequality holds trivially at $n = 0$. Assume

$$\max\left(\left\|\bar{v}_{n-1}^\ell - \bar{v}_{n-1}^{\ell-1}\right\|_p, \left\|\bar{\widehat{v}}_{n-1}^\ell - \bar{\widehat{v}}_{n-1}^{\ell-1}\right\|_p\right) \lesssim N_\ell^{-\beta/2}.$$

By Assumptions 1(i) and 2(ii),

$$\begin{aligned}\left\|\bar{v}_n^\ell - \bar{v}_n^{\ell-1}\right\|_p &\le \left\|\Psi^\ell(\bar{\widehat{v}}_{n-1}^\ell) - \Psi^{\ell-1}(\bar{\widehat{v}}_{n-1}^\ell)\right\|_p + \left\|\Psi^{\ell-1}(\bar{\widehat{v}}_{n-1}^\ell) - \Psi^{\ell-1}(\bar{\widehat{v}}_{n-1}^{\ell-1})\right\|_p \\ &\lesssim N_\ell^{-\beta/2}, \\ \left\|\bar{\widehat{v}}_n^\ell - \bar{\widehat{v}}_n^{\ell-1}\right\|_p &\lesssim |I - K_n H| \left\|\bar{v}_n^\ell - \bar{v}_n^{\ell-1}\right\|_p \lesssim N_\ell^{-\beta/2},\end{aligned}$$

and inequality (3.14) holds by induction. ∎

LEMMA 3.7 (covariance discretization error). *Suppose Assumptions* 1 *and* 2 *hold. For any* $\varepsilon > 0$*, let* L *be defined as in Theorem* 3.2*. Then the following asymptotic inequality holds:*

$$|C_n^L - C_n| \lesssim \varepsilon. \tag{3.15}$$

Proof. It is possible to show that for any symmetric matrix A, the following inequality holds:

$$|A| \le \sum_{j,j'=1}^{d} |A^{jj'}|. \tag{3.16}$$

Furthermore, by adding the terms $\pm\mathbb{E}\big[(v_n)^j\big]\,\mathbb{E}\big[(\bar{v}_n^L)^{j'}\big]$,

$$\begin{aligned}|(C_n^L - C_n)^{jj'}| &= \Big|\mathbb{E}\Big[(\bar{v}_n^L)^j(\bar{v}_n^L)^{j'}\Big] - \mathbb{E}\Big[(v_n)^j(v_n)^{j'}\Big] \\ &\quad - \mathbb{E}\big[(\bar{v}_n^L)^j\big]\,\mathbb{E}\Big[(\bar{v}_n^L)^{j'}\Big] + \mathbb{E}\big[(v_n)^j\big]\,\mathbb{E}\Big[(v_n)^{j'}\Big]\Big| \\ &\le \Big|\mathbb{E}\Big[(\bar{v}_n^L)^j(\bar{v}_n^L)^{j'} - (v_n)^j(v_n)^{j'}\Big]\Big| \\ &\quad + \Big|\mathbb{E}\Big[(\bar{v}_n^L)^{j'}\Big]\Big|\,\big|\mathbb{E}\big[(\bar{v}_n^L)^j - (v_n)^j\big]\big| + |\mathbb{E}\big[(v_n)^j\big]|\,\Big|\mathbb{E}\Big[(\bar{v}_n^L)^{j'} - (v_n)^{j'}\Big]\Big| \\ &\lesssim N_L^{-\alpha} \eqsim \varepsilon.\end{aligned} \tag{3.17}$$

The last inequality follows by Lemma 3.6 and Assumption 2 (i), noting that $\mathcal{F}$ contains all monomials of degree less than or equal to 2. ∎

Notice that

$$C_n^L = \sum_{\ell=0}^{L} \mathrm{Cov}[\bar{v}_n^\ell] - \mathrm{Cov}[\bar{v}_n^{\ell-1}],$$

and

$$C_n^{\mathrm{ML}} = \sum_{\ell=0}^{L} \mathrm{Cov}_{M_\ell}[v_n^\ell] - \mathrm{Cov}_{M_\ell}[v_n^{\ell-1}],$$

with the convention that $v^{-1} = \bar{v}^{-1} := 0$. Consider also the partner covariance to the above:

$$\bar{C}_n^{\mathrm{ML}} = \sum_{\ell=0}^{L} \mathrm{Cov}_{M_\ell}[\bar{v}_n^\ell] - \mathrm{Cov}_{M_\ell}[\bar{v}_n^{\ell-1}].$$

The next two differences are bounded in terms of the single-level differences, using the triangle inequality to extend to the sum.

LEMMA 3.8 (multilevel i.i.d. sample covariance error). *Suppose Assumptions* 1 *and* 2 *hold, and for any* $\varepsilon > 0$, *let* L *and* $\{M_\ell\}_{\ell=0}^L$ *be defined as in Theorem* 3.2. *Then the following asymptotic inequality holds:*

$$\|\bar{C}_n^{\mathrm{ML}} - C_n^L\|_p \lesssim \varepsilon. \tag{3.18}$$

Proof. Notice the following triangle inequality:

$$|\bar{C}_n^{\mathrm{ML}} - C_n^L| \leq \sum_{\ell=0}^{L} |\mathrm{Cov}_{M_\ell}[\bar{v}_n^\ell] - \mathrm{Cov}_{M_\ell}[\bar{v}_n^{\ell-1}] - (\mathrm{Cov}_{M_\ell}[\bar{v}_n^\ell] - \mathrm{Cov}_{M_\ell}[\bar{v}_n^{\ell-1}])|.$$

To avoid needlessly long terms when bounding the summands of the above equation, we now make the assumption in this proof that $\mathbb{E}\big[\bar{v}_n^\ell\big] = 0$, without loss of generality. We may then obtain the rearrangement

$$\mathrm{Cov}_{M_\ell}[\bar{v}_n^\ell] = E_{M_\ell}\big[\bar{v}_n^\ell(\bar{v}_n^\ell)^{\mathsf{T}}\big] - E_{M_\ell}[\bar{v}_n^\ell]\big(E_{M_\ell}[\bar{v}_n^\ell]\big)^{\mathsf{T}},$$

and similarly for the $\ell-1$ term. Using the identity $aa^{\mathsf{T}} - bb^{\mathsf{T}} = \frac{1}{2}[(a+b)(a-b)^{\mathsf{T}} + (a-b)(a+b)^{\mathsf{T}}]$ for $a, b \in \mathbb{R}^d$ on each of the outer products with $\ell, \ell-1$, respectively, and then using (3.16) again for the first term, and Cauchy–Schwartz for the second (and grouping like terms arising from the $(j, j') \to (j', j)$ symmetry of $\frac{1}{2}(a^j b^{j'} + a^{j'} b^j)$), one has

$$\begin{aligned}
&\big|\mathrm{Cov}_{M_\ell}[\bar{v}_n^\ell] - \mathrm{Cov}_{M_\ell}[\bar{v}_n^{\ell-1}] - (\mathrm{Cov}[\bar{v}_n^\ell] - \mathrm{Cov}[\bar{v}_n^{\ell-1}])\big| \\
&\leq \sum_{j \leq j'=1}^{d} \Big| E_{M_\ell}\big[\big(\bar{v}_n^\ell + \bar{v}_n^{\ell-1}\big)^j \big(\bar{v}_n^\ell - \bar{v}_n^{\ell-1}\big)^{j'}\big] - \mathbb{E}\Big[\big(\bar{v}_n^\ell + \bar{v}_n^{\ell-1}\big)^j \big(\bar{v}_n^\ell - \bar{v}_n^{\ell-1}\big)^{j'}\Big] \Big| \\
&\quad + \Big| E_{M_\ell}\big[\big(\bar{v}_n^\ell + \bar{v}_n^{\ell-1}\big)^{j'} \big(\bar{v}_n^\ell - \bar{v}_n^{\ell-1}\big)^{j}\big] - \mathbb{E}\Big[\big(\bar{v}_n^\ell + \bar{v}_n^{\ell-1}\big)^{j'} \big(\bar{v}_n^\ell - \bar{v}_n^{\ell-1}\big)^{j}\Big] \Big| \\
&\quad + \big|E_{M_\ell}[\bar{v}_n^\ell + \bar{v}_n^{\ell-1})^j]\big| \Big|E_{M_\ell}[(\bar{v}_n^\ell - \bar{v}_n^{\ell-1})^{j'}]\Big| \\
&\quad + \Big|E_{M_\ell}[(\bar{v}_n^\ell + \bar{v}_n^{\ell-1})^{j'}]\Big| \big|E_{M_\ell}[(\bar{v}_n^\ell - \bar{v}_n^{\ell-1})^{j}]\big|.
\end{aligned} \tag{3.19}$$

Almost sure convergence follows by the law of large numbers. The rate in L^p is shown now.

First, it will be necessary to recall the Marcinkiewicz–Zygmund inequality: for i.i.d. random variables $X_1, \ldots, X_N \sim X$ with $\|X\|_p < \infty$ for $p \geq 2$, and $\mathbb{E}[X] = 0$,

$$\|E_N[X]\|_p \leq c_p N^{-1/2} \|X\|_p, \tag{3.20}$$

where the constant depends only on p (cf. [8, 22]); in fact, $c_p \leq 3\sqrt{2p}$ (cf. [41]).

Using the Marcinkiewicz–Zygmund inequality, then Hölder's inequality on each of the first two terms on the right-hand side of (3.19), then the reverse order on the last two, and finally Assumption 2, Lemma 3.6, and the fact $\bar{v}_n^\ell \in L^p(\Omega)$ for all $p \geq 2$,

together yield

$$
\begin{aligned}
&\|\mathrm{Cov}_{M_\ell}[\bar{v}_n^\ell] - \mathrm{Cov}_{M_\ell}[\bar{v}_n^{\ell-1}] - (\mathrm{Cov}[\bar{v}_n^\ell] - \mathrm{Cov}[\bar{v}_n^{\ell-1}])\|_p \\
&\quad \le \sum_{j\le j'=1}^{d} M_\ell^{-1/2}\Bigg[c_p\big(\|\big(\bar{v}_n^\ell + \bar{v}_n^{\ell-1}\big)^j\|_{2p}\|\big(\bar{v}_n^\ell - \bar{v}_n^{\ell-1}\big)^{j'}\|_{2p} \\
&\qquad + \|\big(\bar{v}_n^\ell + \bar{v}_n^{\ell-1}\big)^{j'}\|_{2p}\|\big(\bar{v}_n^\ell - \bar{v}_n^{\ell-1})\big)^j\|_{2p}\big) \\
&\qquad + c_{2p}^2 M_\ell^{-1/2}\Big(\|\big(\bar{v}_n^\ell + \bar{v}_n^{\ell-1}\big)^j\|_{2p}\|\big(\bar{v}_n^\ell - \bar{v}_n^{\ell-1}\big)^{j'}\|_{2p} \\
&\qquad + \|\big(\bar{v}_n^\ell + \bar{v}_n^{\ell-1})^{j'}\|_{2p}\|\big(\bar{v}_n^\ell - \bar{v}_n^{\ell-1}\big)^j\|_{2p}\Big)\Bigg] \\
&\quad \lesssim M_\ell^{-1/2} N_\ell^{-\beta/2}.
\end{aligned} \tag{3.21}
$$

Finally, by the triangle inequality, the following bound holds for (3.18) for all $p \ge 2$:

$$
\|\bar{C}_n^{\mathrm{ML}} - C_n^L\|_p \lesssim \sum_{\ell=0}^{L} M_\ell^{-1/2} N_\ell^{-\beta/2} \lesssim \varepsilon. \qquad \square
$$

The previous two lemmas complete the proof of Theorem 3.5. Now we turn to the next term in (3.11), the difference between multilevel ensemble covariances, which is continuous in the individual ensemble members. First it will be necessary to recall (see, e.g., Lemma 4.3 of [35]) that for identically distributed random variables $x_1, \dots, x_N \in \mathbb{R}^d$,

$$
\Big(\mathbb{E}\Big[\Big|E_N\big[|x_n|^p\big]^{1/p}\Big|^q\Big]\Big)^{1/q} \le \|x_n\|_r, \tag{3.22}
$$

where $r = \max\{q, p\}$.

LEMMA 3.9 (continuity of multilevel sample covariances in particles). *Suppose Assumptions* 1 *and* 2 *hold, and for any* $\varepsilon > 0$, *let* L *and* $\{M_\ell\}_{\ell=0}^L$ *be defined as in Theorem* 3.2. *Then the following asymptotic inequality holds for all* $p \ge 2$*:*

$$
\begin{aligned}
\|C_n^{\mathrm{ML}} - \bar{C}_n^{\mathrm{ML}}\|_p \le \sum_{l=0}^{L} \Bigg(&\|v_n^\ell - \bar{v}_n^\ell\|_p + 4\|v_n^\ell - \bar{v}_n^\ell\|_{2p}\|\bar{v}_n^\ell\|_{2p} \\
&+ \|v_n^{\ell-1} - \bar{v}_n^{\ell-1}\|_p + 4\|v_n^{\ell-1} - \bar{v}_n^{\ell-1}\|_{2p}\|\bar{v}_n^{\ell-1}\|_{2p}\Bigg).
\end{aligned} \tag{3.23}
$$

Proof. Recall first that

$$
\big|C_n^{\mathrm{ML}} - \bar{C}_n^{\mathrm{ML}}\big| \le \sum_{\ell=1}^{L} \big|\mathrm{Cov}_{M_\ell}[v_n^\ell] - \mathrm{Cov}_{M_\ell}[v_n^{\ell-1}] - (\mathrm{Cov}_{M_\ell}[\bar{v}_n^\ell] - \mathrm{Cov}_{M_\ell}[\bar{v}_n^{\ell-1}])\big|.
$$

Now the individual terms will be bounded. Note that

$$
\mathrm{Cov}_{M_\ell}[v_n^\ell] = E_{M_\ell}[v_n^\ell (v_n^\ell)^{\mathsf{T}}] - E_{M_\ell}[v_n^\ell]\big(E_{M_\ell}[v_n^\ell]\big)^{\mathsf{T}},
$$

and similarly for $\bar{v}_n^\ell$. Using $a^2 - b^2 = (a-b)^2 + 2b(a-b)$ with $a = u^{\mathsf{T}} v_n^\ell$ and $b = u^{\mathsf{T}} \bar{v}_n^\ell$ and again with $a = (1/M_\ell)\sum_{i=1}^{M_\ell} u^{\mathsf{T}} v_{n,i}^\ell$ and $b = (1/M_\ell)\sum_{i=1}^{M_\ell} u^{\mathsf{T}} \bar{v}_{n,i}^\ell$ for arbitrary

$u \in \mathbb{R}^d$, these terms are rearranged as follows:

$$\begin{aligned} &u^{\mathsf{T}}\big(\mathrm{Cov}_{M_\ell}[v_n^\ell] - \mathrm{Cov}_{M_\ell}[\bar{v}_n^\ell]\big)u \\ &= E_{M_\ell}[\big|u^{\mathsf{T}}(v_n^\ell - \bar{v}_n^\ell)\big|^2] + 2E_{M_\ell}[(u^{\mathsf{T}}\bar{v}_n^\ell)(u^{\mathsf{T}}(v_n^\ell - \bar{v}_n^\ell))] \\ &- \big|E_{M_\ell}[u^{\mathsf{T}}(v_n^\ell - \bar{v}_n^\ell)]\big|^2 - 2E_{M_\ell}[u^{\mathsf{T}}\bar{v}_n^\ell]E_{M_\ell}[u^{\mathsf{T}}(v_n^\ell - \bar{v}_n^\ell)]. \end{aligned}$$

Then, using the Cauchy–Schwartz inequality, the first term of (3.12) is bounded as follows:

$$\begin{aligned} &\big|\mathrm{Cov}_{M_\ell}[v_n^\ell] - \mathrm{Cov}_{M_\ell}[v_n^{\ell-1}] - (\mathrm{Cov}_{M_\ell}[\bar{v}_n^\ell] - \mathrm{Cov}_{M_\ell}[\bar{v}_n^{\ell-1}])\big| \\ &= \big|\mathrm{Cov}_{M_\ell}[v_n^\ell] - \mathrm{Cov}_{M_\ell}[\bar{v}_n^\ell] + \mathrm{Cov}_{M_\ell}[v_n^{\ell-1}] - \mathrm{Cov}_{M_\ell}[\bar{v}_n^{\ell-1}]\big| \\ &\le E_{M_\ell}\big[\,\big|v_n^\ell - \bar{v}_n^\ell\big|^2\,\big] + 4\sqrt{E_{M_\ell}\big[\,\big|\bar{v}_n^\ell\big|^2\,\big]E_{M_\ell}\big[\,\big|v_n^\ell - \bar{v}_n^\ell\big|^2\,\big]} \\ &\quad + E_{M_\ell}\big[\,\big|v_n^{\ell-1} - \bar{v}_n^{\ell-1}\big|^2\,\big] + 4\sqrt{E_{M_\ell}\big[\,\big|\bar{v}_n^{\ell-1}\big|^2\big]E_{M_\ell}\big[\big|v_n^{\ell-1} - \bar{v}_n^{\ell-1}\big|^2\,\big]}. \end{aligned} \tag{3.24}$$

After rearrangement, the triangle inequality, (3.22) with $p = \max\{p, 2\}$, and Hölder's inequality complete the proof. ∎

It has just been shown that the second term of (3.11) is "close in the predicting ensembles." Therefore, the error level of the first term will carry over between observation times by induction. This is made rigorous by the next lemma.

LEMMA 3.10 (distance between ensembles.). *Suppose Assumptions* 1 *and* 2 *hold, and for any* $\varepsilon > 0$*, let* L *and* $\{M_\ell\}_{\ell=0}^L$ *be defined as in Theorem* 3.2*. Then the following asymptotic inequality holds for all* $p \ge 2$*:*

$$\sum_{\ell=0}^{L} \left\| \widehat{v}_n^\ell - \widehat{\bar{v}}_n^\ell \right\|_p \lesssim |\log(\varepsilon)|^n \, \varepsilon. \tag{3.25}$$

Proof. First recall that the assertion holds trivially for $n = 0$. Proceeding by induction, assume for $p \ge 2$,

$$\sum_{\ell=0}^{L} \left\| \widehat{v}_{n-1}^\ell - \widehat{\bar{v}}_{n-1}^\ell \right\|_p \lesssim |\log(\varepsilon)|^{n-1} \, \varepsilon.$$

Then Assumption 1(i) implies the following inequality holds for the prediction:

$$\sum_{\ell=0}^{L} \left\| v_n^\ell - \bar{v}_n^\ell \right\|_p \le c_\Psi \sum_{\ell=0}^{L} \left\| \widehat{v}_{n-1}^\ell - \widehat{\bar{v}}_{n-1}^\ell \right\|_p \lesssim |\log(\varepsilon)|^{n-1} \, \varepsilon. \tag{3.26}$$

Using Lemma 3.4, the following inequalities hold for $\ell = 0, \ldots, L$:

$$|\widehat{v}_n^\ell - \widehat{\bar{v}}_n^\ell| \le |I - K_n H||v_n^\ell - \bar{v}_n^\ell| + c_n |C_n^{\mathrm{ML}} - C_n| \Big(|v_n^\ell - \bar{v}_n^\ell| + |y_n^\ell - H\bar{v}_n^\ell| \Big). \tag{3.27}$$

By Hölder's inequality and since $y_n^\ell, \bar{v}_n^\ell \in L^p(\Omega)$ for all $p \ge 2$,

$$\begin{aligned} \|\widehat{v}_n^\ell - \widehat{\bar{v}}_n^\ell\|_p &\le |I - K_n H| \|v_n^\ell - \bar{v}_n^\ell\|_p \\ &\quad + c_n \|C_n^{\mathrm{ML}} - C_n\|_{2p} \Big(\|v_n^\ell - \bar{v}_n^\ell\|_{2p} + \|y_n^\ell - H\bar{v}_n^\ell\|_{2p} \Big) \\ &\lesssim \|v_n^\ell - \bar{v}_n^\ell\|_p + \|C_n^{\mathrm{ML}} - C_n\|_{2p} \Big(\|v_n^\ell - \bar{v}_n^\ell\|_{2p} + 1 \Big). \end{aligned}$$

Plugging the moment bound (3.26) into the right-hand side of the inequality (3.23) yields that $\|C_n^{\mathrm{ML}} - \bar{C}_n^{\mathrm{ML}}\|_{2p} \lesssim |\log(\varepsilon)|^{n-1}\,\varepsilon$, which in combination with Theorem 3.5 further leads to $\|C_n^{\mathrm{ML}} - C_n\|_{2p} \lesssim |\log(\varepsilon)|^{n-1}\,\varepsilon$. Therefore, summing the above and using (3.26) again for $p, 2p$

$$\begin{aligned}\sum_{\ell=0}^{L} \|\widehat{v}_n^\ell - \bar{\widehat{v}}_n^\ell\|_p &\lesssim \sum_{\ell=0}^{L} \|v_n^\ell - \bar{v}_n^\ell\|_p + \varepsilon\left(\|v_n^\ell - \bar{v}_n^\ell\|_{2p} + \|y_n^\ell - \bar{v}_n^\ell\|_{2p}\right)\\ &\lesssim |\log(\varepsilon)|^{n-1}\,\varepsilon\Big(1 + \sum_{\ell=0}^{L} \|y_n^\ell - \bar{v}_n^\ell\|_{2p}\Big)\\ &\lesssim |\log(\varepsilon)|^{n}\,\varepsilon,\end{aligned} \tag{3.28}$$

where the last inequality of the proof uses that $\|y_n^\ell - H\bar{v}_n^\ell\|_{2p} \lesssim 1$ and $L \eqsim |\log(\varepsilon)|$. ∎

Induction is complete on the distance between the multilevel ensemble and its i.i.d. shadow in L^p, and it remains only to close the argument, which is done next. Note that the induction actually holds for all n, but we are able to neglect the n-dependence of the constant c_n appearing in (3.27) by considering only a finite number N of steps.

Proof of Theorem 3.2. What remains is to verify that provided L and M_ℓ are defined under the constraints in Theorem 3.2, the error bound (3.3) will be obtained for all the functions $\varphi \in \mathcal{F}$ which are locally Lipschitz continuous with at most polynomial growth at infinity; cf. Definition 3.1. Notice that the triangle inequality gives

$$\begin{aligned}\|\mu_n^{\mathrm{ML}}(\varphi) - \mu_n(\varphi)\|_p \le{}& \|\mu_n^{\mathrm{ML}}(\varphi) - \bar{\mu}_n^{\mathrm{ML}}(\varphi)\|_p\\ &+ \|\bar{\mu}_n^{\mathrm{ML}}(\varphi) - \bar{\mu}_n^{L}(\varphi)\|_p + \|\bar{\mu}_n^{L}(\varphi) - \mu_n(\varphi)\|_p,\end{aligned} \tag{3.29}$$

where $\bar{\mu}_n^{\mathrm{ML}}$ denotes the empirical measure associated to the i.i.d. ensemble, and $\bar{\mu}_n^{L}$ denotes the probability measure associated to $\bar{v}^L$. Before treating each term separately, we notice that the two first summands of the right-hand side of the inequality relate to the statistical error, whereas the last relates to the bias.

The first summand of (3.29) satisfies the following bound:

$$\begin{aligned}\|\mu_n^{\mathrm{ML}}(\varphi) - \bar{\mu}_n^{\mathrm{ML}}(\varphi)\|_p &= \Big\|\sum_{\ell=0}^{L} E_{M_\ell}\Big[\varphi(\widehat{v}_n^\ell) - \varphi(\widehat{v}_n^{\ell-1}) - (\varphi(\bar{\widehat{v}}_n^\ell) - \varphi(\bar{\widehat{v}}_n^{\ell-1}))\Big]\Big\|_p\\ &\le \sum_{\ell=0}^{L} \|\varphi(\widehat{v}_n^\ell) - \varphi(\bar{\widehat{v}}_n^\ell)\|_p + \|\varphi(\widehat{v}_n^{\ell-1}) - \varphi(\bar{\widehat{v}}_n^{\ell-1})\|_p\\ &\le C\sum_{\ell=0}^{L}\Big[\big\||\widehat{v}_n^\ell - \bar{\widehat{v}}_n^\ell|\left(1 + |\bar{\widehat{v}}_n^\ell|^\nu\right) + |\widehat{v}_n^\ell - \bar{\widehat{v}}_n^\ell|^{\nu+1}\big\|_p\\ &\quad + \big\||\widehat{v}_n^{\ell-1} - \bar{\widehat{v}}_n^{\ell-1}|\left(1 + |\bar{\widehat{v}}_n^{\ell-1}|^\nu\right) + |\widehat{v}_n^{\ell-1} - \bar{\widehat{v}}_n^{\ell-1}|^{\nu+1}\big\|_p\Big]\\ &\le C\sum_{\ell=0}^{L}\Big[\|\widehat{v}_n^\ell - \bar{\widehat{v}}_n^\ell\|_{2p}\left(1 + \|\bar{\widehat{v}}_n^\ell\|_{2p\nu}^\nu\right) + \|\widehat{v}_n^\ell - \bar{\widehat{v}}_n^\ell\|_{p(\nu+1)}^{\nu+1}\\ &\quad + \|\widehat{v}_n^{\ell-1} - \bar{\widehat{v}}_n^{\ell-1}\|_{2p}\left(1 + \|\bar{\widehat{v}}_n^{\ell-1}\|_{2p\nu}^\nu\right) + \|\widehat{v}_n^{\ell-1} - \bar{\widehat{v}}_n^{\ell-1}\|_{p(\nu+1)}^{\nu+1}\Big]\\ &\lesssim |\log(\varepsilon)|^{n}\,\varepsilon.\end{aligned} \tag{3.30}$$

The second inequality follows directly from the expression (3.1). The third inequality comes from the triangle inequality and Hölder's inequality, and the fourth inequality arises directly from Lemma 3.10 and the boundedness of $\bar{\widehat{v}}_n$ in L^p for $p \geq 2$.

For the second summand of (3.29), notice that we can write $\bar{\mu}_n^L = \sum_{\ell=0}^L \bar{\mu}_n^\ell - \bar{\mu}_n^{\ell-1}$, where $\bar{\mu}_n^\ell$ is the measure associated to the level ℓ limiting process $\bar{v}^\ell$. Then, by virtue of (3.20) and condition (ii) of Assumption 2,

$$\begin{aligned}
\|\bar{\mu}_n^{\mathrm{ML}}(\varphi) - \bar{\mu}_n^L(\varphi)\|_p &\leq \sum_{\ell=0}^L \left\| E_{M_\ell}\left[\varphi(\bar{\widehat{v}}_n^\ell) - \varphi(\bar{\widehat{v}}_n^{\ell-1}) - \mathbb{E}[\varphi(\bar{\widehat{v}}_n^\ell) - \varphi(\bar{\widehat{v}}_n^{\ell-1})]\right]\right\|_p \\
&\leq C \sum_{\ell=0}^L M_\ell^{-1/2} \|\varphi(\bar{\widehat{v}}_n^\ell) - \varphi(\bar{\widehat{v}}_n^{\ell-1})\|_p \\
&\leq C \sum_{\ell=0}^L M_\ell^{-1/2} \|\bar{\widehat{v}}_n^\ell - \bar{\widehat{v}}_n^{\ell-1}\|_p \\
&\lesssim \sum_{\ell=0}^L M_\ell^{-1/2} N_\ell^{-\beta/2} \lesssim \varepsilon.
\end{aligned} \tag{3.31}$$

Finally, for the bias term,

$$\|\bar{\mu}_n^L(\varphi) - \mu_n(\varphi)\|_p = |\bar{\mu}_n^L(\varphi) - \mu_n(\varphi)| = \left|\mathbb{E}\left[\varphi(\bar{\widehat{v}}_n^L) - \varphi(\widehat{v}_n)\right]\right| \lesssim \varepsilon, \tag{3.32}$$

where the last inequality follows from Lemma 3.6 and Assumption 2(i).

Putting together (3.30), (3.31), and (3.32) in (3.29) yields the sought bound in (3.3). ∎

Theorem 3.2 shows the cost-to-ε performance of MLEnKF, and to verify that it generally outperforms EnKF in this performance measure, we end this section with a comparable result on the cost-to-ε perfomance of EnKF.

THEOREM 3.11 (EnKF accuracy versus cost). *Suppose Assumptions* 1, 2(i), *and* 2(iii) *hold. For a given $\varepsilon > 0$, let L and M be defined under the constraints $L \eqsim \log(\varepsilon^{-1})/\alpha$ and $M \eqsim \varepsilon^{-2}$. Then for all functions $\varphi \in \mathcal{F}$ that are locally Lipschitz continuous with at most polynomial growth at infinity (cf. Definition* 3.1*), we have for any $p \geq 2$,*

$$\|\mu_n^{\mathrm{MC}}(\varphi) - \mu_n(\varphi)\|_p \lesssim \varepsilon. \tag{3.33}$$

Here μ_n^{MC} denotes the EnKF empirical measure defined in (2.10)*, where the samples are given by the EnKF predict formulae at resolution level L (i.e., with the numerical integrator Ψ^L), approximating the time $t_n = n$ mean-field EnKF distribution μ_n. The computational cost of the EnKF estimator over the time sequence satisfies*

$$\mathrm{Cost\,(EnKF)} \lesssim \varepsilon^{-(2+\gamma/\alpha)}. \tag{3.34}$$

Sketch of proof. By the triangle inequality

$$\begin{aligned}
\|\mu_n(\varphi) - \mu_n^{\mathrm{MC}}(\varphi)\|_p \leq{}& \left\|\mu_n(\varphi) - \bar{\mu}_n^L(\varphi)\right\|_p, + \left\|\bar{\mu}^L(\varphi) - \bar{\mu}_n^{\mathrm{MC}}(\varphi)\right\|_p \\
&+ \left\|\bar{\mu}_n^{\mathrm{MC}}(\varphi) - \mu_n^{\mathrm{MC}}(\varphi)\right\|_p =: \mathrm{I} + \mathrm{II} + \mathrm{III},
\end{aligned}$$

where $\bar{\mu}_n^{\mathrm{MC}}$ denotes the empricial measure associated to an EnKF ensemble $\{\bar{\widehat{v}}_n^L(\omega_i)\}_{i=1}^M$ and $\bar{\mu}_n^L$ denotes the emprical measure associated to $\bar{\widehat{v}}_n^L$. We bound the terms I, II, and III individually.

For the first term, we have

$$I = \left|\mu_n(\varphi) - \bar{\mu}_n^L(\varphi)\right| \leq \left|\mathbb{E}\left[\varphi(\widehat{v}_n) - \varphi(\bar{\widehat{v}}_n^L)\right]\right| \lesssim \varepsilon,$$

where the last inequality is implied by inequality (3.13) of Lemma 3.6, which it is straightforward to verify holds under Assumption 2(i).

For the second term, we first note that we may assume without loss of generality that $\varphi(0) = 0$. Since φ is locally Lipschitz continuous with at most polynomial growth at infinity, there then exist positive scalars ν, C_φ such that

$$|\varphi(x)| \leq C_\varphi(1 + |x|^\nu).$$

By inequality (3.20) and since $\bar{v}_n^L \in L^p(\Omega)$ for any $p > 1$,

$$II \leq \left\|E_M[\varphi(\bar{\widehat{v}}_n^L)] - \mathbb{E}\left[\varphi(\bar{\widehat{v}}_n^L)\right]\right\|_p \leq M^{-1/2}\left\|\varphi(\bar{\widehat{v}}_n^L)\right\|_p \leq C\varepsilon\left\|1 + \left|\bar{\widehat{v}}_n^L\right|^\nu\right\|_p \lesssim \varepsilon.$$

For the last term, let us first assume that for any $p \geq 2$ and finite n

$$\left\|\widehat{v}_n^L - \bar{\widehat{v}}_n^L\right\| \lesssim \varepsilon, \tag{3.35}$$

for the particle dynamics $\widehat{v}_n^L$ and $\bar{\widehat{v}}_n^L$, respectively, associated to the EnKF ensemble $\{\widehat{v}_{n,i}^L\}_{i=1}^M$ and the mean-field EnKF ensemble $\{\bar{\widehat{v}}_{n,i}^L\}_{i=1}^M$. Then the assumed regularity of φ, that $\widehat{v}_n^L, \bar{\widehat{v}}_n^L \in L^p(\Omega)$ for all $p \geq 2$, and Hölder's inequality yield that

$$\begin{aligned}\text{III} &= \left\|E_M[\varphi(\widehat{v}_n^L) - \varphi(\bar{\widehat{v}}_n^L)]\right\| \leq C_\varphi\left\|\left|\widehat{v}_n^L - \bar{\widehat{v}}_n^L\right|\left(1 + \left|\widehat{v}_n^L\right|^\nu + \left|\bar{\widehat{v}}_n^L\right|^\nu\right)\right\|_p \\ &\lesssim \left\|\widehat{v}_n^L - \bar{\widehat{v}}_n^L\right\|_p \lesssim \varepsilon.\end{aligned}$$

All that remains is to verify (3.35). Since this can be done by very similar steps as in the proof of inequality (3.25), we omit this verification. ∎

Remark 4. Notice that for a given n one can obtain an error $\mathcal{O}(\varepsilon)$ for MLEnKF in (3.3) for an additional cost which is given by replacing ε by $\varepsilon|\log\varepsilon|^{-n}$ in (3.4). Furthermore, it is worth noting that for any m and for any $\delta > 0$, $|\log\varepsilon|^m = \mathcal{O}(\varepsilon^{-\delta})$. Hence one can obtain a cost-of-error rate in (3.4) which is uniform in time and asymptotically superior to EnKF (3.34).

4. Numerical examples. In this section the performance of EnKF and MLEnKF are compared on some very simple numerical examples in terms of computational cost versus approximation error. First, in section 4.1, underlying dynamics from an Ornstein–Uhlenbeck SDE is considered. Next, in section 4.2, the underlying dynamics of GBM is considered. Both of these examples are indeed analytically tractable; however, they are approximated as though they were not. This provides a solid benchmark to compute errors and allows the theory to be illustrated.

4.1. An Ornstein–Uhlenbeck SDE. We first consider the simple Ornstein–Uhlenbeck SDE problem

$$du = -u\,dt + \sigma\,dW_t, \qquad u(0) = 1. \tag{4.1}$$

It has the exact solution

$$u(t) = u(0)e^{-t} + \int_0^t \sigma e^{(s-t)}\,dW_s,$$

and since

$$\int_0^1 \sigma e^{(s-1)} dW_s \sim N\Big(0, \underbrace{\frac{\sigma^2}{2}(1-e^{-2})}_{=:\Sigma}\Big),$$

one SDE realization sampled at the observation times $t_n = n$ is generated by the linear solution operator

$$u_{n+1} = e^{-1}u_n + \xi_n =: \Psi(u_n),$$

where $\xi_n \sim N(0, \Sigma)$ i.i.d. The corresponding noisy observations are given by

$$y_n^{\text{obs}} = u_n + \eta_n$$

with $\eta_n \sim N(0, \Gamma)$ i.i.d.

For the MLEnKF algorithm, a hierarchy of Milstein solution operators $\{\Psi^\ell\}_{\ell=0}^\infty$ are introduced, where the ℓth level solution operator uses a uniform time-step of size $\Delta t^\ell = 2^{-(\ell+1)}$. A numerical integration step takes the form

$$u_{n,m+1}^\ell = u_{n,m}^\ell(1-\Delta t^\ell) + \sigma \Delta W_{n,m}^\ell, \qquad m = 0, 1, \ldots, 2^{\ell+1}-1, \tag{4.2}$$

where the initial condition is given by $v_{n,0}^\ell = \widehat{v}_{n-1}^\ell$,

$$\Delta W_{n,m}^\ell = W(t_n + (m+1)\Delta t^\ell) - W(t_n + m\Delta t^\ell) \sim N(0, \Delta t^\ell),$$

and $u_n^\ell = u_{n,2^{\ell+1}}^\ell$.

Moreover, since the solution operator for (4.1) is linear, the gold standard becomes the conventional Kalman filter update

$$(\widehat{m}_n^\dagger, \widehat{C}_n^\dagger) = \Big((I-K_nH)m_n + K_n y_n^{\text{obs}}, (I-K_nH)C_n\Big), \quad n = 1, 2, \ldots.$$

Problem parameters. In the numerical experiments, $N = 100, 200$, and 400 observation times $\{t_n = n\}_{n=1}^N$ are used, and the covariance parameters are set to $\Gamma = 0.04$ and $\sigma = 0.5$. For a prescribed computational cost $\mathcal{O}(J)$, an EnKF ensemble of size $M = \mathcal{O}\left(J^{2/3}\right)$ is solved by the Milstein method on a mesh $\Delta t = \mathcal{O}\left(J^{1/3}\right)$, and for the MLEnKF method, we set L and M_ℓ according to the constraint in Theorem 3.2.

Approximations of the mean and covariance. In our first numerical experiment we approximate the gold standard mean and covariance for a single observation realization using the respective EnKF methods and measure the approximation error in terms of the root mean square error (RMSE):

$$\sqrt{\sum_{n=1}^N \frac{|\widehat{m}_n^\dagger - \widetilde{m}_n|^2}{N}}, \qquad \sqrt{\sum_{n=1}^N \frac{|\widehat{C}_n^\dagger - \widetilde{C}_n|^2}{N}}, \tag{4.3}$$

with $(\widetilde{m}_n, \widetilde{C}_n)$ denoting a single realization of either the EnKF or the MLEnKF updates approximating the gold standard moments. These observables are sufficiently smooth to reach the rates $\alpha = 1$ and $\beta = 2$ with the Milstein method; cf. [20]. The respective decay rates are numerically verified over a sequence of times in Figure 1. Figure 2 presents a numerical performance study measuring RMSE (4.3) versus computational cost for the respective methods. As is to be expected from Theorem 3.2

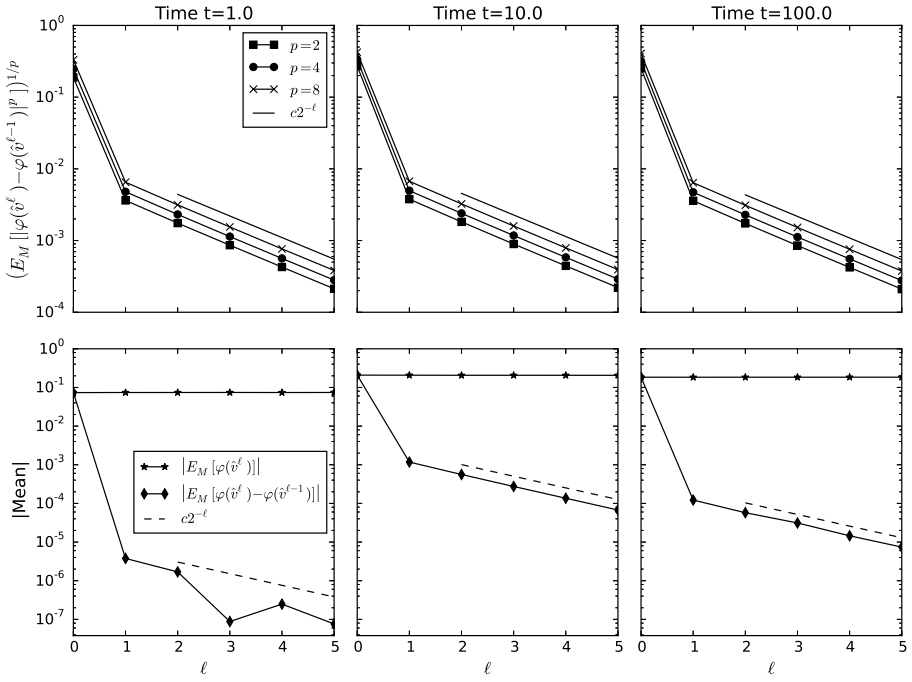

FIG. 1. *Numerical estimates of the decay rates over a sequence of times for the problem presented in section* 4.1 *with* $\varphi(v) = v$. *The computations use* $M = 10^6$ *particles on every level.*

the decay of RMSE for the MLEnKF method as a function of the cost J is roughly $\mathcal{O}(J^{-1/2})$, orders of magnitude faster than the observed and expected EnKF decay rate $\mathcal{O}(J^{-1/3})$. Note that the error growth factor $|\log(\varepsilon)|^n$ from the theoretical bound (3.3) is not visible in the experiments. In fact, the constant is even stable (the shift in cost as measured by runtime is simply due to computation of additional updates), indicating that with a more careful analysis the present results may be extended to an infinite time horizon.

Approximations of the excedence probability. In our second numerical test, we approximate the mean of the observable $\varphi(\widehat{u}_n) := \mathbf{1}\{\widehat{u}_n > 0.1\}$, which corresponds to the excedence probability $\mathbb{P}(\widehat{u}_n > 0.1) = 1 - \Phi((0.1 - \widehat{m}_n^\dagger)/\sqrt{\widehat{C}_n^\dagger})$. The Milstein method achieves the weak rate $\alpha = 1$, but while one may show for $p = 2$ and any $\delta > 0$, $\|\varphi(\Psi^\ell(v)) - \varphi(\Psi^{\ell-1}(v))\|_p \lesssim N_\ell^{(1-\delta)/2}$ (cf. [17, 2]), the low regularity of the observable implies that there does not exist a $\beta > 0$ fulfilling condition (ii) of Assumption 2 for all $p > 2$. A numerical inference of $\beta = 0$ can be made from the numerical estimates of the decay rates in Figure 3, where we see that the decay rate of $\|\varphi(\Psi^\ell(v)) - \varphi(\Psi^{\ell-1}(v))\|_p$ consistently decreases toward 0 as p increases over a sequence of times (while $\alpha \approx 1$). Theorem 3.2 therefore does not cover the given approximation problem. Nonetheless, implementing with the rates $\beta = 1$ and $\alpha = 1$, a numerical comparison of the performance of EnKF and MLEnKF approximating the excedence probability is presented in Figure 4. A near optimal RMSE decay rate, slightly slower than $\mathcal{O}(J^{-1/2})$, is again achieved for the MLEnKF method.

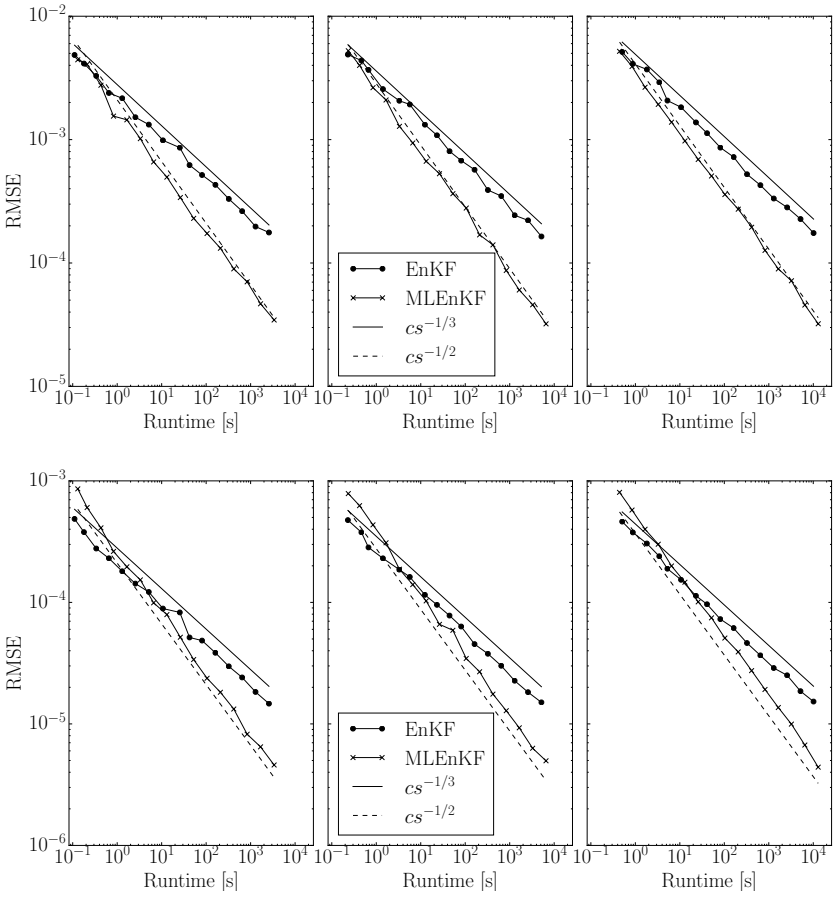

FIG. 2. *Comparison of the accuracy versus computational cost when using the EnKF and MLEnKF methods on the filtering problem presented in section* 4.1. *The error is measured in terms of the RMSE* (4.3) *for the mean (top row) and covariance (bottom row), computed with* $N = 100, 200$, *and* 400 *observation times in the first, second, and third columns, respectively. The computational cost is measured in computer runtime.*

4.2. Drift-alternating GBM. We next consider the SDE

$$du(t+n) = \begin{cases} \sigma^2 u(t+n)dt + \sigma u(t+n)dW(t+n) & \text{if } n \text{ is even} \\ \sigma u(t+n)dW(t+n) & \text{else} \end{cases} \quad \text{for } t \in (0,1) \tag{4.4}$$

and with the initial condition $u(0) = 1$. This equation is analytically tractable as well, and the solution of the transformed equation $z = \log u$ is given via Itô's formula by

$$dz(t+n) = (-1)^n \frac{\sigma^2}{2} dt + \sigma dW(t+n).$$

Defining $\xi_n \sim N(0, \sigma^2)$ i.i.d., one has that

$$z_{n+1} = z_n + (-1)^n \frac{\sigma^2}{2} + \xi_n =: \Psi_n(z_n) \quad \text{with} \quad z_0 = \log u_0 = 0,$$

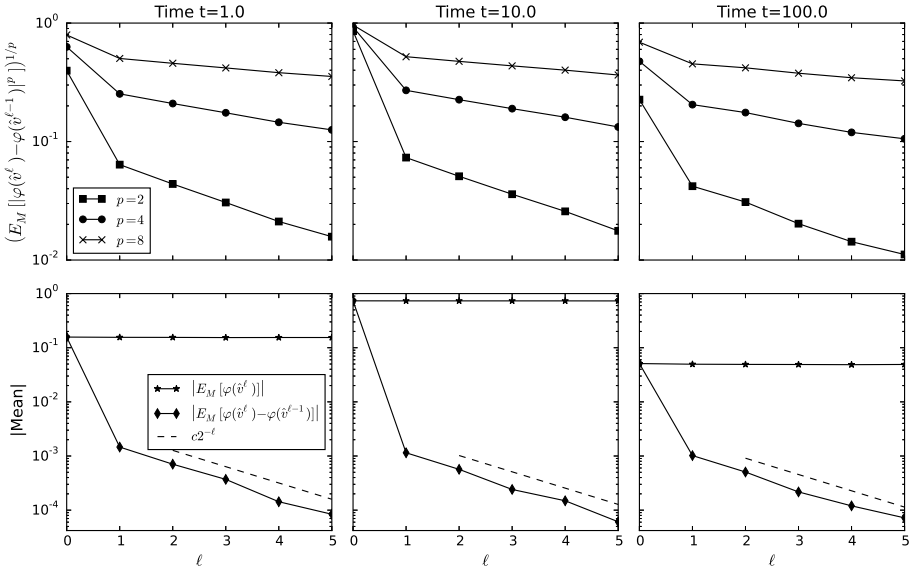

FIG. 3. *Numerical estimates of the decay rates over a sequence of times for the problem presented in section* 4.1 *when approximating the excedence probability* $\mathbb{P}(\widehat{u}_n > 0)$*, i.e., with* $\varphi(v) = \mathbf{1}\{v > 0.1\}$*. The computations use* $M = 10^6$ *particles on every level.*

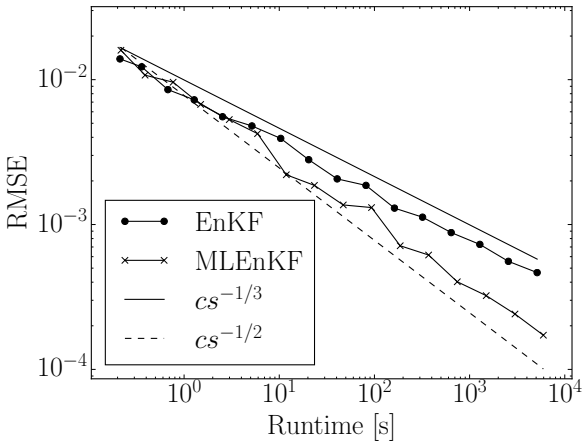

FIG. 4. *Accuracy versus computational cost comparison of the EnKF and MLEnKF methods on the filtering problem presented in section* 4.1 *when approximating the excedence probability* $\mathbb{P}(\widehat{u}_n > 0.1)$ *over* $N = 200$ *observation times. The error is measured in terms of the RMSE and the computational cost in computer runtime.*

and the solution of (4.4) can be obtained via exponentiation: $u_n = e^{z_n}$. Moreover, noisy observations for u_n are introduced on the form

$$\tilde{y}_n = u_n e^{\eta_n},$$

and $\eta_n \sim N(0, \Gamma)$ i.i.d., which, upon defining $y_n^{\text{obs}} = \log \tilde{y}_n$, yields the following relation to noisy observations of z_n:

$$y_n^{\text{obs}} = z_n + \eta_n.$$

As the SDE (4.4) does not fulfill the linear Gaussian constraints (2.6) but $z = \log u$ does, we will here update the ensemble of $z = \log u$ processes. However, to add some artificial difficulty to the problem, the numerical integration is done on the u ensemble:

(i) Numerically integrate a (multilevel or single level) ensemble $u_{n-1} \to u_n$.
(ii) Compute sample mean and covariance of $z_n | Y_{n-1}^{\text{obs}}$ using the $z_n = \log u_n$ ensemble.
(iii) Update the ensemble z_n by the new information provided by the observation y_n^{obs}.
(iv) Compute the initial condition for the ensemble $u_n = e^{z_n}$ and return to (i).

Remark 5. The numerical integration of the GBM process in step (i) above introduces an artificial difficulty in the filtering problem since the integration may by other means be solved exactly. In practice, this of course does not make sense, but our purpose here is simply to numerically validate the performance of the MLEnKF method on a set of simple filtering problems for which reference solutions exist.

Numerical integration of u_n is done by the hierarchy of Euler–Maruyama schemes introduced in (4.2) (applied to the GBM problem, the schemes are Euler–Maruayama, while applied to problems with additive noise, the schemes are Milstein), here with the slightly finer mesh hierarchy $\Delta t^\ell = 2^{-3+\ell}$, since the problem is less stable. The covariance parameters are set to $\sigma = 1/4$ and $\Gamma = 1/16$, $N = 200$, and the numerical method yields the rates $\alpha = 1$, $\beta = 1$ (and $\gamma = 1$). See Figure 5 for a numerical verfication of these decay rates over a sequence of times. In Figure 6, the gold standard mean and covariance of z_n has been approximated by the filtering methods. We observe an RMSE decay rate slightly slower than $\mathcal{O}(s^{-1/2})$ for MLEnKF and $\mathcal{O}(s^{-1/3})$ for EnKF, where s denotes runtime in seconds.

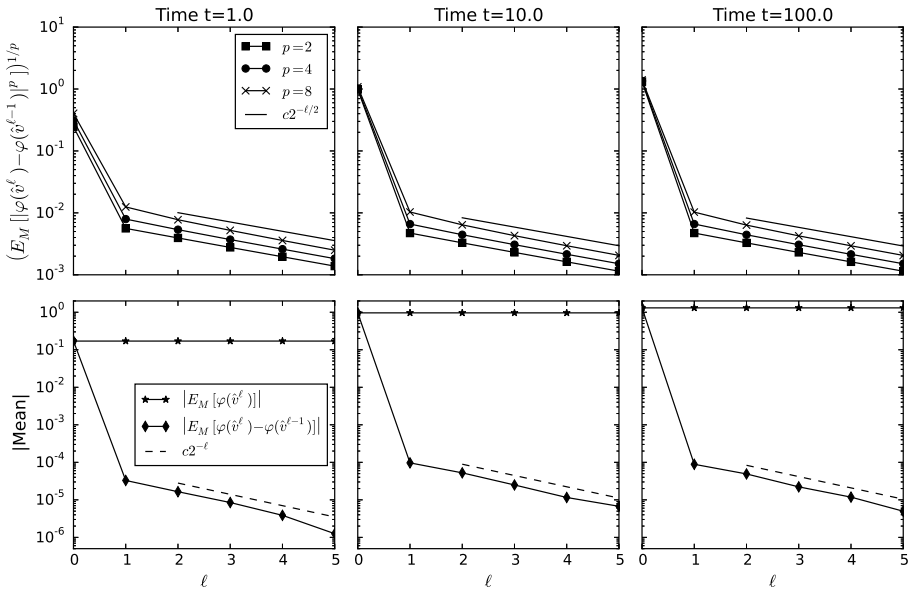

FIG. 5. *Numerical estimates of the decay rates over a sequence of times for the problem presented in section* 4.2 *with* $\varphi(v) = v$. *The computations use* $M = 10^6$ *particles on every level.*

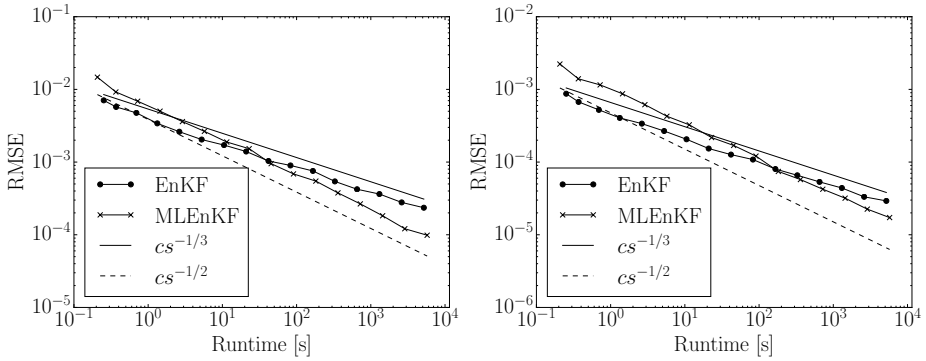

FIG. 6. *Accuracy versus computational cost comparison of the EnKF and MLEnKF methods on the filtering problem presented in section* 4.2. *The error is measured in terms of the RMSE* (4.3) *for the mean (left plot) and covariance (right plot), and the computational cost is measured in computer runtime.*

5. Conclusion. A first attempt, to the knowledge of the authors, at filtering using an MLMC approach is considered in the present work. A proof based on induction of the optimality of the cost as a function of the error, or equivalently the error as a function of the cost, is given. This shows that an optimality result, which is slightly penalized with respect to the vanilla Monte Carlo result, can extend to the case of sequential inference. There is either a logarithmic term which grows with the number of steps or a slightly higher rate $\mathcal{O}(\varepsilon^{-2-\delta})$ for $\delta > 0$. The EnKF is considered, which is consistent only in the case of a linear Gaussian model. However, the mean-field limiting equation may be viewed as a one-step optimal linear (in the observation) filter, and the convergence to this limiting distribution has the desired rate for a more general class of models. Since this work was submitted, two papers have appeared which consider consistent nonlinear filters for similar models. The paper [27] considered multilevel particle filters with optimally coupled multinomial resampling and found the rate of strong convergence β is effectively reduced by a factor of 2 as a result of the resampling. The paper [21] considered multilevel ensemble transform particle filters, which use an optimally coupled deterministic transformation in place of the standard random resampling mechanism, and numerical results indicated the rate may be reduced in some cases and the same in others.

REFERENCES

[1] J. L. ANDERSON, *Localization and sampling error correction in ensemble Kalman filter data assimilation*, Monthly Weather Rev., 140 (2012), pp. 2359–2371.
[2] R. AVIKAINEN, *On irregular functionals of SDEs and the Euler scheme*, Finance Stoch., 13 (2009), pp. 381–401.
[3] A. BAIN AND D. CRISAN, *Fundamentals of Stochastic Filtering*, Springer, New York, 2009.
[4] P. J BICKEL AND E. LEVINA, *Covariance regularization by thresholding*, Ann. Statist., 36 (2008), pp. 2577–2604.
[5] P. J BICKEL AND E. LEVINA, *Regularized estimation of large covariance matrices*, Ann. Statist., 36 (2008), pp. 199–227.
[6] C. BIERIG AND A. CHERNOV, *Convergence analysis of multilevel Monte Carlo variance estimators and application for random obstacle problems*, Numer. Math., 130 (2015), pp. 579–613.
[7] G. BURGERS, P. JAN VAN LEEUWEN, AND G. EVENSEN, *Analysis scheme in the ensemble Kalman filter*, Monthly Weather Rev., 126 (1998), pp. 1719–1724.
[8] O. CAPPÉ, E. MOULINES, AND T. RYDÉN, *Inference in Hidden Markov Models*, Springer, New York, 2005.

[9] J. Carlsson, K. S. Moon, A. Szepessy, R. Tempone, and G. Zouraris, *Stochastic differential equations: Models and numerics*, Lecture notes, KTH, Royal Institute of Technology, 2010, https://people.kth.se/~szepessy/sdepde.pdf.
[10] K. A. Cliffe, M. B. Giles, R. Scheichl, and A. L. Teckentrup, *Multilevel Monte Carlo methods and applications to elliptic PDEs with random coefficients*, Comput. Vis. Sci., 14 (2011), pp. 3–15.
[11] N. Collier, A.-L. Haji-Ali, F. Nobile, E. von Schwerin, and R.Tempone, *A continuation multilevel Monte Carlo algorithm*, BIT, 55 (2015), pp. 399–432.
[12] P. Del Moral, *Feynman-Kac Formulae: Genealogical and Interacting Particle Systems with Applications*, Springer, New York, 2004.
[13] A. Doucet, S. Godsill, and C. Andrieu, *On sequential Monte Carlo sampling methods for Bayesian filtering*, Statist. Comput., 10 (2000), pp. 197–208.
[14] G. Evensen, *Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics*, J. Geophys. Res. Oceans, 99 (1994), pp. 10143–10162.
[15] G. Evensen, *The ensemble Kalman filter: Theoretical formulation and practical implementation*, Ocean Dynamics, 53 (2003), pp. 343–367.
[16] R. P. Fedorenko, *A relaxation method for solving elliptic difference equations*, Zh. Vychisl. Mat. Mat. Fiz., 1 (1961), pp. 922–927.
[17] M. Giles, *Improved multilevel Monte Carlo convergence using the Milstein scheme*, in Monte Carlo and Quasi-Monte Carlo Methods 2006, A. Keller, S. Heinrich, and H. Niederreiter, eds., Springer, Berlin, 2008, pp. 343–358.
[18] M. B. Giles, *Multilevel Monte Carlo path simulation*, Oper. Res., 56 (2008), pp. 607–617.
[19] M. B. Giles and L. Szpruch, *Antithetic multilevel Monte Carlo estimation for multi-dimensional SDEs without Lévy area simulation*, Ann. Appl. Probab., 24 (2014), pp. 1585–1620.
[20] C. Graham and D. Talay, *Stochastic Simulation and Monte Carlo Methods: Mathematical Foundations of Stochastic Simulation*, Stoch. Model. Appl. Probab. 68, Springer, Heidelberg, 2013.
[21] A. Gregory, C. Cotter, and S. Reich, *Multilevel Ensemble Transform Particle Filtering*, preprint, arXiv:1509.00325, 2015.
[22] A. Gut, *Probability: A Graduate Course*, Probab. Theory. Stochastic Process. 200, Springer, Berlin, 2005.
[23] W. Hackbusch, *Multi-Grid Methods and Applications*, Springer Ser. Comput. Math. 4, Springer-Verlag, Berlin, 1985.
[24] S. Heinrich, *Multilevel Monte Carlo Methods*, in Large-Scale Scientific Computing, Springer, New York, 2001, pp. 58–67.
[25] V. H. Hoang, C. Schwab, and A. M. Stuart, *Complexity analysis of accelerated MCMC methods for Bayesian inversion*, Inverse Problems, 29 (2013), 085010.
[26] H. Hoel, E. von Schwerin, A. Szepessy, and R. Tempone, *Implementation and analysis of an adaptive multilevel Monte Carlo algorithm*, Monte Carlo Methods Appl., 20 (2014), pp. 1–41.
[27] A. Jasra, K. Kamatani, K. J. H. Law, and Y. Zhou, *Multilevel Particle Filter*, preprint, arXiv:1510.04977, 2015.
[28] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*, vol. 63, Academic Press, New York, 1970.
[29] R. E. Kalman et al., *A new approach to linear filtering and prediction problems*, J. Basic Engineering, 82 (1960), pp. 35–45.
[30] E. Kalnay, *Atmospheric Modeling, Data Assimilation and Predictability*, Cambridge University Press, Cambridge, UK, 2003.
[31] C. Ketelsen, R. Scheichl, and A. L. Teckentrup, *A Hierarchical Multilevel Markov Chain Monte Carlo Algorithm with Applications to Uncertainty Quantification in Subsurface Flow*, preprint, arXiv:1303.7343, 2013.
[32] P. E. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations*, Appl. Math. (N. Y.) 23, Springer-Verlag, Berlin, 1992.
[33] K. Law, A. Stuart, and K. Zygalakis, *Data Assimilation: A Mathematical Introduction*, Springer-Verlag, Berlin, 2015.
[34] K. J. H. Law, H. Tembine, and R. Tempone, *Deterministic Mean-Field Ensemble Kalman Filtering*, preprint, arXiv:1409.0628v4, 2014.
[35] F. Le Gland, V. Monbet, V.-D. Tran, et al., *Large sample asymptotics for the ensemble Kalman filter*, in The Oxford Handbook of Nonlinear Filtering, Oxford University Press, New York, 2011, pp. 598–631.

[36] O. Ledoit and M. Wolf, *A well-conditioned estimator for large-dimensional covariance matrices*, J. Multivariate Anal., 88 (2004), pp. 365–411.
[37] D. G. Luenberger, *Optimization by Vector Space Methods*, Wiley, New York, 1968.
[38] J. Mandel, L. Cobb, and J. D. Beezley, *On the convergence of the ensemble Kalman filter*, Appl. Math., 56 (2011), pp. 533–541.
[39] S. Mishra and C. Schwab, *Sparse tensor multi-level Monte Carlo finite volume methods for hyperbolic conservation laws with random initial data*, Math. Comp., 81 (2012), pp. 1979–2018.
[40] O. Pajonk, B. V. Rosić, A. Litvinenko, and H. G. Matthies, *A deterministic filter for non-Gaussian Bayesian estimation: Applications to dynamical system estimation with noisy measurements*, Phys. D, 241 (2012), pp. 775–788.
[41] Y.-F. Ren and H.-Y. Liang, *On the best constant in Marcinkiewicz-Zygmund inequality*, Statist. Probab. Lett., 53 (2001), pp. 227–233.