MULTILEVEL ENSEMBLE KALMAN FILTERING BASED ON A SAMPLE AVERAGE OF INDEPENDENT ENKF ESTIMATORS

Håkon Hoel

Chair of Mathematics for Uncertainty Quantification
RWTH Aachen University, Aachen, Germany

Gaukhar Shaimerdenova*

Applied Mathematics and Computational Sciences
KAUST, Thuwal, Saudi Arabia

Raúl Tempone

Chair of Mathematics for Uncertainty Quantification
RWTH Aachen University, Aachen, Germany
Applied Mathematics and Computational Sciences
KAUST, Thuwal, Saudi Arabia

(Communicated by Kody Law)

ABSTRACT. We introduce a new multilevel ensemble Kalman filter method (MLEnKF) which consists of a hierarchy of independent samples of ensemble Kalman filters (EnKF). This new MLEnKF method is fundamentally different from the preexisting method introduced by Hoel, Law and Tempone in 2016, and it is suitable for extensions towards multi-index Monte Carlo based filtering methods. Robust theoretical analysis and supporting numerical examples show that under appropriate regularity assumptions, the MLEnKF method has better complexity than plain vanilla EnKF in the large-ensemble and fine-resolution limits, for weak approximations of quantities of interest. The method is developed for discrete-time filtering problems with finite-dimensional state space and linear observations polluted by additive Gaussian noise.

1. **Introduction.** We develop a new multilevel ensemble Kalman filter method (MLEnKF) for the setting of finite-dimensional state space and discrete-time partial observations polluted by additive Gaussian noise. Our method makes use of recent hierarchical variance-reduction techniques [20, 14, 18] to improve the asymptotic efficiency of weak approximations of filtering distributions compared to standard ensemble Kalman filtering (EnKF). We consider settings where solutions of nonlinear dynamics models must be approximated by numerical methods.

The herein introduced MLEnKF method consists of a hierarchy of independent samples of pairwise-coupled EnKF estimators where, in particular, the Kalman gains of every EnKF sample thus also are independent. Our method is fundamentally different from the original MLEnKF [22], which consists of a hierarchy of

---

* Corresponding author: Gaukhar Shaimerdenova.

coupled ensembles on different resolution levels, all sharing one global "multilevel" Kalman gain.

The motivations for developing the new MLEnKF method are threefold. First, the method is closer to classic EnKF, and we therefore believe it will be easier to implement for practitioners. Second, imposing slightly stricter regularity assumptions, we can prove better asymptotic efficiency results for this method than for the original one. And third, creating a rigorous convergence theory for the new MLEnKF method is a stepping stone towards a multi-index Ensemble Kalman filtering (MIEnKF) method; See [19] for highly efficient approximations of McKean–Vlasov dynamics by the multi-index Monte Carlo method, and Appendix D for a sketch of the said extension to MIEnKF.

The main theoretical contributions of this work are Theorems 1 and 2, which respectively derive L^p-convergence rates for weak approximations in the large-ensemble and fine-numerical-resolution limits with EnKF. Theorem 2 is novel, and while Theorem 1 is similar to [22, Theorem 3.11]; but, to the best of our knowledge, this is the first fully proved convergence result for EnKF in the said limits (i.e., in both limits simultaneously). Estimates for EnKF's and MLEnKF's asymptotic computational cost versus accuracy for the respective methods are provided in Corollaries 1 and 2, respectively. From these estimates we conclude that MLEnKF asymptotically outperforms EnKF whenever Assumptions 1 and 2 hold.

1.1. **Literature review.** The EnKF method was first introduced in the seminal work [12], and due to its ease of use and impressive performance in high dimensions, it quickly became a popular method for weather prediction, ocean-atmosphere science, and oil reservoir simulations [29, 26, 1]. The standard version of EnKF with perturbed observations, which is the method we will study and extend to the multi-level setting in this work, first appeared in [25] and ensuing analysis [6] showed that adding artificial noise to the observations may be viewed as a consistency step for avoiding covariance deflation of the empirical filtering distribution. L^p-convergence of the first and second sample moments in the large-ensemble limit for EnKF was first treated by a short argument in [37] for the linear model setting and the result was subsequently extended to a set of nonlinear filtering problems by a more technical argument in [36] (in the sense of deriving L^p-convergence rates for weak approximations of sufficiently smooth quantities of interest).

The multilevel Monte Carlo (MLMC) method was introduced for efficient weak approximations of random fields in [20] and for stochastic differential equations in [14]. The MLEnKF method was developed for finite-dimensional settings in [22] and countable, infinite-dimensional settings in [7]. In [13] a multilevel hybrid EnKF method was developed for solving reservoir history matching problems. Multi-level particle filters using a multilevel-coupled resampling algorithm was introduced in [27]. The multilevel transform particle filter, applying optimal transportation mapping in the analysis step, was treated in [17, 16]. In the context of Bayesian inverse problems, multilevel sequential Monte Carlo methods have been studied in [3, 2, 38, 34] and a multilevel Markov Chain Monte Carlo method was developed in [10]. The multi-fidelity Monte Carlo method [39] is a recent and close kin of MLMC that differs from MLMC by having fixied its estimator's finest resolution level and by having more extensive coupling of samples than only pairwise. This often reduces the estimator's statistical error even more effectively than MLMC. A multi-fidelity Monte Carlo EnKF method was developed in [40].

Under sufficient regularity, EnKF converges to the so-called mean-field Kalman filter in the large-ensemble limit. In nonlinear-dynamics or observation settings, however, the mean-field Kalman filter is not equal to the Bayes filter [35, 36, 11, 41]. Due to this discrepancy between EnKF and the Bayes filter even in the large-ensemble limit, due to the large uncertainty in the estimation of the model error and due to the constraints imposed by challenging high-dimensional problems and limited computational budgets, a considerable number of works have, instead of studying the large-ensemble limit, focused on the large-time and/or continuous-time limit of the fixed-ensemble-size EnKF cf. [30, 46, 42, 43, 5, 9, 33]. However, a recurring problem for fixed-ensemble-size EnKF is to determine how large the ensemble ought to be to equilibrate the model error with the other error contributions (statistical error and bias). And the large-ensemble limit convergence rates for EnKF and MLEnKF that are presented in this work may be helpful for determining the ensemble size of a finite-ensemble EnKF method.

1.2. **Organization of this work.** Section 2 describes the problem, the MLEnKF method, and L^p-convergence results for EnKF and MLEnKF towards the mean-field EnKF. Section 3 presents numerical performance studies of EnKF and MLEnKF for two different problems. Appendix A presents a brief overview of the mean-field EnKF method. Appendix B contains proofs of the main theoretical results, Appendix C describes an algorithm for obtaining pseudo-reference solutions for nonlinear filtering problems, and Appendix D sketches an extension from MLEnKF to multi-index EnKF.

2. Problem setting and main results.

2.1. **Problem setting.** Let $(\Omega, (\mathcal{F}_t), \mathcal{F} = \mathcal{F}_\infty, \mathbb{P})$ denote a complete filtered probability space, and for any $k \in \mathbb{N}$ and $p \geq 1$, let $L_t^p(\Omega, \mathbb{R}^k)$ denote the space of $\mathcal{F}_t \backslash \mathcal{B}^k$–measurable functions $u : \Omega \to \mathbb{R}^k$ such that $\mathbb{E}[|u|^p] < \infty$. Here, $\mathcal{B}^k$ represents the Borel σ-algebra on $\mathbb{R}^k$ and u is said to be $\mathcal{F}_t \backslash \mathcal{B}^k$–measurable if and only if $u^{-1}(B) \in \mathcal{F}_t$ for all $B \in \mathcal{B}^k$. For a given state-space dimension $d \in \mathbb{N}$ and initial data $u_0 \in \cap_{p \geq 2} L_0^p(\Omega, \mathbb{R}^d)$, we consider the discrete stochastic dynamics for $n = 0, 1, \ldots$

$$u_{n+1}(\omega) = \Psi_n(u_n, \omega), \qquad \omega \in \Omega, \tag{1}$$

for a sequence of mappings $\Psi_n : \mathbb{R}^d \times \Omega \to \mathbb{R}^d$. The dynamics is associated with the stochastic differential equation (SDE)

$$\Psi_n(u_n, \omega) = u_n + \int_n^{n+1} a(u_t)dt + \int_n^{n+1} b(u_t)dW_t(\omega), \tag{2}$$

with coefficients $a : \mathbb{R}^d \to \mathbb{R}^d$, $b : \mathbb{R}^d \to \mathbb{R}^{d \times d_W}$ and the driving noise $W : [0, \infty) \times \Omega \to \mathbb{R}^{d_W}$ denoting a d_W-dimensional standard Wiener process. We further assume the coefficients a and b are sufficiently smooth so that

$$u_n \in L_n^2(\Omega, \mathbb{R}^d) \implies u_{n+1} \in L_{n+1}^2(\Omega, \mathbb{R}^d),$$

and we note, from now on suppressing the dependence on ω whenever confusion is not possible, that u_{n+1} may be expressed as the n-fold composition of Ψ with itself evaluated at u_0:

$$u_{n+1} = \Psi_n \circ \Psi_{n-1} \circ \cdots \circ \Psi_0(u_0).$$

Associated with the dynamics (1), there exists a series of noisy observations

$$y_n = H u_n + \eta_n, \quad n = 1, 2, \ldots, \tag{3}$$

where $H \in \mathbb{R}^{d_O \times d}$ and $\eta_1, \eta_2, \ldots$ is a sequence of independent and identically distributed (iid) random variables satisfying $\eta_1 \sim N(0, \Gamma)$ with positive definite covariance matrix $\Gamma \in \mathbb{R}^{d_O \times d_O}$ and the independence property $\{\eta_j\}_{j \geq 1} \perp \{u_k\}_{k \geq 0}$. The filtration $\mathcal{F}_t$ is the completion of the smallest σ-algebra generated by u_0, $\{W_s\}_{s \in [0,t]}$ and $\{\eta_j\}_{j=1}^{\lfloor t \rfloor}$, with $\lfloor t \rfloor = \max\{k \in \mathbb{Z} \mid k \leq t\}$. That is,

$$\mathcal{F}_t = \overline{\sigma\Big(\sigma(u_0) \cup \sigma(\{W_s\}_{s \in [0,t]}) \cup \sigma(\{\eta_j\}_{j=1}^{\lfloor t \rfloor})\Big)}.$$

with $\overline{\sigma}$ denoting the completion of σ.

The series of observations up to time k is denoted by

$$Y_k := \begin{cases} (y_1, y_2, \ldots, y_k) & \text{if} \quad k \geq 1, \\ \emptyset & \text{if} \quad k = 0. \end{cases} \tag{4}$$

For notational convenicence and since no observations have been made at time 0, we defined $Y_0 = \emptyset$ above. In consistency with the propery that Y_0 holds no information we write $\mathbb{P}_{u_0|Y_0} := \mathbb{P}_{u_0}$. The Bayes filter is a sequential procedure for determining the (conditional) prediction $\mathbb{P}_{u_k|Y_{k-1}}$ and update $\mathbb{P}_{u_k|Y_k}$. Assuming that the densities of said distributions exist, the stochastic dynamics (1) and the conditional independence $(u_k|u_{k-1}) \perp Y_{k-1}$ yield the proportionality

$$\rho_{u_k|Y_{k-1}}(u) = \int_{\mathbb{R}^d} \rho_{u_k, u_{k-1}|Y_{k-1}}(u, v) dv \propto \int_{\mathbb{R}^d} \rho_{u_k|u_{k-1}}(u) \rho_{u_{k-1}|Y_{k-1}}(v) \, dv,$$

and Bayesian inference implies that

$$\rho_{u_k|Y_k}(u) \propto \mathcal{L}_{u_k|y_k}(u) \rho_{u_k|Y_{k-1}}(u),$$

where the likelihood function is given by

$$\mathcal{L}_{u_k|y_k}(u) = \frac{\exp(-|\Gamma^{-1/2}(y_k - Hu)|^2/2)}{\sqrt{(2\pi)^p |\det(\Gamma)|}}.$$

In other words, if the updated distribution $\mathbb{P}_{u_{k-1}|y_{k-1}}$ is known and suitable regularity assumptions hold, then the prediction and updated distribution at the next time can be computed up to a proportionality constant (although this step may be computationally intractable).

In settings where u_0 is a Gaussian random variable and the dynamics Ψ is linear with additive Gaussian noise, the Kalman filter [28] solves the above filtering problem exactly. When Ψ is nonlinear, however, it is often not possible to solve the filtering problem exactly and one must resort to approximation methods. EnKF is a nonlinear and ensemble-based filtering method that preserves that tends to be particularly efficient when the "true" dimension of a filtering problem is far smaller than the state-space dimension.

For linear-Gaussian filtering problems, the empirical measure of EnKF converges towards the exact (Bayes filter) density $\mathbb{P}_{u_k|Y_k}$ in the large-ensemble limit [37]. More generally, for instance when Ψ is nonlinear, EnKF converges to the mean-field EnKF in the large-ensemble limit, cf. Section A. As a consequence of EnKF employing a Gaussian-like update of its particles, the mean-field EnKF will in many cases not be equal to the Bayes filter [35, 11]. But, to the best of our knowledge, there does not exist any thorough scientific comparison of the mean-field EnKF and the Bayes Filter, and Figure 1 shows that for the nonlinear dynamics Ψ defined by the SDE

$$du = -(u + \pi \cos(\pi u/5)/5)dt + \sigma dW \tag{5}$$

and (1), the dissipative/contractive properties of the associated Fokker-Planck equation can produce prediction densities for the respective filtering methods that are indistinguishable to the naked eye.
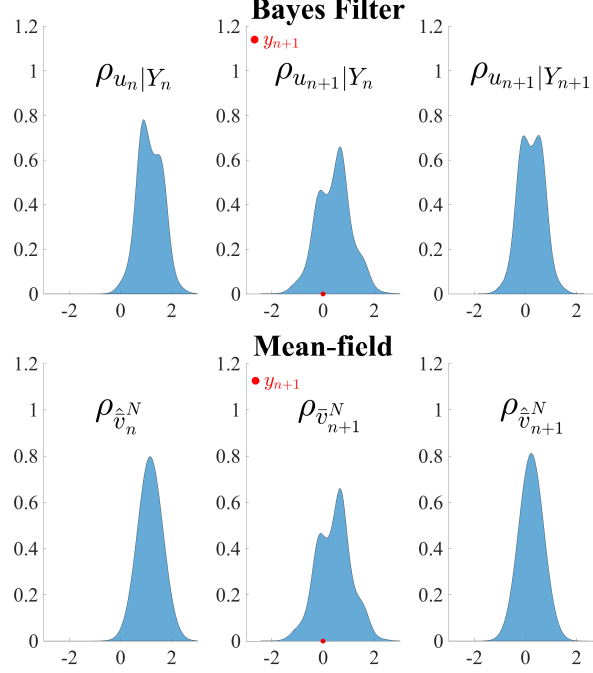

FIGURE 1. Illustration, based on the nonlinear dynamics (5), of the contracting property which can produce almost identical prediction densities (middle panels) for the Bayes filter and MFEnKF even when the preceding updated densities differ notably.

Objective. For a given quantity of interest (QoI) $\varphi : \mathbb{R}^d \to \mathbb{R}$, our objective is to construct an efficient filtering method for computing

$$\mathbb{E}^{\bar{\mu}_n}[\varphi(u)] = \int_{\mathbb{R}^d} \varphi(u)\bar{\mu}_n(du), \tag{6}$$

where $\bar{\mu}_n$ denotes the mean-field EnKF updated measure at time n, cf. Appendix A. In the best of worlds, one would rather seek a method computing the exact expectation of φ with respect to the Bayes filter posterior, i.e., $\mathbb{E}[\varphi(u_n) \mid Y_n]$, but since the said posterior $\mathbb{P}_{u_n|Y_n}$ is generally not attainable by EnKF-based filtering methods, and since both EnKF and MLEnKF converge weakly towards mean-field EnKF, cf. Theorems 1 and 2, we will in this work focus on the simpler (but still very challenging) goal (6).

Notation 1.

- *For $f, g : (0, \infty) \to [0, \infty)$ the notation $f \lesssim g$ implies there exists a $C > 0$ such that*

$$f(x) \leq Cg(x), \quad \forall x \in (0, \infty).$$

- *The notation $f \eqsim g$ implies that $f \lesssim g$ and $g \lesssim f$.*

- *For $r, s \in \mathbb{N}$, $|x|$ denotes the Euclidean norm of a vector $x \in \mathbb{R}^s$ and for $A \in \mathbb{R}^{r \times s}$, $|A|_2 := \sup_{|x|=1} |Ax|$.*
- *For $\mathcal{F} \backslash \mathcal{B}^d$-measurable functions $u : \Omega \to \mathbb{R}^d$ and $p \geq 1$,*

$$\|u\|_p := \|u\|_{L^p(\Omega, \mathbb{R}^d)} = \left(\int_\Omega |u(\omega)|^p \, \mathbb{P}(d\omega) \right)^{1/p}.$$

- *For any $\kappa \in \mathbb{N}_0^r$, with $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$, and any sufficiently smooth functions of the form $f : \mathbb{R}^r \to \mathbb{R}$ and $g : \mathbb{R}^r \times \Omega \to \mathbb{R}$, the respective ∂^κ-partial derivatives are defined by*

$$\partial^\kappa f(x) = \frac{\partial^{|\kappa|_1} f}{\partial x_1^{\kappa_1} \ldots \partial x_r^{\kappa_d}}(x) \quad and \quad \partial^\kappa g(x, \omega) = \frac{\partial^{|\kappa|_1} g}{\partial x_1^{\kappa_1} \ldots \partial x_r^{\kappa_d}}(x, \omega)$$

  for all $x \in \mathbb{R}^r$ and $\mathbb{P}-$almost all $\omega \in \Omega$, where $|\kappa|_1 := \sum_{i=1}^r \kappa_i$. This extends to vector-valued mappings of the form $f : \mathbb{R}^r \to \mathbb{R}^s$ by component-wise partial derivatives: $\partial^\kappa f(x) = (\partial^\kappa f_1(x), \partial^\kappa f_2(x), \ldots, \partial^\kappa f_s(x))$, and so on.
- *For $k \in \mathbb{N}_0$, $C_B^k(\mathbb{R}^r, \mathbb{R}^s)$ denotes the space of k-times continuously differentiable functions $\eta : \mathbb{R}^r \to \mathbb{R}^s$ for which η and all of its partial derivatives of order up to and including k is uniformly bounded.*
- *$C_P^k(\mathbb{R}^r, \mathbb{R}^s)$ denotes the space of k-times continuously differentiable functions whose partial derivatives of order up to and including k have polynomial growth.*
- *For a mapping $\eta : \mathbb{R}^r \to \mathbb{R}^s$, its Jacobian is denoted $D\eta$, and its Hessian is $D^2\eta$.*
- *$\mathbb{R}_r[x_1, \ldots, x_d]$ denotes the set of polynomials in d variables that are of total degree smaller or equal to $r \in \mathbb{N}_0$ and have coefficients in $\mathbb{R}$.*

2.2. **Numerical approximation of the stochastic dynamics.** We assume that for every time $n \geq 0$, there exists a collection of progressively more accurate numerical solvers $\{\Psi_n^N : \mathbb{R}^d \times \Omega \to \mathbb{R}^d\}_{N \in \mathbb{N}}$ satisfying the following assumptions:

Assumption 1.
The initial distribution $u_0 \in \cap_{p \geq 2} L_0^p(\Omega, \mathbb{R}^d)$ and for any $n \in \mathbb{N}_0$,

(i) *for any $p \geq 2$, there exists a $c_p > 0$ such that*

$$\left\| \Psi_n^N(u) \right\|_p \leq c_p (1 + \|u\|_p) \quad \forall \big(N \in \mathbb{N} \quad \& \quad u \in \cap_{p \geq 2} L_n^p(\Omega, \mathbb{R}^d) \big);$$

(ii) *there exists an $\alpha > 0$ and a set of mappings $\mathbb{F}$ with $\mathbb{R}_2[x_1, \ldots, x_d] \subset \mathbb{F} \subset C_P^2(\mathbb{R}^d, \mathbb{R})$ such that if*

$$\left| \mathbb{E} \left[\varphi(u^N) - \varphi(u) \right] \right| \leq c_\varphi N^{-\alpha} \quad \forall (N \geq 1 \quad \& \quad \varphi \in \mathbb{F}),$$

  for some $u \in \cap_{p \geq 2} L_n^p(\Omega, \mathbb{R}^d)$ and $\{u^N\}_N \subset \cap_{p \geq 2} L_n^p(\Omega, \mathbb{R}^d)$ and (observable-dependent constant) $c_\varphi > 0$, then there exists another (observable-dependent constant) $\tilde{c}_\varphi > 0$ such that

$$\left| \mathbb{E} \left[\varphi(\Psi_n^N(u^N)) - \varphi(\Psi_n(u)) \right] \right| \leq \tilde{c}_\varphi N^{-\alpha} \quad \forall (N \geq 1 \quad \& \quad \varphi \in \mathbb{F});$$

(iii) *for any $p \geq 2$, there exists a $c_p > 0$ such that*

$$\left\| \Psi_n^N(u) - \Psi_n^N(v) \right\|_p < c_p \|u - v\|_p \quad \forall \big(N \in \mathbb{N} \quad \& \quad u, v \in \cap_{p \geq 2} L_n^p(\Omega, \mathbb{R}^d) \big);$$

(iv) *there exists a $c_3 > 0$ such that the computational cost of the numerical solution Ψ_n^N is $\mathbb{P}$-almost surely bounded by*

$$\text{Cost}(\Psi_n^N(u)) \leq c_3 N \quad \forall\big(N \in \mathbb{N} \quad \& \quad u \in \cap_{p \geq 2} L_n^p(\Omega, \mathbb{R}^d)\big);$$

Remark 1. If the dynamics (1) is given by the SDE (2) where the coefficients $a : \mathbb{R}^d \to \mathbb{R}^d$, $b : \mathbb{R}^d \to \mathbb{R}^{d \times d_W}$ satisfy $\partial^\kappa a_j, \partial^\kappa b_{jk} \in C_B(\mathbb{R}^d, \mathbb{R})$ for all $1 \leq j \leq d$, $1 \leq k \leq d_W$ and $|\kappa| \geq 1$, and if Ψ_n^N denotes the Euler–Maruyama numerical solution of (2) using N uniform timesteps, then Assumption 1 holds with $\alpha = 1$ and $\mathbb{F} = C_P^4(\mathbb{R}^d, \mathbb{R})$, cf. [15, Chapter 7] and [31, Thm 14.5.2]. Noting that the uniformly-bounded constraint is only imposed on partial derivatives of the coefficients, and not the coefficients themselves, the stated convergence rates for Euler–Maruyama apply for instance to the nonlinear dynamics (5).

2.3. **EnKF.**

For a solver Ψ^N with fixed resolution $N \geq 1$ and a fixed ensemble size $P \geq 1$, the EnKF method consists of an ensemble of P particles that are iteratively simulated forward and updated in a way that can be viewed as a nonlinear extension of Kalman filtering. The evolution of the EnKF ensemble over the times $n = 0, 1, \ldots$ can be described as follows:

We denote the updated ensemble at time n by $\hat{v}_{n,1:P}^{N,P} := \{\hat{v}_{n,i}^{N,P}\}_{i=1}^P$ and the prediction ensemble at the same time by $v_{n,1:P}^{N,P} := \{v_{n,i}^{N,P}\}_{i=1}^P$. At time 0, $\hat{v}_{0,1:P}^{N,P} := \{\hat{v}_{0,i}^{N,P}\}_{i=1}^P$, consists of P independent and $\mathbb{P}_{u_0|Y_0}$-distributed particles, where we assume the initial distribution can be sampled exactly. The empirical measure induced by the ensemble $\hat{v}_{0,1:P}^{N,P}$ may be viewed as the EnKF approximation of $\mathbb{P}_{u_0|Y_0}$. Given an updated ensemble at time $n \geq 0$, the prediction ensemble at time $n + 1$ is computed by simulating each particle forward

$$v_{n+1,i}^{N,P} = \Psi_n^N(\hat{v}_{n,i}^{N,P}) \quad \text{for} \quad i = 1, 2, \ldots, P.$$

Thereafter, the updated ensemble at time $n + 1$ is computed through assimilating the new measurement y_{n+1} through the Kalman-filter-like and particle-wise formula

$$\hat{v}_{n+1,i}^{N,P} = (I - K_{n+1}^{N,P} H)v_{n+1,i}^{N,P} + K_{n+1}^{N,P} \tilde{y}_{n+1,i} \quad \text{for} \quad i = 1, 2, \ldots, P.$$

Here

$$K_{n+1}^{N,P} = C_{n+1}^{N,P} H^{\mathbf{T}}(H C_{n+1}^{N,P} H^{\mathbf{T}} + \Gamma)^{-1}$$

denotes the Kalman gain,

$$C_{n+1}^{N,P} = \overline{\text{Cov}}[v_{n+1}^{N,P}] := \sum_{i=1}^P \frac{v_{n+1,i}^{N,P} \left(v_{n+1,i}^{N,P}\right)^{\mathbf{T}}}{P} - \sum_{i=1}^P \frac{v_{n+1,i}^{N,P}}{P} \left(\sum_{i=1}^P \frac{v_{n+1,i}^{N,P}}{P}\right)^{\mathbf{T}} \quad (7)$$

denotes the *biased* sample covariance of the ensemble $\hat{v}_{n,1:P}^{N,P}$,

$$\tilde{y}_{n+1,i} = y_{n+1} + \eta_{n+1,i}, \quad \text{for} \quad i = 1, 2, \ldots, P$$

are iid perturbed observations with $\eta_{n+1,1} \sim N(0, \Gamma)$ and $\eta_{j,i} \perp u_k$ for all $j \geq 1$, $k \geq 0$. Perturbed observations were originally introduced in [6] to correct an unwanted covariance-deflation feature in the original formulation of EnKF [12].

We note that for all filtering methods considered in this work, the iteratively augmented observation sequence $y_1, y_2, y_3, \ldots$ is assumed to be given, i.e., *non-random*. The sources of randomness in the EnKF filter are therefore the driving noise in the dynamics and the perturbed observations. For EnKF with dynamics resolution N, randomness thus enter through $\{\Psi_n^{N,P}\}_n$ and $\{\eta_{n,i}\}_{n,i}$. Observe further that

since the particles $\hat{v}_{n,1}^{N,P}, \hat{v}_{n,2}^{N,P}, \ldots, \hat{v}_{n,P}^{N,P}$ are identically distributed, the distribution of $\hat{v}_n^{N,P}$ will depend on the ensemble size P and the model resolution N.

If Assumption 1 holds, then Theorem 1 implies that $\hat{v}_{n,1:P}^{N,P} \subset \cap_{p\geq 2} L_n^p(\Omega, \mathbb{R}^d)$ for any $P \in \mathbb{N}$, $N \in \mathbb{N}$, and $n \in \mathbb{N}_0$. This ensures the existence of the EnKF empirical measure

$$\mu_n^{N,P}(dv) = \frac{1}{P}\sum_{i=1}^{P} \delta(dv; \hat{v}_{n,i}^{N,P}),$$

with the Dirac measure

$$\delta(dv; y) := \begin{cases} 1 & \text{if} \quad y \in dv \\ 0 & \text{otherwise.} \end{cases}$$

Notation 2. *The expectation of a QoI $\varphi : \mathbb{R}^d \to \mathbb{R}$ with respect to the a probability measure $\mu : \mathcal{B}^d \to [0,1]$ is denoted by*

$$\mu[\varphi] := \int_{\mathbb{R}^d} \varphi(v)\,\mu(dv).$$

Note that the expectation with respect to the EnKF empirical measure takes the form of a sample average

$$\mu_n^{N,P}[\varphi] = \frac{1}{P}\sum_{i=1}^{P} \varphi(\hat{v}_{n,i}^{N,P}),$$

and that $\mu_n^{N,P}[\varphi]$ in fact is a random variable since it is a function of the ensemble's P particles $\hat{v}_{n,i}^{N,P}$ (where P obviously is a finite number). For the mean-field measure $\bar{\mu}_n = \mu_n^{\infty,\infty}$, on the other hand, $\bar{\mu}_n[\varphi]$ is a deterministic value; this is a consequence of $P = \infty$.

2.4. **MLEnKF.** The MLEnKF estimator that we introduce in this work is a sample average of pairwise coupled EnKF estimators $\{\mu_n^{N_\ell,P_\ell}[\varphi]\}_{\ell=0}^{L}$ on a hierarchy of resolution levels $\{N_\ell\}_{\ell=0}^{L} \subset \mathbb{N}$ and ensemble-size levels $\{P_\ell\}_{\ell=0}^{L} \subset \mathbb{N}$. Here, L refers to the finest resolution level of the MLEnKF estimator, and we impose the following exponential-growth constraints on the resolution and ensemble-size sequences:

$$P_{\ell+1} = 2P_\ell \quad \text{and} \quad N_\ell \eqsim 2^{s\ell} \quad \text{for some } s > 0.$$

Before writing the explicit form of the MLEnKF estimator, note that since our goal, the value we want to approximate, is deterministic, cf. Notation 2, the following equality holds

$$\bar{\mu}_n[\varphi] = \mu_n^{\infty,\infty}[\phi] = \mathbb{E}\left[\mu_n^{\infty,\infty}[\phi]\right].$$

This motivates the approximation

$$\bar{\mu}_n[\varphi] \approx \mathbb{E}\left[\mu_n^{N_L,P_L}[\phi]\right] = \mathbb{E}\left[\mu_n^{N_0,P_0}[\varphi]\right] + \sum_{\ell=1}^{L} \mathbb{E}\left[\mu_n^{N_\ell,P_\ell}[\varphi] - \mu_n^{N_{\ell-1},P_{\ell-1}}[\varphi]\right],$$

where the last equality holds due to the linearity of the expectation operator. In the expectations for $\ell \geq 1$ we seek to employ pairwise coupling between the fine-level estimator $\mu_n^{N_\ell,P_\ell}[\varphi]$ and the coarse-level estimator $\mu_n^{N_{\ell-1},P_{\ell-1}}[\varphi]$ through associating the driving noise and perturbed observations for each particle on level ℓ uniquely to one particle on level $\ell - 1$. As the number of particles on the neighboring levels are not equal ($P_\ell = 2P_{\ell-1}$), this is clearly not possible in the current form of expectations of differences, so we replace $\mu_n^{N_{\ell-1},P_{\ell-1}}[\varphi]$ by the average of two identically

distributed random variables $\mu_n^{N_{\ell-1},P_{\ell-1},1}[\varphi]$ and $\mu_n^{N_{\ell-1},P_{\ell-1},2}[\varphi]$. Importantly, these satisfy

$$\mathbb{E}\left[\frac{\mu_n^{N_{\ell-1},P_{\ell-1},1}[\varphi] + \mu_n^{N_{\ell-1},P_{\ell-1},2}[\varphi]}{2}\right] = \mathbb{E}\left[\mu_n^{N_{\ell-1},P_{\ell-1}}[\varphi]\right] \quad k = 1,2,$$

which implies that the following equality holds

$$\mathbb{E}\left[\mu_n^{N_L,P_L}[\phi]\right] = \mathbb{E}\left[\mu_n^{N_0,P_0}[\varphi]\right]$$
$$+ \sum_{\ell=1}^{L} \mathbb{E}\left[\mu_n^{N_\ell,P_\ell}[\varphi] - \frac{\mu_n^{N_{\ell-1},P_{\ell-1},1}[\varphi] + \mu_n^{N_{\ell-1},P_{\ell-1},2}[\varphi]}{2}\right].$$

By approximating the ℓ-th the expectation by a sample average using M_ℓ iid pairwise-coupled samples

$$\left\{\mu_n^{N_\ell,P_\ell,m}[\varphi] - \frac{\mu_n^{N_{\ell-1},P_{\ell-1},1,m}[\varphi] + \mu_n^{N_{\ell-1},P_{\ell-1},2,m}[\varphi]}{2}\right\}_{m=1}^{M_\ell},$$

we obtain the MLEnKF estimator

$$\mu^{ML}[\varphi] := \sum_{m=1}^{M_0} \frac{\mu_n^{N_0,P_0,m}[\varphi]}{M_0}$$
$$+ \sum_{\ell=1}^{L}\sum_{m=1}^{M_\ell} \frac{\mu_n^{N_\ell,P_\ell,m}[\varphi]}{M_\ell} - \frac{\left(\mu_n^{N_{\ell-1},P_{\ell-1},1,m} + \mu_n^{N_{\ell-1},P_{\ell-1},2,m}\right)[\varphi]}{2M_\ell}. \tag{8}$$

As for classic MLMC estimators, the degrees of freedom L and $\{M_\ell\}_{\ell=0}^{L} \subset \mathbb{N}$ are determined through a constrained optimization problem where one seeks to equilibrate the variance contribution from all of the sample averages and the bias error cf. Corollary 2. The purpose of pairwise-coupled particles is to reduce the variance of the MLEnKF estimator and thereby improve the efficiency of the method by reducing the number of samples $\{M_\ell\}$ needed to control the approximation error.

To describe the coupling between $\mu_n^{N_\ell,P_\ell}$ and $(\mu_n^{N_{\ell-1},P_{\ell-1},1}, \mu_n^{N_{\ell-1},P_{\ell-1},2})$, we first introduce the following notation for the particles of the three underlying EnKF ensembles: for $i = 1, 2\ldots, P_\ell$,

$$\hat{v}_{n,i}^{\ell,\mathbf{f}} := \hat{v}_{n,i}^{N_\ell,P_\ell}$$
$$\hat{v}_{n,i}^{\ell,\mathbf{c}} := \begin{cases} \hat{v}_{n,i}^{\ell,\mathbf{c}_1} := \hat{v}_{n,i}^{N_{\ell-1},P_{\ell-1},1} & \text{if} \quad i \le P_{\ell-1} \\ \hat{v}_{n,i}^{\ell,\mathbf{c}_2} := \hat{v}_{n,i-P_{\ell-1}}^{N_{\ell-1},P_{\ell-1},2} & \text{if} \quad P_{\ell-1} < i \le P_\ell, \end{cases}$$

where $\hat{v}_{n,i}^{N_{\ell-1},P_{\ell-1},k}$ refers to the i-th updated particle at time n of the EnKF measure $\mu_n^{N_{\ell-1},P_{\ell-1},k}$, for $k = 1, 2$. The superscript $\mathbf{f}$ refers to fine-level particles simulated with numerical resolution N_ℓ whereas $\mathbf{c}$ refers to the coarse-level particles simulated with numerical resolution $N_{\ell-1}$. The set of all the fine- and coarse-level particles are respectively denoted by

$$\hat{v}_{n,1:P_\ell}^{\ell,\mathbf{f}} := \{\hat{v}_{n,i}^{\ell,\mathbf{f}}\}_{i=1}^{P_\ell} \quad \text{and} \quad \hat{v}_{n,1:P_\ell}^{\ell,\mathbf{c}} := \{\hat{v}_{n,i}^{\ell,\mathbf{c}}\}_{i=1}^{P_\ell},$$

with the EnKF ensemble $\hat{v}_{n,1:P_\ell}^{\ell,\mathbf{f}}$ inducing the empirical measure $\mu_n^{\ell,\mathbf{f}} := \mu_n^{N_\ell,P_\ell}$. The set of all the coarse-level particles is a union of two EnKF ensembles with $P_{\ell-1}$ particles in each, namely $\hat{v}_{n,1:P_\ell}^{\ell,\mathbf{c}} = \hat{v}_{n,1:P_{\ell-1}}^{\ell,\mathbf{c}_1} \cup \hat{v}_{n,1:P_{\ell-1}}^{\ell,\mathbf{c}_2}$, where $\hat{v}_{n,1:P_{\ell-1}}^{\ell,\mathbf{c}_1} := \{\hat{v}_{n,i}^{\ell,\mathbf{c}}\}_{i=1}^{P_{\ell-1}}$ and $\hat{v}_{n,1:P_{\ell-1}}^{\ell,\mathbf{c}_2} := \{\hat{v}_{n,i}^{\ell,\mathbf{c}}\}_{i=P_{\ell-1}+1}^{P_\ell}$. The EnKF ensemble $\hat{v}_{n,1:P_\ell}^{\ell,\mathbf{c}_k}$ induces the empirical

measure $\mu_n^{\ell,\mathbf{c}_k} := \mu_n^{N_{\ell-1},P_{\ell-1},k}$ for $k = 1,2$. By defining $\mu_n^{\ell,\mathbf{c}} := \left(\mu_n^{\ell,\mathbf{c}_1} + \mu_n^{\ell,\mathbf{c}_2}\right)/2$ and imposing the convention

$$\mu_n^{0,\mathbf{c}} = \mu_n^{0,\mathbf{c}_1} = \mu_n^{0,\mathbf{c}_2} = 0 \quad \forall n \geq 0,$$

the MLEnKF estimator (8) can be written

$$\mu_n^{\mathrm{ML}}[\varphi] = \sum_{\ell=0}^{L} \sum_{m=1}^{M_\ell} \frac{\mu_n^{\ell,\mathbf{f},m}[\varphi] - \mu_n^{\ell,\mathbf{c},m}[\varphi]}{M_\ell} \tag{9}$$

where the sequence $\{(\mu_n^{\ell,\mathbf{f},m}[\varphi], \mu_n^{\ell,\mathbf{c},m}[\varphi])\}_m$ are iid draws from the joint distribution of coupled random variables $\mathbb{P}_{(\mu_n^{\ell,\mathbf{f}}[\varphi], \mu_n^{\ell,\mathbf{c}}[\varphi])}$.

2.4.1. *One prediction-update iteration of pairwise coupled ensembles.* The fine-level initial ensemble $\{\hat{v}_{0,i}^{\ell,\mathbf{f}}\}_{i=1}^{P_\ell}$ for $\ell = 0, 1, \ldots, L$ is independent and identically $\mathbb{P}_{u_0|Y_0}$-distributed and it is pairwise coupled to the coarse-level initial ensemble through setting $\hat{v}_{0,i}^{\ell,\mathbf{f}} = \hat{v}_{0,i}^{\ell,\mathbf{c}}$. For pairwise-coupled updated-state particles $\hat{v}_{n,i}^{\ell,\mathbf{f}}$ and $\hat{v}_{n,i}^{\ell,\mathbf{c}}$ at time n, we denote by the next-time prediction state respectively by $v_{n+1,i}^{\ell,\mathbf{f}}$ and $v_{n+1,i}^{\ell,\mathbf{c}}$. The prediction state is obtained by simulation of model dynamics on neighboring resolutions:

$$\left.\begin{array}{l} v_{n+1,i}^{\ell,\mathbf{f}} = \Psi_n^{N_\ell}(\hat{v}_{n,i}^{\ell,\mathbf{f}}) \\ v_{n+1,i}^{\ell,\mathbf{c}} = \Psi_n^{N_{\ell-1}}(\hat{v}_{n,i}^{\ell,\mathbf{c}}) \end{array}\right\} \quad i = 1, \ldots, P_\ell. \tag{10}$$

The simulations are pairwise coupled through sharing the same driving noise W. Then the sample covariance matrices and Kalman gains for the respective ensembles are computed by

$$\begin{aligned} C_{n+1}^{\ell,\mathbf{f}} &= \overline{\mathrm{Cov}}[v_{n+1}^{\ell,\mathbf{f}}], & K_{n+1}^{\ell,\mathbf{f}} &= C_{n+1}^{\ell,\mathbf{f}} H^{\mathbf{T}} (HC_{n+1}^{\ell,\mathbf{f}} H^{\mathbf{T}} + \Gamma)^{-1}, \\ C_{n+1}^{\ell,\mathbf{c}_1} &= \overline{\mathrm{Cov}}[v_{n+1}^{\ell,\mathbf{c}_1}], & K_{n+1}^{\ell,\mathbf{c}_1} &= C_{n+1}^{\ell,\mathbf{c}_1} H^{\mathbf{T}} (HC_{n+1}^{\ell,\mathbf{c}_1} H^{\mathbf{T}} + \Gamma)^{-1}, \\ C_{n+1}^{\ell,\mathbf{c}_2} &= \overline{\mathrm{Cov}}[v_{n+1}^{\ell,\mathbf{c}_2}], & K_{n+1}^{\ell,\mathbf{c}_2} &= C_{n+1}^{\ell,\mathbf{c}_2} H^{\mathbf{T}} (HC_{n+1}^{\ell,\mathbf{c}_2} H^{\mathbf{T}} + \Gamma)^{-1}, \end{aligned}$$

where the sample covariances are defined in a similar fashion as in for EnKF:

$$\overline{\mathrm{Cov}}[v_{n+1}^{\ell,\mathbf{f}}] := \sum_{i=1}^{P_\ell} \frac{v_{n+1,i}^{\ell,\mathbf{f}} \left(v_{n+1,i}^{\ell,\mathbf{f}}\right)^{\mathbf{T}}}{P_\ell} - \sum_{i=1}^{P_\ell} \frac{v_{n+1,i}^{\ell,\mathbf{f}}}{P_\ell} \left(\sum_{i=1}^{P_\ell} \frac{v_{n+1,i}^{\ell,\mathbf{f}}}{P_\ell}\right)^{\mathbf{T}},$$

$$\overline{\mathrm{Cov}}[v_{n+1}^{\ell,\mathbf{c_j}}] := \sum_{i=1}^{P_{\ell-1}} \frac{v_{n+1,i}^{\ell,\mathbf{c_j}} \left(v_{n+1,i}^{\ell,\mathbf{c_j}}\right)^{\mathbf{T}}}{P_{\ell-1}} - \sum_{i=1}^{P_{\ell-1}} \frac{v_{n+1,i}^{\ell,\mathbf{c_j}}}{P_{\ell-1}} \left(\sum_{i=1}^{P_{\ell-1}} \frac{v_{n+1,i}^{\ell,\mathbf{c_j}}}{P_{\ell-1}}\right)^{\mathbf{T}}, \quad j = 1,2. \tag{11}$$

And the new observation y_{n+1} is assimilated into the filter in the update step:

$$\begin{aligned} \left.\begin{array}{l} \tilde{y}_{n+1,i}^{\ell} = y_{n+1} + \eta_{n+1,i}^{\ell} \\ \hat{v}_{n+1,i}^{\ell,\mathbf{f}} = (I - K_{n+1}^{\ell,\mathbf{f}} H) v_{n+1,i}^{\ell,\mathbf{f}} + K_{n+1}^{\ell,\mathbf{f}} \tilde{y}_{n+1,i}^{\ell} \end{array}\right\} \quad & i = 1, ..., P_\ell, \\ \left.\begin{array}{l} \hat{v}_{n+1,i}^{\ell,\mathbf{c}_1} = (I - K_{n+1}^{\ell,\mathbf{c}_1} H) v_{n+1,i}^{\ell,\mathbf{c}_1} + K_{n+1}^{\ell,\mathbf{c}_1} \tilde{y}_{n+1,i}^{\ell} \\ \hat{v}_{n+1,i}^{\ell,\mathbf{c}_2} = (I - K_{n+1}^{\ell,\mathbf{c}_2} H) v_{n+1,i}^{\ell,\mathbf{c}_2} + K_{n+1}^{\ell,\mathbf{c}_2} \tilde{y}_{n+1,i+P_{\ell-1}}^{\ell} \end{array}\right\} \quad & i = 1, ..., P_{\ell-1}, \end{aligned} \tag{12}$$

where $\{\eta_{n,i}^{\ell}\}_{n,\ell,i}$ are iid $N(0,\Gamma)$-distributed random variables. We note that the pairwise coupling of fine- and coarse-level particles is obtained through them sharing the same initial condition, driving noise W in the dynamics (10), and perturbed observations. And to further elaborate on the relation between pairwise-coupled

particles and pairwise-coupled EnKF estimators, the numerator in the (ℓ, m)-th summand/sample of (9) takes the following form when represented in terms of its particles:

$$\mu_n^{\ell,\mathbf{f},m}[\varphi] - \mu_n^{\ell,\mathbf{c},m}[\varphi] = \begin{cases} P_0^{-1} \sum_{i=1}^{P_0} \varphi(\hat{v}_{n,i}^{0,\mathbf{f},m}) & \text{if} \quad \ell = 0 \\ P_\ell^{-1} \sum_{i=1}^{P_\ell} \varphi(\hat{v}_{n,i}^{\ell,\mathbf{f},m}) - \varphi(\hat{v}_{n,i}^{\ell,\mathbf{c},m}) & \text{if} \quad \ell \geq 1. \end{cases}$$

The motivation for pairwise-coupled particles is primarily to reduce the variance of the above sample/summand numerators $\varphi(\hat{v}_n^{\ell,\mathbf{f},m}) - \varphi(\hat{v}_n^{\ell,\mathbf{c},m})$, and ultimately to improve the efficiency of your MLEnKF estimator. Provided the variance of these terms are substantially reduced, the number of samples $\{M_\ell\}$ needed to achieve a certain accuracy with your MLEnKF estimator may be lowered substantially, and this will improve the efficiency of your MLEnKF estimator. See Figure 2 for an illustration of one prediction-update iteration of the full MLEnKF estimator.
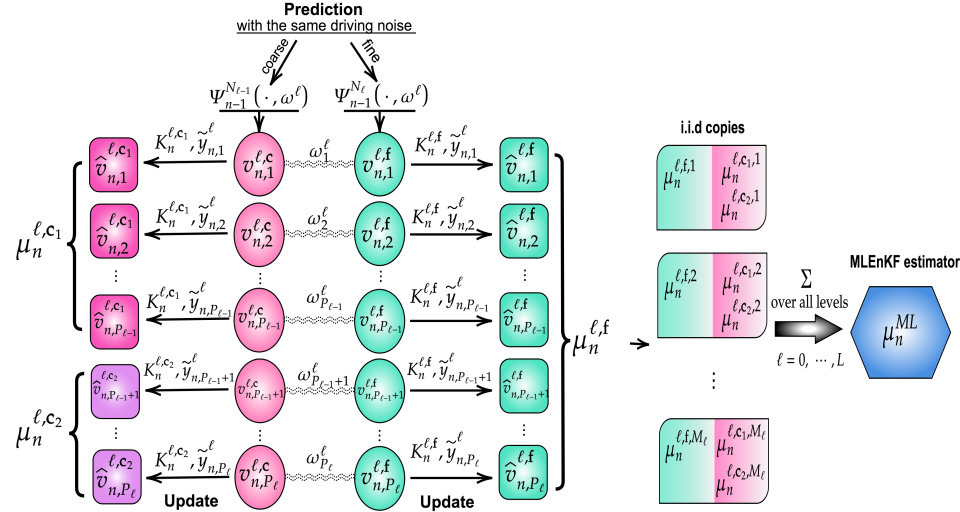

FIGURE 2. One prediction-update iteration of the MLEnKF estimator described in Section 2.4.1. Green and pink ovals represent fine- and coarse-level prediction-state particles, respectively, sharing the same initial condition and driving noise ω^ℓ and the respective squares represent fine- and coarse-level updated-state particles sharing the perturbed observartions. The MLEnKF estimator is obtained by iid copies of pairwise-coupled samples, cf (9).

Remark 2. For the herein-introduced MLEnKF method all the EnKF estimators $\{\mu^{\ell,\cdot,m}[\varphi]\}_{\ell,m}$ are independent for different $(\ell, m)-$samples. This in particular implies that also all Kalman gains and particles from different (ℓ, m)-samples are independent. For the original MLEnKF method [22], on the other hand, all the particles in the full estimator are updated using the one shared Kalman gain. Therefore, the particles in the original MLEnKF method are all correlated, while all particles from different (ℓ, m)-samples in the herein-introduced estimator are independent. The more extensive independence properties of the new MLEnKF estimator simplifies the proofs of asymptotic convergence results and in some settings also improves theoretical complexity bounds. See Corollary 2 and Remark 6 for a comparison of the theoretical complexity of the two methods.

2.5. **Main results.** We now present the main convergence and complexity results for EnKF and MLEnKF. The proofs for these results are provided in Appendix B. We will need the following assumptions:

Assumption 2. *For any exponentially-growing sequence $\{N_\ell\}$ of natural numbers, there exists a constant $c_\Psi > 0$ and $\beta > 0$ such that for all $\ell \in \mathbb{N}_0 \cup \{\infty\}$, $n \in \mathbb{N}_0$, and $p \geq 2$,*

(i) *for all $|\kappa|_1 \leq 1$,*
$$\left\| \partial^\kappa \Psi_n^{N_\ell}(u) \right\|_p \leq c_\Psi (1 + \|u\|_p) \quad \forall u \in L_n^p(\Omega, \mathbb{R}^d),$$

(ii) *for all $|\kappa|_1 = 2$,*
$$\left\| \partial^\kappa \Psi_n^{N_\ell}(u) \right\|_p \leq c_\Psi (1 + \|u\|_{2p}) \quad \forall u \in L_n^{2p}(\Omega, \mathbb{R}^d),$$

(iii) *for all $|\kappa|_1 \leq 1$,*
$$\left\| \partial^\kappa \Psi_n^{N_{\ell+1}}(u) - \partial^\kappa \Psi_n^{N_\ell}(u) \right\|_p \leq c_\Psi (1 + \|u\|_p) N_\ell^{-\beta/2} \quad \forall u \in L_n^p(\Omega, \mathbb{R}^d).$$

Remark 3. If the dynamics (1) is given by the SDE in Remark 1 and Ψ_n^N denotes the corresponding Euler-Maruyama numerical solution, then Assumption 2 holds with $\beta = 1$, cf. [45, 23, 24, 21].

Theorem 1 (Convergence of EnKF). *If Assumption 1 holds, then for any $\varphi \in \mathbb{F}$, $p \geq 2$ and $n \in \mathbb{N}_0$,*
$$\|\mu_n^{N,P}[\varphi] - \bar{\mu}_n[\varphi]\|_p \lesssim P^{-1/2} + N^{-\alpha}, \tag{13}$$
where the hidden constant in $\lesssim$ depends on p, n and d (but not on N and P).

Corollary 1. *If Assumption 1 holds, then $N \eqsim \epsilon^{-1/\alpha}$ and $P \eqsim \epsilon^{-2}$ ensure that*
$$\|\mu_n^{N,P}[\varphi] - \bar{\mu}_n[\varphi]\|_p \lesssim \epsilon$$
for any $\varphi \in \mathbb{F}$, $p \geq 2$ and $n \in \mathbb{N}_0$. The resulting computational cost is
$$\mathrm{Cost(EnKF)} \eqsim \epsilon^{-(2+\frac{1}{\alpha})}.$$

Remark 4. The proof of Theorem 1 shows that Theorem 1 and Corollary 1 apply even if the constraint $\mathbb{F} \subset C_P^2(\mathbb{R}^d, \mathbb{R})$ in Assumption 1(ii) is weakened to $\mathbb{F} \subset C_P^1(\mathbb{R}^d, \mathbb{R})$.

Theorem 2 (Convergence of MLEnKF). *If Assumptions 1 and 2 hold, then for any $L \geq 1$ and triplet of sequences $\{M_\ell\}, \{N_\ell\}, \{P_\ell\} \subset \mathbb{N}$ with $P_\ell \eqsim 2^\ell$ and exponentially-growing N_ℓ, it holds for any $\varphi \in \mathbb{F}$, $n \geq 0$, and $p \geq 2$ that*
$$\|\mu_n^{\mathrm{ML}}[\varphi] - \overline{\mu}_n[\varphi]\|_p \lesssim B_L + \sum_{\ell=0}^L \sqrt{V_\ell/M_\ell},$$
where $B_L := N_L^{-\beta/2} P_L^{-1/2} + P_L^{-1} + N_L^{-\alpha}$ and $V_\ell = (N_\ell^{-\beta/2} P_\ell^{-1/2} + P_\ell^{-1})^2$ may be loosely associated with the bias and statistical error, respectively, and the hidden constant in $\lesssim$ depends on p, n and d (but not on $\{M_\ell\}, \{N_\ell\}, \{P_\ell\}$ and L).

Remark 5. In Appendix B, the Theorems 1 and 2 are for simplicity proved for EnKF and MLEnKF methods using biased sample covariances, precisely as introduced in (7) and (11), respectively. However, the proofs can easily be extended, without any changes in convergence rates, to methods that instead use unbiased sample covariances, cf. Remark 7.

Corollary 2. *If Assumptions 1 and 2 hold, then for any $\epsilon, s > 0$, the configuration $L = \left\lceil \frac{\log_2(\epsilon^{-1})}{\min(1,(1+\beta s)/2,\alpha s)} \right\rceil$, $P_\ell \eqsim 2^\ell$, $N_\ell \eqsim 2^{s\ell}$ and*

$$
M_\ell \eqsim \begin{cases} \epsilon^{-2} 2^{-\frac{3+2s+\min(\beta s,1)}{3}\ell} + 1 & \text{if } \min(\beta s, 1) > s, \\ \epsilon^{-2} L^2 2^{-(1+s)\ell} + 1, & \text{if } \min(\beta s, 1) = s, \quad (14) \\ \epsilon^{-2-\frac{2(s-\min(\beta s,1))}{3\min(1,(1+\beta s)/2,\alpha s)}} 2^{-\frac{3+2s+\min(\beta s,1)}{3}\ell} + 1, & \text{if } \min(\beta s, 1) < s, \end{cases}
$$

ensures that for any $\varphi \in \mathbb{F}$, $n \geq 0$ and $p \geq 2$,

$$
\|\mu_n^{\mathrm{ML}}[\varphi] - \bar{\mu}_n[\varphi]\|_p \lesssim \epsilon. \tag{15}
$$

Moreover, if one sets the parameter s such that

$$
s \in \begin{cases} [\alpha^{-1}, 1) & \text{if } \beta > 1 \quad \& \quad \alpha > 1, \\ [\alpha^{-1}, 1] & \text{if } (\beta > 1 \quad \& \quad \alpha = 1) \text{ or } (\beta = 1 \text{ and } \alpha \geq 1), \\ \{\alpha^{-1}\} & \text{if } (\beta \geq 1 \quad \& \quad \alpha < 1) \text{ or } (\beta < 1 \text{ and } \alpha \leq \beta), \\ \{(2\alpha - \beta)^{-1}\} & \text{if } \beta < 1 \quad \& \quad \alpha > \beta, \end{cases} \tag{16}
$$

then the cost of the MLEnKF estimator is bounded by

$$
\mathrm{Cost(MLEnKF)} \eqsim \begin{cases} \epsilon^{-2} & \text{if } \beta > 1 \quad \& \quad \alpha > 1, \\ \epsilon^{-2} |\log(\epsilon)|^3 & \text{if } (\beta > 1 \quad \& \quad \alpha = 1) \text{ or } (\beta = 1 \text{ and } \alpha \geq 1), \\ \epsilon^{-(1+1/\alpha)} & \text{if } (\beta \geq 1 \quad \& \quad \alpha < 1) \text{ or } (\beta < 1 \text{ and } \alpha \leq \beta), \\ \epsilon^{-(2+(1-\beta)/\alpha)} & \text{if } \beta < 1 \quad \& \quad \alpha > \beta. \end{cases}
$$

$$(17)$$

The optimization of the parameter s is stated in terms of inclusion sets in (16) as that may be useful for implementations. For further details on this optimization and the computational cost of the MLEnKF estimator, see Appendix B.3. The a priori cost-versus-accuracy results of Corollaries 1 and 2 show that in settings where both rates are sharp, MLEnKF will asymptotically have better complexity than EnKF.

Remark 6. For comparison, we recall in compact form the cost-versus-accuracy result for the original MLEnKF [22, Theorem 3.2]: For any $\epsilon > 0$ and sufficiently smooth QoI φ one may choose the estimator's degrees of freedom so that

$$
\|\mu_n^{\mathrm{ML}^{\mathrm{ORG}}}[\varphi] - \bar{\mu}_n[\varphi]\|_p \lesssim \epsilon, \tag{18}
$$

where $\mu_n^{\mathrm{ML}^{\mathrm{ORG}}}[\varphi]$ denotes the original MLEnKF estimator. The computational cost of achieving this accuracy is bounded by

$$
\mathrm{Cost(MLEnKF^{ORG})} \eqsim \begin{cases} |\log(\epsilon)|^{2n} \epsilon^{-2}, & \text{if } \beta > 1, \\ |\log(\epsilon)|^{2n+3} \epsilon^{-2}, & \text{if } \beta = 1, \\ \left(\epsilon / |\log(\epsilon)|^{2n}\right)^{-\left(2+\frac{1-\beta}{\alpha}\right)}, & \text{if } \beta < 1. \end{cases} \tag{19}
$$

To achieve the same $\mathcal{O}(\epsilon)$-accuracy for the two MLEnKF methods, the theoretical bounds imply that for any $n \geq 2$ and any admissible (α, β)-values with $\alpha \geq 1$,

$$
\lim_{\epsilon \to 0} \frac{\mathrm{Cost(MLEnKF^{NEW})}}{\mathrm{Cost(MLEnKF^{ORG})}} = 0.
$$

In other words, if our theoretical bounds are sharp for both methods, then the new MLEnKF method asymptotically outperforms the original one in terms of computational cost versus accuracy. We note however that the regularity assumptions imposed for the new method's hierarchy of numerical solvers, cf. Assumption 2, are more restrictive than those needed for the original method, and that we have previously made the conjecture that the error bounds for the original method are not sharp, cf. [22, Remark 3].

3. **Numerical examples.** In this section, we numerically compare the performance of MLEnKF and EnKF and (numerically) verify the results predicted by our theory. We will consider a couple of test problems with dynamics of the form

$$u_{n+1} = \Psi(u_n) := u_n + \int_n^{n+1} -V'(u_t)dt + \int_n^{n+1} \sigma dW_t, \tag{20}$$

where the potential $V \in C_P^\infty(\mathbb{R})$ satisfies $V'' \in C_B^\infty(\mathbb{R})$, the diffusion coefficient is constant; $\sigma = 0.5$, and linear observations (3) with $H = 1$ and $\Gamma = 0.1$. For any $N \geq 1$, Ψ^N will in this section denote the Milstein numerical scheme with uniform timestep $\Delta t = 1/N$. This yields the convergence rates $\alpha = 1$, $\beta = 2$.

For any $x \in \mathbb{R}$, let Round(x) denote the nearest integer to x. Given an input RMSE constraint $\epsilon > 0$, we seek to equilibrate the magnitude of all error terms (cf. (13) for EnKF and Corollary 2 for MLEnKF) by setting the degrees of freedom to

$$P = \text{Round}(8\epsilon^{-2}) \quad \text{and} \quad N = \text{Round}(\epsilon^{-1}) \tag{21}$$

for EnKF, and for MLEnKF we set $L = \text{Round}(\log_2(\epsilon^{-1})) - 1$, $N_\ell = 2^{\ell+1}$, $P_\ell = 10 \times 2^\ell$, and

$$M_\ell = \begin{cases} 2\text{Round}(\epsilon^{-2}L^2 2^{-3}) & \text{if} \quad \ell = 0 \\ \text{Round}(\epsilon^{-2}L^2 2^{-2\ell-3}) & \text{if} \quad 1 \leq \ell \leq L. \end{cases}$$

Over a sequence of iterations $n = 1, \ldots, \mathcal{N}$, we compare the performance of the methods in terms of error versus computer runtime (wall-clock time), where the error is measured in time-averaged-root-mean-squared error (RMSE):

$$\text{RMSE(EnKF)} := \sqrt{\frac{1}{100(\mathcal{N}+1)} \sum_{i=1}^{100} \sum_{n=0}^{\mathcal{N}} \left| \mu_{n,i}^{N,P}[\varphi] - \bar{\mu}_n[\varphi] \right|^2} \lesssim \text{Runtime}^{-1/3},$$

where for a given $\epsilon > 0$ with $\alpha = 1$ and $\beta = 2$, the last approximate asymptotic inequality follows from Corollary 1

$$\text{RMSE(EnKF)} \approx \left(\sum_{n=0}^{\mathcal{N}} \frac{\left\| \mu_n^{N,P}[\varphi] - \bar{\mu}_n[\varphi] \right\|_2^2}{\mathcal{N}+1} \right)^{1/2} \lesssim \epsilon \approx \text{Cost(EnKF)}^{-1/3}.$$

Similarly,

$$\text{RMSE(MLEnKF)} := \sqrt{\frac{1}{100(\mathcal{N}+1)} \sum_{i=1}^{100} \sum_{n=0}^{\mathcal{N}} \left| \mu_{n,i}^{\text{ML}}[\varphi] - \bar{\mu}_n[\varphi] \right|^2}$$

$$\approx \left(\sum_{n=0}^{\mathcal{N}} \frac{\left\| \mu_n^{\text{ML}}[\varphi] - \bar{\mu}_n[\varphi] \right\|_2^2}{\mathcal{N}+1} \right)^{1/2}$$

$$\lesssim \log(10 + \text{Runtime})^{3/2} \text{Runtime}^{-1/2},$$

where the last inequality follows from Corollary 2, since it implies that

$$\text{RMSE(MLEnKF)} \lesssim \epsilon \approx \log(\text{Cost(MLEnKF)})^{3/2}\text{Cost(MLEnKF)}^{-1/2}.$$

Note here that $\{\mu_{.,i}^{N,P}[\varphi]\}_{i=1}^{100}$ and $\{\mu_{.,i}^{\text{ML}}[\varphi]\}_{i=1}^{100}$ are computed using iid sequences of EnKF and MLEnKF empirical measures, respectively.

Reference solutions and computer architecture. For the first test problem we will consider, the dynamics Ψ is linear. Therefore, we compute the exact reference solution $\bar{\mu}_n[\varphi]$ straightforwardly using the Kalman filter. The second test problem we will consider has nonlinear dynamics, and we need to approximate the reference solution. A pseudo-reference solution is then obtained by the deterministic mean-field EnKF (DMFEnKF) algorithm described in [35], cf. Appendix C.

The numerical simulations were computed in parallel on five cores on a Mac Pro with Quad-Core Intel Xeon X-5550 8-core processor with 16 GB of RAM. The computer code is written in the Julia programming language [4], and it can be downloaded from https://github.com/GaukharSH/mlenkf.

3.1. **Ornstein-Uhlenbeck process.** We consider the SDE (20) with $V(u) = u^2/2$, and the initial condition is $u_0 \sim N(0,\Gamma)$. This is an Ornstein-Uhlenbeck (OU) process with the exact solution

$$\Psi(u_n) = u_n e^{-t} + \int_0^t \sigma e^{s-t} dW_s.$$

For a sequence of inputs, $\epsilon = [2^{-4}, 2^{-5}, 2^{-6}, 2^{-7}, 2^{-8}]$ for EnKF and MLEnKF, $\epsilon = [2^{-4}, 2^{-4.5}, 2^{-5}, \ldots, 2^{-9}]$ for the original MLEnKF, we study the performance of the three methods in terms of runtime versus RMSE over timeframes $\mathcal{N} = 10$ and $\mathcal{N} = 20$ observation times in Figure 3. For both QoI considered (the mean and the variance), the observations are in agreement with Corollaries 1 and 2, and Remark 6. The new MLEnKF method performs similarly as the original MLEnKF method and that both methods asymptotically outperform EnKF.

3.2. **Double-well SDE.** We consider the SDE (20) with the double-well potential

$$V(u) = \frac{1}{2 + 4u^2} + \frac{u^2}{4}$$

and $u_0 \sim N(0,\Gamma)$. This SDE is the metastable motion of particles between the two wells, cf. [8, 44]. Figure 4 shows the dynamics of a particle over $\mathcal{N} = 20$ observation times. We observe that the particle remains in either one of the wells for a relatively long time; in other words, transitions between wells happen quite rarely. Figure 5 illustrates the well transition of an EnKF ensemble over a few predict-update iterations when the observations and the majority of the ensemble's particles are located in opposite wells for the first few iterations. Using the same ϵ-input sequences as in the preceding example, we study the performance of the three methods in terms of runtime versus RMSE over timeframes of $\mathcal{N} = 10$ and $\mathcal{N} = 20$ observation times in Figure 6. Once again, the work rates are in agreement with theory (Corollaries 1, 2 and Remark 6), that the new MLEnKF method performs similarly as the original MLEnKF method, and that both methods asymptotically outperform EnKF.
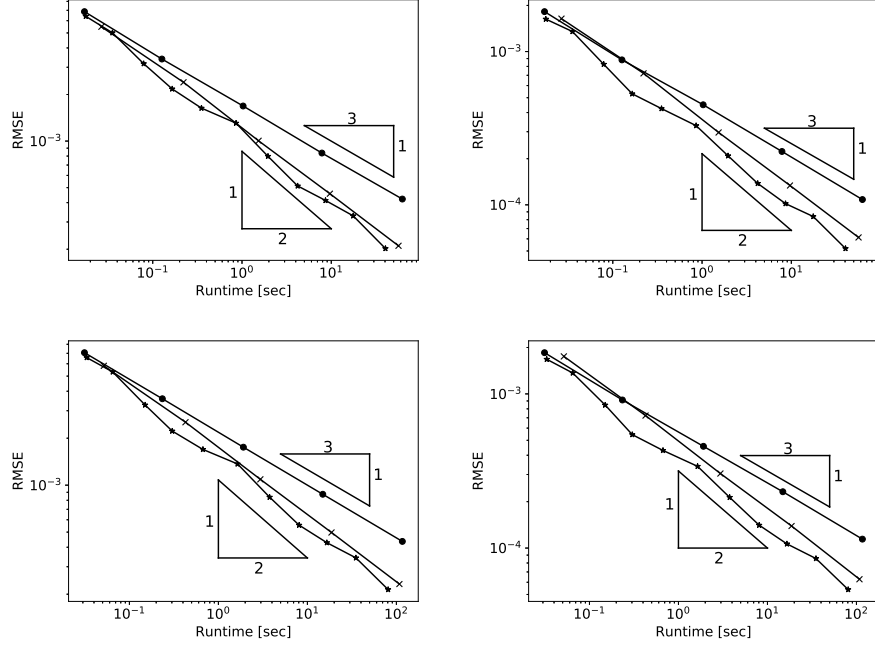
FIGURE 3. Top row: comparison of the runtime versus RMSE for the QoIs mean (left) and variance (right) over $\mathcal{N} = 10$ observation times for the problem in Section 3.1. The solid-crossed line represents MLEnKF, the solid-asterisk line represents the original MLEnKF and the bottom reference triangle with the slope $\frac{1}{2}$, the solid-bulleted line represents EnKF and the upper reference triangle with the slope $\frac{1}{3}$. Bottom row: similar plots over $\mathcal{N} = 20$ observation times.
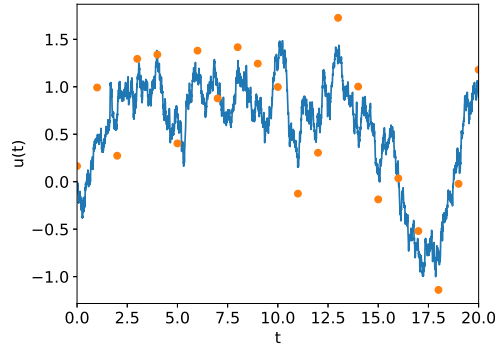
FIGURE 4. Realization of the double-well SDE from Section 3.2 over time $\mathcal{N} = 20$ observation times (solid line) and observations (dots).

Appendix A. Mean-field EnKF.

The large-ensemble-limit of EnKF is referred to as the mean-field EnKF (MFEnKF). In the mean-field limit, the EnKF Kalman gain becomes a deterministic matrix. Consequently, there is no mixing between ensemble particles and the particles become iid. To describe the dynamics of
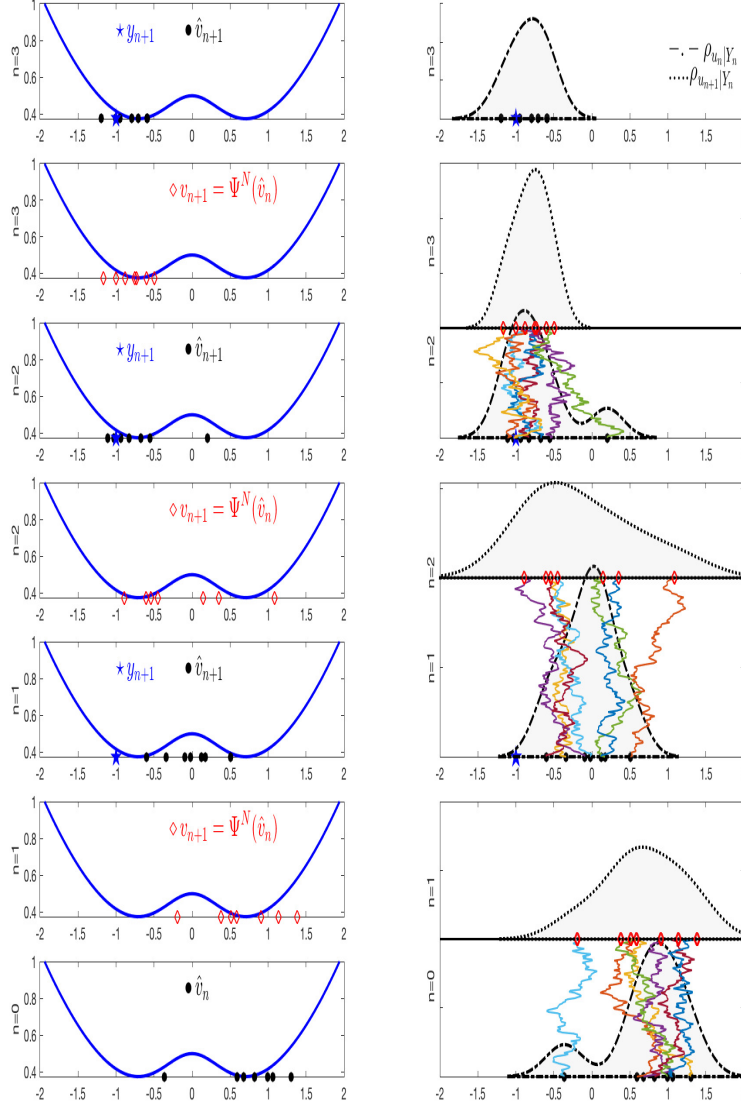
FIGURE 5. Left column: Well transition of the EnKF ensemble when the measurements are located in the opposite well. Right column: Animation of the particle paths of the corresponding EnKF ensemble during the well-transition, and the resulting kernel density estimations of the EnKF prediction and update densities (Section 3.2). For practical purposes, EnKF with only 7 particles is not very robust. We use so few particles in this computation for the sole purpose of obtaining a visually clear illustration of the ensemble transition.

MFEnKF, it thus suffices to consider a single particle. For fixed dynamics resolution Ψ^N, let $\hat{\bar{v}}_n^N$ denote the updated state of a mean-field particle at time n. The MFEnKF prediction and update dynamics is given by

$$\bar{v}_{n+1}^N = \Psi_n^N(\hat{\bar{v}}_n^N), \tag{22}$$

FIGURE 6. Top row: comparison of the runtime versus RMSE for the QoIs mean (left) and variance (right) over $\mathcal{N} = 10$ observation times for the problem in Section 3.2. The solid-crossed line represents MLEnKF, the solid-asterisk line represents the original MLEnKF and the bottom reference triangle with the slope $\frac{1}{2}$, the solid-bulleted line represents EnKF and the upper reference triangle with the slope $\frac{1}{3}$. Bottom row: similar plots over $\mathcal{N} = 20$ observation times
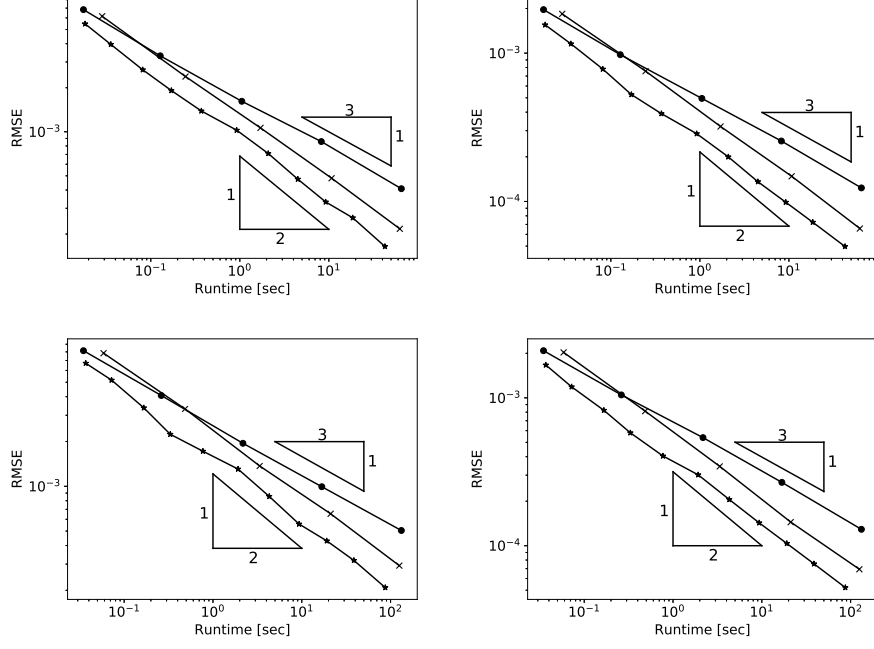
and
$$\hat{\bar{v}}_{n+1}^N = (I - \bar{K}_{n+1}^N H)\bar{v}_{n+1}^N + \bar{K}_{n+1}^N \tilde{y}_{n+1},$$
where
$$\bar{K}_{n+1}^N = \bar{C}_{n+1}^N H^{\mathbf{T}} \left(H\bar{C}_{n+1}^N H^{\mathbf{T}} + \Gamma\right)^{-1}$$
denotes the *deterministic* mean-field Kalman gain,
$$\bar{C}_{n+1}^N = \mathbb{E}\left[\left(\bar{v}_{n+1}^N - \mathbb{E}\left[\bar{v}_{n+1}^N\right]\right)\left(\bar{v}_{n+1}^N - \mathbb{E}\left[\bar{v}_{n+1}^N\right]\right)^{\mathbf{T}}\right]$$
is the mean-field prediction covariance, and
$$\tilde{y}_{n+1} = y_{n+1} + \tilde{\eta}_{n+1}$$
denotes the perturbed observation with $\{\tilde{\eta}_n\}_{n=1}^N$ being independent and $N(0, \Gamma)$-distributed random variables satisfying $\tilde{\eta}_j \perp u_k$ for all $j, k \geq 0$.

By Assumption 1 (i) with $|\kappa|_1 = 0$, and using that $L_n^p(\Omega, \mathbb{R}^d) \subset L^p(\Omega, \mathbb{R}^d)$, it follows straightforwardly by induction that $\bar{v}_n^N, \hat{\bar{v}}_n^N \in \cap_{p \geq 2} L^p(\Omega, \mathbb{R}^d)$ uniformly in $N \in \mathbb{N} \cup \{\infty\}$:

$$
\begin{aligned}
\hat{\bar{v}}_0^N \sim \mathbb{P}_{u_0|Y_0} &\implies \bar{v}_1^N = \Psi_0^N(\hat{\bar{v}}_0^N) \in \cap_{p \geq 2} L^p(\Omega, \mathbb{R}^d) \\
&\implies |\bar{C}_1^N|_2 < \infty \implies |\bar{K}_1|_2 < \infty \\
&\implies \|\hat{\bar{v}}_1^N\|_p \leq c_p(1 + |H|_2)\left(\|\bar{v}_1^N\|_p + \|\tilde{y}_1\|_p\right) < \infty \quad \forall p \geq 2 \\
&\implies \dots \implies \hat{\bar{v}}_n^N \in \cap_{p \geq 2} L^p(\Omega, \mathbb{R}^d).
\end{aligned}
\tag{23}
$$

This ensures the existence of the mean-field measure $\bar{\mu}_n^N$, i.e., the distribution $\hat{\bar{v}}_n^N \sim \bar{\mu}_n^N$ satisfies $\hat{\bar{v}}_n^N \in \cap_{p \geq 2} L^p(\Omega, \mathbb{R}^d)$ for all $N \in \mathbb{N} \cup \{\infty\}$ and $n \in \mathbb{N}_0$. For evaluating the expectation of a QoI $\varphi : \mathbb{R}^d \to \mathbb{R}$ wrt to the mean-field measure $\bar{\mu}_n^N$, we recall from Notation 2 that

$$\bar{\mu}_n^N[\varphi] = \int_{\mathbb{R}^d} \varphi(v) \bar{\mu}_n^N(dv).$$

For a subset of nonlinear dynamics Ψ with additive Gaussian noise, it has been shown that in the large-ensemble limit, EnKF with perturbed observations converges to the Kalman filtering distribution with the standard rate $\mathcal{O}(P^{-1/2})$, cf. [36]. That is, in L^p for all QoI $\varphi : \mathbb{R}^d \to \mathbb{R}$ with sufficient regularity $n \in \mathbb{N}_0$, and $p \geq 2$

$$\left\| \mu_n^{\infty,P}[\varphi] - \bar{\mu}_n[\varphi] \right\|_p \leq c_{\varphi,p} P^{-1/2} \quad \text{for some} \quad C_{\varphi,\Psi^\infty} > 0,$$

where $\bar{\mu}_n := \bar{\mu}_n^\infty$. The result was extended to a subset of fully non-Gaussian models [35] with numerical approximations of the dynamics in [22], i.e., $\mu_n^{N,P}[\varphi] \to \bar{\mu}_n[\varphi]$ as $P, N \to \infty$ (see also Theorem 1).

Appendix B. Theoretical proofs.

B.1. EnKF.
In this section, we present a collection of theoretical results for EnKF that culminates with proof of Theorem 1.

Notation. The sample average of a QoI $\varphi \in \mathbb{F}$ applied to an ensemble of P identically distributed particles $\{\hat{v}_{n,i}^{N,P}\}_{i=1}^P$ is defined as

$$E_P[\varphi(\hat{v}_n^{N,P})] := \frac{1}{P} \sum_{i=1}^P \varphi(\hat{v}_{n,i}^{N,P}) \quad (= \mu_n^{N,P}[\varphi]),$$

and similarly

$$E_P[\hat{v}_n^{N,P}] := \frac{1}{P} \sum_{i=1}^P \hat{v}_{n,i}^{N,P} \quad \text{and} \quad E_P[(\hat{v}_n^{N,P})(\hat{v}_n^{N,P})^{\mathbf{T}}] := \frac{1}{P} \sum_{i=1}^P \left(\hat{v}_{n,i}^{N,P}\right)\left(\hat{v}_{n,i}^{N,P}\right)^{\mathbf{T}}.$$

$$(24)$$

Lemma 1 (Difference between EnKF and MFEnKF Kalman gains). *If Assumption 1 holds, then*

$$\left\| K_n^{N,P} - \bar{K}_n^N \right\|_p \leq |\Gamma^{-1}|_2 |H|_2 \left(1 + 2|\bar{K}_n^N H|_2\right) \left\| C_n^{N,P} - \bar{C}_n^N \right\|_p \lesssim \left\| C_n^{N,P} - \bar{C}_n^N \right\|_p$$

for any $n \in \mathbb{N}_0$ and $p \geq 2$.

Proof. The proof of the first inequality is analogous to [22, Lemma 3.4], and the latter inequality follows from $\bar{v}_n^N \in \cap_{p \geq 2} L^p(\Omega, \mathbb{R}^d)$, uniformly in $N \geq 1$, cf. (23). $\square$

We next bound the difference between $C_n^{N,P} = \overline{\mathrm{Cov}}[v_n^{N,P}]$ and $\bar{C}_n^N = \mathrm{Cov}[\bar{v}_n^N]$.

Lemma 2 (Difference between EnKF and MFEnKF covariance matrices). *If Assumption 1 holds, then*

$$\left\| C_n^{N,P} - \bar{C}_n^N \right\|_p \lesssim \left\| v_n^{N,P} - \bar{v}_n^N \right\|_{2p} + P^{-\frac{1}{2}}$$

for any $n \in \mathbb{N}_0$ and $p \geq 2$.

Proof. In order to bound $\left\|C_n^{N,P} - \bar{C}_n^N\right\|_p$, we introduce the auxiliary mean-field ensemble $\hat{\bar{v}}_{n,1:P}^{N,P} = \{\hat{\bar{v}}_{n,i}^{N,P}\}_{i=1}^P$ consisting of P identically distributed particles whose prediction/update dynamics is given by the mean-field Kalman gain, namely

$$
\begin{aligned}
\bar{v}_{n+1,i}^{N,P} &= \Psi_n^N(\hat{\bar{v}}_{n,i}^{N,P}), \\
\hat{\bar{v}}_{n+1,i}^{N,P} &= (I - \bar{K}_{n+1}^N H)\bar{v}_{n+1,i}^{N,P} + \bar{K}_{n+1}^N \tilde{y}_{n+1,i},
\end{aligned}
\tag{25}
$$

with the initial condition $\hat{\bar{v}}_{0,i}^{N,P} = \hat{v}_{0,i}^{N,P}$ for $i = 1, ..., P$. The auxiliary ensemble satisfies the following two crucial properties: $\bar{v}_n^{N,P} \stackrel{D}{=} \bar{v}_n$ and $\bar{v}_{n,i}^{N,P}$ has the same initial data, driving noise W and perturbed observations as the i-th EnKF particle $v_{n,i}^{N,P}$, cf. Section 2.3. In other words, $\bar{v}_{n,i}^{N,P}$ is shadowing $v_{n,i}^{N,P}$.

Introducing the auxiliary ensemble covariance based on the mean-field prediction particles,

$$
\tilde{\bar{C}}_n^{N,P} := \overline{\mathrm{Cov}}[\bar{v}_n^{N,P}],
\tag{26}
$$

the triangle inequality implies that

$$
\left\|C_n^{N,P} - \bar{C}_n^N\right\|_p \leq \left\|C_n^{N,P} - \tilde{\bar{C}}_n^{N,P}\right\|_p + \left\|\tilde{\bar{C}}_n^{N,P} - \bar{C}_n^N\right\|_p.
\tag{27}
$$

For the first term, recalling the notation (24), we have

$$
\begin{aligned}
\left\|C_n^{N,P} - \tilde{\bar{C}}_n^{N,P}\right\|_p \leq\ & \left\|E_P[v_n^{N,P}((v_n^{N,P})^{\mathbf{T}} - (\bar{v}_n^{N,P})^{\mathbf{T}})]\right\|_p \\
& + \left\|E_P[(v_n^{N,P} - \bar{v}_n^{N,P})(\bar{v}_n^{N,P})^{\mathbf{T}}]\right\|_p \\
& + \left\|E_P[\bar{v}_n^{N,P}]E_P[(\bar{v}_n^{N,P})]^{\mathbf{T}} - E_P[\bar{v}_n^{N,P}]E_P[(v_n^{N,P})]^{\mathbf{T}}\right\|_p \\
& + \left\|E_P[\bar{v}_n^{N,P}]E_P[(v_n^{N,P})]^{\mathbf{T}} - E_P[v_n^{N,P}]E_P[(v_n^{N,P})]^{\mathbf{T}}\right\|_p.
\end{aligned}
$$

Jensen's and Hölder's inequalities and $\bar{v}_n^{N,P} \in L^p(\Omega, \mathbb{R}^d)$ yield that

$$
\begin{aligned}
\left\|C_n^{N,P} - \tilde{\bar{C}}_n^{N,P}\right\|_p &\leq 2\left\|v_n^{N,P}\right\|_{2p}\left\|v_n^{N,P} - \bar{v}_n^{N,P}\right\|_{2p} + 2\left\|\bar{v}_n^{N,P}\right\|_{2p}\left\|v_n^{N,P} - \bar{v}_n^{N,P}\right\|_{2p} \\
&\leq 2\left\|v_n^{N,P} - \bar{v}_n^{N,P}\right\|_{2p}^2 + 4\left\|\bar{v}_n^{N,P}\right\|_{2p}\left\|v_n^{N,P} - \bar{v}_n^{N,P}\right\|_{2p} \\
&\lesssim \max\left(\left\|v_n^{N,P} - \bar{v}_n^{N,P}\right\|_{2p}^2, \left\|v_n^{N,P} - \bar{v}_n^{N,P}\right\|_{2p}\right).
\end{aligned}
$$

For the second term in (27), we have

$$
\begin{aligned}
\left\|\tilde{\bar{C}}_n^{N,P} - \bar{C}_n^N\right\|_p \leq\ & \left\|E_P[\bar{v}_n^{N,P}(\bar{v}_n^{N,P})^{\mathbf{T}}] - \mathbb{E}\left[\bar{v}_n^N(\bar{v}_n^N)^{\mathbf{T}}\right]\right\|_p \\
& + \left\|E_P[\bar{v}_n^{N,P}]E_P[(\bar{v}_n^{N,P})^{\mathbf{T}}] - \mathbb{E}\left[\bar{v}_n^N\right]\mathbb{E}\left[(\bar{v}_n^N)^{\mathbf{T}}\right]\right\|_p \\
& \lesssim P^{-\frac{1}{2}}\left\|\bar{v}_n^{N,P}(\bar{v}_n^{N,P})^{\mathbf{T}}\right\|_p \\
& \lesssim P^{-\frac{1}{2}}\left\|\bar{v}_n^{N,P}\right\|_{2p}^2,
\end{aligned}
$$

where the penultimate inequality is obtained by the Marcinkiewicz-Zygmund (M-Z) inequality [32, Theorem 5.2]. The property $\bar{v}_n^{N,P} \in L^p(\Omega, \mathbb{R}^d)$, together with the above inequalities imply that

$$
\left\|C_n^{N,P} - \bar{C}_n^N\right\|_p \lesssim \max(\left\|v_n^{N,P} - \bar{v}_n^{N,P}\right\|_{2p}, \left\|v_n^{N,P} - \bar{v}_n^{N,P}\right\|_{2p}^2) + P^{-\frac{1}{2}}.
\tag{28}
$$

Finally, by the proof of Lemma 3, it follows that $\left\|v_n^{N,P} - \bar{v}_n^{N,P}\right\|_{2p} \lesssim 1$ and thus

$$
\left\|C_n^{N,P} - \bar{C}_n^N\right\|_p \lesssim \left\|v_n^{N,P} - \bar{v}_n^{N,P}\right\|_{2p} + P^{-\frac{1}{2}}.
$$

$\square$

Remark 7. Lemma 2 straightforwardly extends to EnKF methods using an unbiased sample prediction covariance $\mathcal{C}_n^{N,P}$. By the following relationship between biased and unbiased sample covariances

$$\mathcal{C}_n^{N,P} := \frac{P}{P-1} C_n^{N,P},$$

we obtain that

$$\left\| \mathcal{C}_n^{N,P} - \bar{C}_n^N \right\|_p \le \left\| \mathcal{C}_n^{N,P} - C_n^{N,P} \right\|_p + \left\| C_n^{N,P} - \bar{C}_n^N \right\|_p$$

$$= \frac{P}{P-1} \left\| C_n^{N,P} - \bar{C}_n^N \right\|_p + \frac{1}{P-1} \left\| \bar{C}_n^{N,P} \right\|_p.$$

The first term in the last equality is bounded by Lemma 2, and the bound for the second term follows from $\bar{v}_n^{N,P} \in L^p(\Omega, \mathbb{R}^d)$. In the proof of Theorem 1, the error contribution from the sample covariance, be that a biased or an unbiased one, only enters through Lemma 2. This implies that the theorem holds and the convergence rate is not affected when replacing biased with unbiased sample covariances in the EnKF method.

Moreover, in the proof of Theorem 2 for MLEnKF, the error contribution from sample covariances only enter through Corollary 3 and Lemma 8. By a similar argument as above, the rates of the said corollary and lemma are not affected by replacing biased sample covariances by unbiased ones and Theorem 2 will also holds with the same convergence rate.

Lemma 3 (Distance between ensembles)**.** *If Assumption 1 holds, then*

$$\max \left(\left\| \hat{v}_n^{N,P} - \hat{\bar{v}}_n^N \right\|_p, \left\| v_n^{N,P} - \bar{v}_n^N \right\|_p \right) \lesssim P^{-1/2} \quad \text{for any} \quad n \in \mathbb{N}_0 \quad \text{and} \quad p \ge 2.$$

Proof. Since $\hat{v}_0^{N,P} = \hat{\bar{v}}_0^N$, the first half of the statement holds for $n = 0$. Assume that for some $n \ge 1$, $\|\hat{v}_{n-1}^{N,P} - \hat{\bar{v}}_{n-1}^N\|_p \lesssim P^{-1/2}$ holds for all $p \ge 2$. Assumption 1(iii) then implies that

$$\left\| v_n^{N,P} - \bar{v}_n^N \right\|_p \lesssim \left\| \hat{v}_{n-1}^{N,P} - \hat{\bar{v}}_{n-1}^N \right\|_p \lesssim P^{-\frac{1}{2}}. \tag{29}$$

By Hölder's inequality and Lemma 1,

$$\left\| \hat{v}_n^{N,P} - \hat{\bar{v}}_n^N \right\|_p \lesssim |I - \bar{K}_n^N H|_2 \left\| v_n^{N,P} - \bar{v}_n^N \right\|_p$$
$$+ \left\| C_n^{N,P} - \bar{C}_n^N \right\|_{2p} \left(\left\| v_n^{N,P} - \bar{v}_n^N \right\|_{2p} + \left\| \tilde{y}_n - H\bar{v}_n^N \right\|_{2p} \right). \tag{30}$$

Since $\bar{v}_n^N, \tilde{y}_n^\ell \in \cap_{r \ge 2} L^r(\Omega)$, inequalities (29) and (28) imply that

$$\left\| \hat{v}_n^{N,P} - \hat{\bar{v}}_n^N \right\|_p \lesssim (|I - \bar{K}_n^N H|_2 + \left\| \tilde{y}_n - H\bar{v}_n^N \right\|_{2p}) P^{-1/2} + P^{-1} \lesssim P^{-1/2}.$$

The argument holds for any $p \ge 2$, and the proof follows by induction. $\square$

Before proving the main convergence result for EnKF, we introduce the notation $\bar{\mu}_n^{N,P}[\varphi]$ to denote the average of the QoI φ over the empirical measure associated with the auxiliary mean-field ensemble $\{\hat{\bar{v}}_{n,i}^N\}_{i=1}^P$, cf. (25), and we recall that $\bar{\mu}_n^N[\varphi]$ denotes the evaluation of empirical measure on QoI φ associated with the MFEnKF $\hat{\bar{v}}_n^N$, cf. Section A.

Proof of Theorem 1. By the triangle inequality,

$$\left\| \mu_n^{N,P}[\varphi] - \bar{\mu}_n[\varphi] \right\|_p \le \left\| \mu_n^{N,P}[\varphi] - \bar{\mu}_n^{N,P}[\varphi] \right\|_p + \left\| \bar{\mu}_n^{N,P}[\varphi] - \bar{\mu}_n^N[\varphi] \right\|_p$$
$$+ \left\| \bar{\mu}_n^N[\varphi] - \bar{\mu}_n[\varphi] \right\|_p.$$

By Assumption 1(ii), $\varphi \in \mathbb{F} \subset C_P^2(\mathbb{R}^d, \mathbb{R})$, and Lemma 3 implies that

$$\left\| \mu_n^{N,P}[\varphi] - \bar{\mu}_n^{N,P}[\varphi] \right\|_p = \left\| \frac{1}{P} \sum_{i=1}^P \varphi(\hat{v}_{n,i}^{N,P}) - \varphi(\hat{\bar{v}}_{n,i}^N) \right\|_p \lesssim \left\| \hat{v}_n^{N,P} - \hat{\bar{v}}_n^N \right\|_{2p} \lesssim P^{-1/2}.$$

For the second term, using the Marcinkiewicz–Zygmund (M-Z) inequality, that $\varphi(\hat{\bar{v}}_n^N) \in \cap_{p \ge 2} L^p(\Omega, \mathbb{R})$ and that $\hat{\bar{v}}_{n,1}^{N,P}, \ldots, \hat{\bar{v}}_{n,P}^{N,P}$ are iid with $\hat{\bar{v}}_{n,i}^{N,P} \overset{D}{=} \hat{\bar{v}}_n^N$ imply that

$$\left\| \bar{\mu}_n^{N,P}[\varphi] - \bar{\mu}_n^N[\varphi] \right\|_p = \left\| \sum_{i=1}^P \frac{\varphi(\hat{\bar{v}}_{n,i}^N) - \mathbb{E}\left[\varphi(\hat{\bar{v}}_n^N)\right]}{P} \right\|_p \lesssim P^{-1/2}.$$

For the last term, we have that

$$\left\| \bar{\mu}_n^N[\varphi] - \bar{\mu}_n[\varphi] \right\|_p = \left| \mathbb{E}\left[\varphi(\hat{\bar{v}}_n^N) - \varphi(\hat{\bar{v}}_n)\right] \right|,$$

and it remains to prove by induction that the right-hand side is $\mathcal{O}(N^{-\alpha})$.

Since $\hat{\bar{v}}_0^N \overset{D}{=} \hat{\bar{v}}_0$, it holds that $\left| \mathbb{E}\left[\varphi(\hat{\bar{v}}_0^N) - \varphi(\hat{\bar{v}}_0)\right] \right| = 0$. Assume that for some $k \ge 1$, there exists an observable dependent constant $c_\varphi > 0$ such that

$$\left| \mathbb{E}\left[\varphi(\hat{\bar{v}}_{k-1}^N) - \varphi(\hat{\bar{v}}_{k-1})\right] \right| \le c_\varphi N^{-\alpha} \quad \forall \varphi \in \mathbb{F}.$$

Then, Assumption 1(ii) implies there exists a $\tilde{c}_\varphi > 0$ such that

$$\left| \mathbb{E}\left[\varphi(\bar{v}_k^N) - \varphi(\bar{v}_k)\right] \right| \le \tilde{c}_\varphi N^{-\alpha} \quad \forall \varphi \in \mathbb{F}. \tag{31}$$

In order to bound $\left| \mathbb{E}\left[\varphi(\hat{\bar{v}}_k^N) - \varphi(\hat{\bar{v}}_k)\right] \right|$, we first recall that

$$\hat{\bar{v}}_k^N = (I - \bar{K}_k^N H)\bar{v}_k^N + \bar{K}_k^N y_n + \bar{K}_k^N \tilde{\eta}_k$$
$$\hat{\bar{v}}_k = (I - \bar{K}_k H)\bar{v}_k + \bar{K}_k y_n + \bar{K}_k \tilde{\eta}_k,$$

with $\tilde{\eta}_k \sim N(0, \Gamma)$ and introduce the functions $\tilde{\varphi}^N, \tilde{\varphi} \in \mathbb{F}$ defined by

$$\tilde{\varphi}^N(x) = \frac{1}{\sqrt{(2\pi)^p |\det(\Gamma)|}} \int_{\mathbb{R}^p} \varphi\Big((I - \bar{K}_k^N H)x + \bar{K}_k^N y_k + \bar{K}_k^N z\Big) e^{-z^{\mathbf{T}} \Gamma^{-1} z/2} \, dz$$

and

$$\tilde{\varphi}(x) = \frac{1}{\sqrt{(2\pi)^p |\det(\Gamma)|}} \int_{\mathbb{R}^p} \varphi\Big((I - \bar{K}_k H)x + \bar{K}_k y_k + \bar{K}_k z\Big) e^{-z^{\mathbf{T}} \Gamma^{-1} z/2} \, dz.$$

It then follows by the mean-value theorem that

$$\left| \mathbb{E}\left[\varphi(\hat{\bar{v}}_k^N) - \varphi(\hat{\bar{v}}_k)\right] \right| \le \left| \mathbb{E}\left[\tilde{\varphi}^N(\bar{v}_k^N) - \tilde{\varphi}^N(\bar{v}_k)\right] \right| + \left| \mathbb{E}\left[\tilde{\varphi}^N(\bar{v}_k) - \tilde{\varphi}(\bar{v}_k)\right] \right|$$
$$\lesssim N^{-\alpha} + |\bar{K}_k - \bar{K}_k^N|_2.$$

From [22, Lemma 3.4], we have that

$$\bar{K}_k - \bar{K}_k^N = \bar{K}_k H(\bar{C}_k^N - \bar{C}_k) H^{\mathbf{T}} (H \bar{C}_k^N H^{\mathbf{T}} + \Gamma)^{-1} + (\bar{C}_k - \bar{C}_k^N) H^{\mathbf{T}} (H \bar{C}_k^N H^{\mathbf{T}} + \Gamma)^{-1},$$

which, since Γ is positive definite and thus $|(H\bar{C}_k^N H^{\mathbf{T}} + \Gamma)^{-1}|_2 \lesssim |\Gamma^{-1}|_2 < \infty$, implies that

$$
\begin{aligned}
|\bar{K}_k - \bar{K}_k^N|_2 &\lesssim |\bar{C}_k^N - \bar{C}_k|_2 \\
&\leq \left| \mathbb{E}\left[\bar{v}_k^N (\bar{v}_k^N)^{\mathbf{T}} - \bar{v}_k(\bar{v}_k)^{\mathbf{T}} \right] \right|_2 + \left| \mathbb{E}\left[\bar{v}_k^N \right] \mathbb{E}\left[(\bar{v}_k^N)^{\mathbf{T}} \right] - \mathbb{E}\left[\bar{v}_k \right] \mathbb{E}\left[(\bar{v}_k)^{\mathbf{T}} \right] \right|_2 \\
&\lesssim N^{-\alpha}.
\end{aligned}
$$

Since all monomials of degree 1 and 2 are contained in $\mathbb{F}$, the last inequality follows from (31) and the equivalence of the Euclidean and Frobenius norms in any dimension $d < \infty$. It holds by induction that for any $n \geq 0$,

$$
|\mathbb{E}[\varphi(\hat{v}_n^N) - \varphi(\hat{\bar{v}}_n)]| \lesssim N^{-\alpha}.
$$

$\square$

B.2. MLEnKF. In this section, we present a collection of theoretical results for MLEnKF, including the proof of Theorem 2.

In order to obtain a connection between $\{\hat{v}_{n,i}^{\ell,\mathbf{f}}\}_{i=1}^{P_\ell}$ and $\{\hat{v}_{n,i}^{\ell,\mathbf{c}}\}_{i=1}^{P_\ell} = \{\hat{v}_{n,i}^{\ell,\mathbf{c}_1}\}_{i=1}^{P_{\ell-1}} \cup \{\hat{v}_{n,i}^{\ell,\mathbf{c}_2}\}_{i=1}^{P_{\ell-1}}$ in the superindex $\mathbf{c}_j \leftrightarrow \mathbf{f}_j$, we introduce

$$
\hat{v}_{n,i}^{\ell,\mathbf{f}_1} := \hat{v}_{n,i}^{\ell,\mathbf{f}} \quad \text{and} \quad \hat{v}_{n,i}^{\ell,\mathbf{f}_2} := \hat{v}_{n,i+P_{\ell-1}}^{\ell,\mathbf{f}} \quad \text{for} \quad i = 1, \ldots, P_{\ell-1},
$$

and

$$
\mu_n^{\ell,\mathbf{f}_j}[\varphi] := E_{P_{\ell-1}}[\varphi(\hat{v}_n^{\ell,\mathbf{f}_j})] = \frac{1}{P_{\ell-1}} \sum_{i=1}^{P_{\ell-1}} \varphi(\hat{v}_{n,i}^{\ell,\mathbf{f}_j}) \quad \text{for} \quad j = 1, 2.
$$

The random variable $\hat{v}_n^{\ell,\mathbf{f}_j}$ has the same driving noise and perturbed observations as $\hat{v}_n^{\ell,\mathbf{c}_j}$ and we note that

$$
\left(\mu_n^{\ell,\mathbf{f}} - \frac{\mu_n^{\ell,\mathbf{c}_1} + \mu_n^{\ell,\mathbf{c}_2}}{2} \right)[\varphi] = E_{P_\ell}[\varphi(\hat{v}_n^{\ell,\mathbf{f}}) - \varphi(\hat{v}_n^{\ell,\mathbf{c}})]
$$

$$
= \frac{1}{2} \sum_{j=1}^{2} \left(\mu_n^{\ell,\mathbf{f}_j} - \mu_n^{\ell,\mathbf{c}_j} \right)[\varphi].
$$

We further introduce the auxiliary mean-field MLEnKF ensemble denoted by $\{\hat{\bar{v}}_{n,1:P_\ell}^{\ell,\mathbf{f}}, \hat{\bar{v}}_{n,1:P_\ell}^{\ell,\mathbf{c}}\}_{\ell=0}^{L}$ where the dynamics for $\hat{\bar{v}}_{n,i}^{\ell,\mathbf{f}}$ and $\hat{\bar{v}}_{n,i}^{\ell,\mathbf{c}}$ are coupled through driving noise-sharing dynamics

$$
\bar{v}_{n+1,i}^{\ell,\mathbf{f}} = \Psi_n^{N_\ell}(\hat{\bar{v}}_{n,i}^{\ell,\mathbf{f}}), \qquad \bar{v}_{n+1,i}^{\ell,\mathbf{c}} = \Psi_n^{N_\ell}(\hat{\bar{v}}_{n,i}^{\ell,\mathbf{c}}) \tag{32}
$$

and perturbed-observation-sharing updates

$$
\hat{\bar{v}}_{n+1,i}^{\ell,\mathbf{f}} = (I - \bar{K}_n^{\ell,\mathbf{f}} H)\bar{v}_{n,i}^{\ell,\mathbf{f}} + \bar{K}_n^{\ell,\mathbf{f}} \tilde{y}_{n,i}^{\ell}, \qquad \hat{\bar{v}}_{n+1,i}^{\ell,\mathbf{c}} = (I - \bar{K}_n^{\ell,\mathbf{c}} H)\bar{v}_{n,i}^{\ell,\mathbf{c}} + \bar{K}_n^{\ell,\mathbf{c}} \tilde{y}_{n,i}^{\ell}, \tag{33}
$$

where $\bar{K}_n^{\ell,\mathbf{f}} := \bar{K}_n^{N_\ell}$ and $\bar{K}_n^{\ell,\mathbf{c}} := \bar{K}_n^{N_{\ell-1}}$, cf. Section A, and with initial conditions equal identical to MLEnKF:

$$
\hat{\bar{v}}_{0,i}^{0,\mathbf{f}} := \hat{v}_{0,i}^{\ell,\mathbf{f}}
$$

$$
\hat{\bar{v}}_{0,i}^{\ell,\mathbf{f}} = \hat{\bar{v}}_{0,i}^{\ell,\mathbf{c}} := \hat{v}_{0,i}^{\ell,\mathbf{f}} = \hat{v}_{0,i}^{\ell,\mathbf{c}} \quad \ell \geq 1.
$$

The particles $(\hat{\bar{v}}_{n,i}^{\ell,\mathbf{f}}, \hat{\bar{v}}_{n,i}^{\ell,\mathbf{c}})$ are thus coupled through sharing the same initial condition, driving noise W, and perturbed observations. The particle pair is also coupled to the MLEnKF pair $(\hat{v}_{n,i}^{\ell,\mathbf{f}}, \hat{v}_{n,i}^{\ell,\mathbf{c}})$ in all the same ways. In other words, $(\hat{\bar{v}}_{n,i}^{\ell,\mathbf{f}}, \hat{\bar{v}}_{n,i}^{\ell,\mathbf{c}})$ is shadowing $(\hat{v}_{n,i}^{\ell,\mathbf{f}}, \hat{v}_{n,i}^{\ell,\mathbf{c}})$. The ensemble $\hat{\bar{v}}_{n,1:P_\ell}^{\ell,\mathbf{f}} = \hat{\bar{v}}_{n,1:P_{\ell-1}}^{\ell,\mathbf{f}_1} \cup \hat{\bar{v}}_{n,1:P_{\ell-1}}^{\ell,\mathbf{f}_2}$ induces an

empirical measure $\bar{\mu}_n^{\ell,\mathbf{f}} = (\bar{\mu}_n^{\ell,\mathbf{f_1}} + \bar{\mu}_n^{\ell,\mathbf{f_2}})/2$ and $\hat{v}_{n,1:P_\ell}^{\ell,\mathbf{c}} = \hat{v}_{n,1:P_{\ell-1}}^{\ell,\mathbf{c_1}} \cup \hat{v}_{n,1:P_{\ell-1}}^{\ell,\mathbf{c_2}}$ induces $\bar{\mu}_n^{\ell,\mathbf{c}} = (\bar{\mu}_n^{\ell,\mathbf{c_1}} + \bar{\mu}_n^{\ell,\mathbf{c_2}})/2$, with the convention that $\bar{\mu}_n^{0,\mathbf{c}} = \bar{\mu}_n^{0,\mathbf{c_1}} = \bar{\mu}_n^{0,\mathbf{c_2}} = 0$ for all $n \geq 0$. Employing sequences $\{M_\ell\}, \{N_\ell\} \subset \mathbb{N}$ and $L \in \mathbb{N}$ with exactly the same values as those used for the MLEnKF estimator (9) we seek to shadow, we define the auxiliary mean-field MLEnKF estimator by

$$\bar{\mu}_n^{\text{ML}}[\varphi] = \sum_{\ell=0}^{L} \frac{1}{M_\ell} \sum_{m=1}^{M_\ell} \left(\bar{\mu}_n^{\ell,\mathbf{f},m} - \bar{\mu}_n^{\ell,\mathbf{c},m} \right)[\varphi] \tag{34}$$

where for $m = 1, \ldots, M_\ell$, $\left(\bar{\mu}_n^{\ell,\mathbf{f},m} - \bar{\mu}_n^{\ell,\mathbf{c},m} \right)[\varphi]$ are iid realizations that are coupled to/shadowing $\left(\mu_n^{\ell,\mathbf{f},m} - \mu_n^{\ell,\mathbf{c},m} \right)[\varphi]$ by sharing the same underlying randomness (initial conditions, driving noise and perturbed observations). And, consequently, $\bar{\mu}_n^{\text{ML}}[\varphi]$ is shadowing $\mu_n^{\text{ML}}[\varphi]$.

Proof of Theorem 2. By the triangle inequality,

$$\left\| \mu_n^{\text{ML}}[\varphi] - \bar{\mu}_n[\varphi] \right\|_p \leq \left\| \mu_n^{\text{ML}}[\varphi] - \bar{\mu}_n^{\text{ML}}[\varphi] \right\|_p + \left\| \bar{\mu}_n^{\text{ML}}[\varphi] - \bar{\mu}_n^{N_L}[\varphi] \right\|_p \\ + \left\| \bar{\mu}_n^{N_L}[\varphi] - \bar{\mu}_n[\varphi] \right\|_p. \tag{35}$$

By (9) and (34), $\Delta\tilde{\mu}_n^\ell[\varphi] := \mathbb{E}\left[\left(\bar{\mu}_n^{\ell,\mathbf{f}} - \mu_n^{\ell,\mathbf{f}} + \mu_n^{\ell,\mathbf{c}} - \bar{\mu}_n^{\ell,\mathbf{c}} \right)[\varphi] \right]$ and the M-Z inequality,

$$\|\mu_n^{\text{ML}}[\varphi] - \bar{\mu}_n^{\text{ML}}[\varphi]\|_p \leq \left\| \sum_{\ell=0}^{L} \Delta\tilde{\mu}_n^\ell[\varphi] \right\|_p$$

$$+ \left\| \sum_{\ell=0}^{L} \sum_{m=1}^{M_\ell} \frac{\left(\bar{\mu}_n^{\ell,\mathbf{f},m} - \mu_n^{\ell,\mathbf{f},m} + \mu_n^{\ell,\mathbf{c},m} - \bar{\mu}_n^{\ell,\mathbf{c},m} \right)[\varphi] - \Delta\tilde{\mu}_n^\ell[\varphi]}{M_\ell} \right\|_p$$

$$\lesssim \left\| \mathbb{E}\left[\left(\bar{\mu}_n^{L,\mathbf{f}} - \mu_n^{L,\mathbf{f}} \right)[\varphi] \right] \right\|_p + \sum_{\ell=0}^{L} M_\ell^{-1/2} \left\| \left(\bar{\mu}_n^{\ell,\mathbf{f}} - \mu_n^{\ell,\mathbf{f}} + \mu_n^{\ell,\mathbf{c}} - \bar{\mu}_n^{\ell,\mathbf{c}} \right)[\varphi] \right\|_p,$$

where we used that $\mathbb{E}\left[\mu_n^{\ell,\mathbf{c}} \right] = \mathbb{E}\left[\mu_n^{\ell-1,\mathbf{f}} \right]$ and $\mathbb{E}\left[\bar{\mu}_n^{\ell,\mathbf{c}} \right] = \mathbb{E}\left[\bar{\mu}_n^{\ell-1,\mathbf{f}} \right]$ for $\ell \geq 1$ in the last inequality. The $L^p(\Omega)$-convergence $\mu_n^{L,\mathbf{f}}[\varphi] \to \bar{\mu}_n^{L,\mathbf{f}}[\varphi]$ as $L \to \infty$, cf. Theorem 1, further implies that

$$\left(\bar{\mu}_n^{L,\mathbf{f}} - \mu_n^{L,\mathbf{f}} \right)[\varphi] = - \sum_{\ell=L+1}^{\infty} \mathbb{E}\left[\left(\bar{\mu}_n^{\ell,\mathbf{f}} - \mu_n^{\ell,\mathbf{f}} + \mu_n^{\ell,\mathbf{c}} - \bar{\mu}_n^{\ell,\mathbf{c}} \right)[\varphi] \right],$$

and Jensen's inequality and Lemma 11 yield

$$\|\mu_n^{\text{ML}}[\varphi] - \bar{\mu}_n^{\text{ML}}[\varphi]\|_p \lesssim \sum_{\ell=0}^{L} M_\ell^{-1/2} \left\| \left(\bar{\mu}_n^{\ell,\mathbf{f}} - \mu_n^{\ell,\mathbf{f}} + \mu_n^{\ell,\mathbf{c}} - \bar{\mu}_n^{\ell,\mathbf{c}} \right)[\varphi] \right\|_p$$

$$+ \sum_{\ell=L+1}^{\infty} \left\| \left(\bar{\mu}_n^{\ell,\mathbf{f}} - \mu_n^{\ell,\mathbf{f}} + \mu_n^{\ell,\mathbf{c}} - \bar{\mu}_n^{\ell,\mathbf{c}} \right)[\varphi] \right\|_p$$

$$\lesssim N_L^{-\beta/2} P_L^{-1/2} + P_L^{-1} + \sum_{\ell=0}^{L} M_\ell^{-1/2} (N_\ell^{-\beta/2} P_\ell^{-1/2} + P_\ell^{-1}).$$

For the second term in (35), note first that

$$\mathbb{E}\left[\bar{\mu}_n^{N_L}[\varphi] \right] = \mathbb{E}\left[\bar{\mu}_n^{L,\mathbf{f}}[\varphi] \right] = \sum_{\ell=0}^{L} \mathbb{E}\left[(\bar{\mu}_n^{\ell,\mathbf{f}} - \bar{\mu}_n^{\ell,\mathbf{c}})[\varphi] \right] = \sum_{\ell=0}^{L} \mathbb{E}\left[\varphi(\hat{v}_n^{\ell,\mathbf{f}}) - \varphi(\hat{v}_n^{\ell,\mathbf{c}}) \right].$$

By applying the M-Z inequality twice (first in M_ℓ and thereafter in P_ℓ) and Lemma 6,

$$\left\|\bar{\mu}_n^{\mathrm{ML}}[\varphi] - \bar{\mu}_n^{N_L}[\varphi]\right\|_p \leq \sum_{\ell=0}^{L} \left\|\sum_{m=1}^{M_\ell} \frac{(\bar{\mu}_n^{\ell,\mathbf{f},m} - \bar{\mu}_n^{\ell,\mathbf{c},m})[\varphi] - \mathbb{E}\left[(\bar{\mu}_n^{\ell,\mathbf{f}} - \bar{\mu}_n^{\ell,\mathbf{c}})[\varphi]\right]}{M_\ell}\right\|_p$$

$$\lesssim \sum_{\ell=0}^{L} M_\ell^{-1/2} \left\|E_{P_\ell}[\varphi(\hat{\bar{v}}_n^{\ell,\mathbf{f}}) - \varphi(\hat{\bar{v}}_n^{\ell,\mathbf{c}}) - \mathbb{E}\left[\varphi(\hat{\bar{v}}_n^{\ell,\mathbf{f}}) - \varphi(\hat{\bar{v}}_n^{\ell,\mathbf{c}})\right]\right\|_p$$

$$\lesssim \sum_{\ell=0}^{L} M_\ell^{-1/2} P_\ell^{-1/2} \left\|\varphi(\hat{\bar{v}}_n^{\ell,\mathbf{f}}) - \varphi(\hat{\bar{v}}_n^{\ell,\mathbf{c}})\right\|_p$$

$$\lesssim \sum_{\ell=0}^{L} M_\ell^{-1/2} P_\ell^{-1/2} N_\ell^{-\beta/2}.$$

For the third term in (35), it follows by (31) that

$$\left\|\bar{\mu}_n^{N_L}[\varphi] - \bar{\mu}_n[\varphi]\right\|_p = \left|\mathbb{E}\left[\varphi(\hat{\bar{v}}_n^{L,\mathbf{f}}) - \varphi(\hat{\bar{v}}_n)\right]\right| \lesssim N_L^{-\alpha}.$$

$\square$

Lemma 4 (Continuity of mean-field Kalman gains). *It holds that*

$$|\bar{K}_n^{\ell,\mathbf{c}} - \bar{K}_n^{\ell,\mathbf{f}}|_2 \leq |\Gamma^{-1}|_2 |H|_2 \left(1 + 2|\bar{K}_n^{\ell,\mathbf{f}} H|_2\right) |\bar{C}_n^{\ell,\mathbf{c}} - \bar{C}_n^{\ell,\mathbf{f}}|_2 \lesssim |\bar{C}_n^{\ell,\mathbf{c}} - \bar{C}_n^{\ell,\mathbf{f}}|_2.$$

Proof. The proof is analogous to [22, Lemma 3.4]. $\square$

Lemma 5 (Continuity of mean-field covariance matrices). *If Assumptions 1 and 2 hold, then for any triplet of sequences $\{M_\ell\}, \{N_\ell\}, \{P_\ell\} \subset \mathbb{N}$ described in Section 2.4, $L \geq 0$ and $n \geq 1$ it holds that*

$$|\bar{C}_n^{\ell,\mathbf{c}} - \bar{C}_n^{\ell,\mathbf{f}}|_2 \lesssim \left\|\bar{v}_n^{\ell,\mathbf{c}} - \bar{v}_n^{\ell,\mathbf{f}}\right\|_4.$$

Proof. Using that $\bar{v}_n^{\ell,\mathbf{c}}, \bar{v}_n^{\ell,\mathbf{f}} \in L^4(\Omega, \mathbb{R}^d)$, Hölder's and Jensen's inequalities yield that

$$\begin{aligned}\left|\bar{C}_n^{\ell,\mathbf{c}} - \bar{C}_n^{\ell,\mathbf{f}}\right|_2 &\leq \left|\mathbb{E}\left[\bar{v}_n^{\ell,\mathbf{c}}(\bar{v}_n^{\ell,\mathbf{c}})^{\mathbf{T}} - \bar{v}_n^{\ell,\mathbf{f}}(\bar{v}_n^{\ell,\mathbf{f}})^{\mathbf{T}}\right]\right|_2 \\ &\quad + \left|\mathbb{E}\left[\bar{v}_n^{\ell,\mathbf{c}}\right] \mathbb{E}\left[(\bar{v}_n^{\ell,\mathbf{c}})^{\mathbf{T}}\right] - \mathbb{E}\left[\bar{v}_n^{\ell,\mathbf{f}}\right] \mathbb{E}\left[(\bar{v}_n^{\ell,\mathbf{f}})^{\mathbf{T}}\right]\right|_2 \\ &\lesssim \left\|\bar{v}_n^{\ell,\mathbf{c}} - \bar{v}_n^{\ell,\mathbf{f}}\right\|_4.\end{aligned}$$

$\square$

In the remaining part of this section, we will assume that Assumptions 1 and 2 hold, and that the sequences $\{N_\ell\}, \{P_\ell\} \subset \mathbb{N}$ satisfy the constraints given in Section 2.4 (namely, $P_\ell = 2P_{\ell-1}$ and $\{N_\ell\}$ exponentially increasing).

Lemma 6 (Stability of mean-field particles). *For any $n \geq 0$ and $p \geq 2$, it holds that*

$$\max\left(\left\|\hat{\bar{v}}_n^{\ell,\mathbf{f}} - \hat{\bar{v}}_n^{\ell,\mathbf{c}}\right\|, \left\|\bar{v}_{n+1}^{\ell,\mathbf{f}} - \bar{v}_{n+1}^{\ell,\mathbf{c}}\right\|_p\right) \lesssim N_\ell^{-\beta/2},$$

for multilevel mean-field prediction and update particles defined as in (32) and (33).

Proof. Since $\hat{\bar{v}}_0^{\ell,\mathbf{f}} = \hat{\bar{v}}_0^{\ell,\mathbf{c}}$, we may assume that for some $n \geq 1$,

$$\left\|\hat{\bar{v}}_{n-1}^{\ell,\mathbf{f}} - \hat{\bar{v}}_{n-1}^{\ell,\mathbf{c}}\right\|_p \lesssim N_\ell^{-\beta/2}.$$

Assumption 1(iii) and Assumption 2(iii) imply that

$$\left\|\bar{v}_n^{\ell,\mathbf{f}} - \bar{v}_n^{\ell,\mathbf{c}}\right\|_p \le \left\|\Psi_{n-1}^{N_\ell}(\hat{\bar{v}}_{n-1}^{\ell,\mathbf{f}}) - \Psi_{n-1}^{N_{\ell-1}}(\hat{\bar{v}}_{n-1}^{\ell,\mathbf{f}})\right\|_p + \left\|\Psi_{n-1}^{N_{\ell-1}}(\hat{\bar{v}}_{n-1}^{\ell,\mathbf{f}}) - \Psi_{n-1}^{N_{\ell-1}}(\hat{\bar{v}}_{n-1}^{\ell,\mathbf{c}})\right\|_p$$
$$\lesssim N_\ell^{-\beta/2},$$

and Lemmas 4 and 5 and $|\bar{v}_n^{\ell,\mathbf{f}}|, |\tilde{y}_n^{\ell}| \in \cap_{r \ge 2} L^r(\Omega)$ yield that

$$\left\|\hat{\bar{v}}_n^{\ell,\mathbf{f}} - \hat{\bar{v}}_n^{\ell,\mathbf{c}}\right\|_p \le \left|I - \bar{K}_n^{\ell,\mathbf{c}} H\right|_2 \left\|\bar{v}_n^{\ell,\mathbf{f}} - \bar{v}_n^{\ell,\mathbf{c}}\right\|_p + \left|\bar{K}_n^{\ell,\mathbf{c}} - \bar{K}_n^{\ell,\mathbf{f}}\right|_2 \left\|H\bar{v}_n^{\ell,\mathbf{f}} + \tilde{y}_n^\ell\right\|_p$$
$$\lesssim N_\ell^{-\beta/2}.$$

The statement holds by induction. $\qquad\square$

Corollary 3 (Continuity of mean-field and EnKF covariance matrices). *For any* $n \ge 0$ *and* $p \ge 2$, *it holds that*

$$\left\|C_n^{\ell,\mathbf{f}} - \bar{C}_n^{\ell,\mathbf{f}}\right\|_p \lesssim \left\|v_n^{\ell,\mathbf{f}} - \bar{v}_n^{\ell,\mathbf{f}}\right\|_{2p} + P_\ell^{-1/2},$$
$$\left\|C_n^{\ell,\mathbf{c}_k} - \bar{C}_n^{\ell,\mathbf{c}}\right\|_p \lesssim \left\|v_n^{\ell,\mathbf{c}} - \bar{v}_n^{\ell,\mathbf{c}}\right\|_{2p} + P_\ell^{-1/2}, \quad k = 1,2,$$

for multilevel prediction particles defined as in (10) *and* (32).

Proof. Since $C_n^{\ell,\mathbf{f}} = C_n^{N_\ell,P_\ell}$, $\bar{C}_n^{\ell,\mathbf{f}} = \bar{C}_n^{N_\ell}$, $C_n^{\ell,\mathbf{c}_k} \overset{D}{=} C_n^{N_{\ell-1},P_{\ell-1}}$ and $\bar{C}_n^{\ell,\mathbf{c}} = \bar{C}_n^{N_{\ell-1}}$, the result follows from Lemma 2. $\qquad\square$

Corollary 4 (Distance between ensembles II). *For any* $n \ge 0$ *and* $p \ge 2$, *the following asymptotic inequality holds*

$$\max\left(\left\|\hat{v}_n^{\ell,\mathbf{f}} - \hat{\bar{v}}_n^{\ell,\mathbf{f}}\right\|_p, \left\|\hat{v}_n^{\ell,\mathbf{c}} - \hat{\bar{v}}_n^{\ell,\mathbf{c}}\right\|_p\right) \lesssim P_\ell^{-1/2}.$$

for multilevel update particles defined as in (12) *and* (33).

Proof. Since $\hat{v}_n^{\ell,\mathbf{f}} - \hat{\bar{v}}_n^{\ell,\mathbf{f}} \overset{D}{=} \hat{v}_n^{N_\ell,P_\ell} - \hat{\bar{v}}_n^{N_\ell}$ and $\hat{v}_n^{\ell,\mathbf{c}} - \hat{\bar{v}}_n^{\ell,\mathbf{c}} \overset{D}{=} \hat{v}_n^{N_{\ell-1},P_\ell} - \hat{\bar{v}}_n^{N_{\ell-1}}$, the result follows from Lemma 3. $\qquad\square$

Lemma 7 (Continuity of Kalman gain double differences). *For any* $n \ge 0$ *and* $p \ge 2$, *it holds that*

$$\left\|K_n^{\ell,\mathbf{f}} - \frac{K_n^{\ell,\mathbf{c}_1} + K_n^{\ell,\mathbf{c}_2}}{2} - \bar{K}_n^{\ell,\mathbf{f}} + \bar{K}_n^{\ell,\mathbf{c}}\right\|_p \lesssim \left\|C_n^{\ell,\mathbf{f}} - \frac{C_n^{\ell,\mathbf{c}_1} + C_n^{\ell,\mathbf{c}_2}}{2} - \bar{C}_n^{\ell,\mathbf{f}} + \bar{C}_n^{\ell,\mathbf{c}}\right\|_p$$
$$+ P_\ell^{-1/2} N_\ell^{-\beta/2} + P_\ell^{-1}.$$
$$(36)$$

Proof. From the proof of [22, Lemma 3.4], one may deduce that

$$K_n^{\ell,\mathbf{f}} - \bar{K}_n^{\ell,\mathbf{f}} = K_n^{\ell,\mathbf{f}} H (\bar{C}_n^{\ell,\mathbf{f}} - C_n^{\ell,\mathbf{f}}) H^{\mathbf{T}} (H\bar{C}_n^{\ell,\mathbf{f}} H^{\mathbf{T}} + \Gamma)^{-1}$$
$$+ (C_n^{\ell,\mathbf{f}} - \bar{C}_n^{\ell,\mathbf{f}}) H^{\mathbf{T}} (H\bar{C}_n^{\ell,\mathbf{f}} H^{\mathbf{T}} + \Gamma)^{-1},$$
$$K_n^{\ell,\mathbf{c}_j} - \bar{K}_n^{\ell,\mathbf{c}} = K_n^{\ell,\mathbf{c}_j} H (\bar{C}_n^{\ell,\mathbf{c}} - C_n^{\ell,\mathbf{c}_j}) H^{\mathbf{T}} (H\bar{C}_n^{\ell,\mathbf{c}} H^{\mathbf{T}} + \Gamma)^{-1}$$
$$+ (C_n^{\ell,\mathbf{c}_j} - \bar{C}_n^{\ell,\mathbf{c}}) H^{\mathbf{T}} (H\bar{C}_n^{\ell,\mathbf{c}} H^{\mathbf{T}} + \Gamma)^{-1},$$

and

$$(H\bar{C}_n^{\ell,\mathbf{f}} H^{\mathbf{T}} + \Gamma)^{-1} = (H\bar{C}_n^{\ell,\mathbf{c}} H^{\mathbf{T}} + \Gamma)^{-1}$$
$$+ (H\bar{C}_n^{\ell,\mathbf{c}} H^{\mathbf{T}} + \Gamma)^{-1} (\bar{C}_n^{\ell,\mathbf{c}} - \bar{C}_n^{\ell,\mathbf{f}}) H^{\mathbf{T}} (H\bar{C}_n^{\ell,\mathbf{f}} H^{\mathbf{T}} + \Gamma)^{-1}.$$

The above equations imply that

$$K_n^{\ell,\mathbf{f}} - \bar{K}_n^{\ell,\mathbf{f}} = K_n^{\ell,\mathbf{f}} H(\bar{C}_n^{\ell,\mathbf{f}} - C_n^{\ell,\mathbf{f}}) H^{\mathbf{T}} (H\bar{C}_n^{\ell,\mathbf{c}} H^{\mathbf{T}} + \Gamma)^{-1}$$
$$+ K_n^{\ell,\mathbf{f}} H(\bar{C}_n^{\ell,\mathbf{f}} - C_n^{\ell,\mathbf{f}}) H^{\mathbf{T}} (H\bar{C}_n^{\ell,\mathbf{c}} H^{\mathbf{T}} + \Gamma)^{-1} (\bar{C}_n^{\ell,\mathbf{c}} - \bar{C}_n^{\ell,\mathbf{f}}) H^{\mathbf{T}} (H\bar{C}_n^{\ell,\mathbf{f}} H^{\mathbf{T}} + \Gamma)^{-1}$$
$$+ (C_n^{\ell,\mathbf{f}} - \bar{C}_n^{\ell,\mathbf{f}}) H^{\mathbf{T}} (H\bar{C}_n^{\ell,\mathbf{f}} H^{\mathbf{T}} + \Gamma)^{-1},$$

and

$$K_n^{\ell,\mathbf{f}} - \frac{K_n^{\ell,\mathbf{c}_1} + K_n^{\ell,\mathbf{c}_2}}{2} - (\bar{K}_n^{\ell,\mathbf{f}} - \bar{K}_n^{\ell,\mathbf{c}})$$

$$= K_n^{\ell,\mathbf{f}} - \bar{K}_n^{\ell,\mathbf{f}} - \frac{1}{2}\left(K_n^{\ell,\mathbf{c}_1} - \bar{K}_n^{\ell,\mathbf{c}} + K_n^{\ell,\mathbf{c}_2} - \bar{K}_n^{\ell,\mathbf{c}}\right)$$

$$= (I - K_n^{\ell,\mathbf{f}} H)\left(C_n^{\ell,\mathbf{f}} - \frac{C_n^{\ell,\mathbf{c}_1} + C_n^{\ell,\mathbf{c}_2}}{2} - \bar{C}_n^{\ell,\mathbf{f}} + \bar{C}_n^{\ell,\mathbf{c}}\right) H^{\mathbf{T}} (H\bar{C}_n^{\ell,\mathbf{c}} H^{\mathbf{T}} + \Gamma)^{-1}$$

$$+ (I - K_n^{\ell,\mathbf{f}} H)(\bar{C}_n^{\ell,\mathbf{f}} - C_n^{\ell,\mathbf{f}}) H^{\mathbf{T}} (H\bar{C}_n^{\ell,\mathbf{c}} H^{\mathbf{T}} + \Gamma)^{-1} H(\bar{C}_n^{\ell,\mathbf{c}} - C_n^{\ell,\mathbf{f}}) H^{\mathbf{T}} \times$$

$$(H\bar{C}_n^{\ell,\mathbf{f}} H^{\mathbf{T}} + \Gamma)^{-1} - (K_n^{\ell,\mathbf{c}_1} - K_n^{\ell,\mathbf{f}}) H(\bar{C}_n^{\ell,\mathbf{c}} - \frac{C_n^{\ell,\mathbf{c}_1} + C_n^{\ell,\mathbf{c}_2}}{2}) H^{\mathbf{T}} (H\bar{C}_n^{\ell,\mathbf{c}} H^{\mathbf{T}} + \Gamma)^{-1}$$

$$- \frac{1}{2}(K_n^{\ell,\mathbf{c}_2} - K_n^{\ell,\mathbf{c}_1}) H(\bar{C}_n^{\ell,\mathbf{c}} - C_n^{\ell,\mathbf{c}_2}) H^{\mathbf{T}} (H\bar{C}_n^{\ell,\mathbf{c}} H^{\mathbf{T}} + \Gamma)^{-1}.$$

By the positive definiteness of Γ, Lemmas 4, 5 and 6, and corollaries 3 and 4,

$$\left\| K_n^{\ell,\mathbf{f}} - \frac{K_n^{\ell,\mathbf{c}_1} + K_n^{\ell,\mathbf{c}_2}}{2} - (\bar{K}_n^{\ell,\mathbf{f}} - \bar{K}_n^{\ell,\mathbf{c}}) \right\|_p \lesssim \left\| C_n^{\ell,\mathbf{f}} - \frac{C_n^{\ell,\mathbf{c}_1} + C_n^{\ell,\mathbf{c}_2}}{2} - \bar{C}_n^{\ell,\mathbf{f}} + \bar{C}_n^{\ell,\mathbf{c}} \right\|_p$$

$$+ \left\| \bar{C}_n^{\ell,\mathbf{f}} - C_n^{\ell,\mathbf{f}} \right\|_p \left| \bar{C}_n^{\ell,\mathbf{c}} - \bar{C}_n^{\ell,\mathbf{f}} \right|_2 + \left\| K_n^{\ell,\mathbf{c}_1} - K_n^{\ell,\mathbf{f}} \right\|_p \left\| \bar{C}_n^{\ell,\mathbf{c}} - \frac{C_n^{\ell,\mathbf{c}_1} + C_n^{\ell,\mathbf{c}_2}}{2} \right\|_p$$

$$+ \frac{1}{2} \left\| K_n^{\ell,\mathbf{c}_2} - K_n^{\ell,\mathbf{c}_1} \right\|_p \left\| \bar{C}_n^{\ell,\mathbf{c}} - C_n^{\ell,\mathbf{c}_2} \right\|_p$$

$$\lesssim \left\| C_n^{\ell,\mathbf{f}} - \frac{C_n^{\ell,\mathbf{c}_1} + C_n^{\ell,\mathbf{c}_2}}{2} - \bar{C}_n^{\ell,\mathbf{f}} + \bar{C}_n^{\ell,\mathbf{c}} \right\|_p + P_\ell^{-1/2} N_\ell^{-\beta/2} + P_\ell^{-1}$$

where the last inequality follows from

$$\left\| K_n^{\ell,\mathbf{c}_1} - K_n^{\ell,\mathbf{f}} \right\|_p \lesssim \left\| C_n^{\ell,\mathbf{c}_1} - C_n^{\ell,\mathbf{f}} \right\|_p$$
$$\leq \left\| C_n^{\ell,\mathbf{c}_1} - \bar{C}_n^{\ell,\mathbf{c}} \right\|_p + \left| \bar{C}_n^{\ell,\mathbf{c}} - \bar{C}_n^{\ell,\mathbf{f}} \right|_2 + \left\| \bar{C}_n^{\ell,\mathbf{f}} - C_n^{\ell,\mathbf{f}} \right\|_p$$

and

$$\left\| K_n^{\ell,\mathbf{c}_1} - K_n^{\ell,\mathbf{c}_2} \right\|_p \lesssim \left\| C_n^{\ell,\mathbf{c}_1} - C_n^{\ell,\mathbf{c}_2} \right\|_p \leq \left\| C_n^{\ell,\mathbf{c}_1} - \bar{C}_n^{\ell,\mathbf{c}} \right\|_p + \left\| \bar{C}_n^{\ell,\mathbf{c}} - C_n^{\ell,\mathbf{c}_2} \right\|_p. \quad (37)$$

$\square$

We next show how to bound the first term on the right hand side of (36).

Lemma 8 (Continuity of covariance matrix double differences)**.** *For any $n \geq 0$ and $p \geq 2$, the following asymptotic inequality holds:*

$$\left\| C_n^{\ell,\mathbf{f}} - \frac{C_n^{\ell,\mathbf{c}_1} + C_n^{\ell,\mathbf{c_2}}}{2} - \bar{C}_n^{\ell,\mathbf{f}} + \bar{C}_n^{\ell,\mathbf{c}} \right\|_p$$

$$\lesssim \left\| E_{P_\ell} \left[\left(v_n^{\ell,\mathbf{f}_j} - v_n^{\ell,\mathbf{c}_j} - \bar{v}_n^{\ell,\mathbf{f}_j} + \bar{v}_n^{\ell,\mathbf{c}_j} \right) \left(\bar{v}_n^{\ell,\mathbf{f}_j} \right)^{\mathbf{T}} \right] \right\|_p$$

$$+ \left\| E_{P_\ell} [v_n^{\ell,\mathbf{f}_j} - v_n^{\ell,\mathbf{c}_j} - \bar{v}_n^{\ell,\mathbf{f}_j} + \bar{v}_n^{\ell,\mathbf{c}_j}] \right\|_{2p}$$

$$+ P_\ell^{-1/2} N_\ell^{-\beta/2} + P_\ell^{-1}.$$

Proof. Let us first recall that

$$C_n^{\ell,\mathbf{f}} = \overline{\mathrm{Cov}}[v_n^{\ell,\mathbf{f}}], \quad \bar{C}_n^{\ell,\mathbf{f}} = \mathrm{Cov}[\bar{v}_n^{\ell,\mathbf{f}}],$$
$$C_n^{\ell,\mathbf{c}_j} = \overline{\mathrm{Cov}}[v_n^{\ell,\mathbf{c}_j}], \quad \bar{C}_n^{\ell,\mathbf{c}} = \mathrm{Cov}[\bar{v}_n^{\ell,\mathbf{c}}],$$

and introduce the following covariance matrices for the auxiliary mean-field multi-level EnKF ensemble

$$\tilde{\bar{C}}_n^{\ell,\mathbf{f}} := \overline{\mathrm{Cov}}[\bar{v}_n^{\ell,\mathbf{f}}], \quad \tilde{\bar{C}}_n^{\ell,\mathbf{c}_j} := \overline{\mathrm{Cov}}[\bar{v}_n^{\ell,\mathbf{c}_j}].$$

By the triangle inequality,

$$\left\| C_n^{\ell,\mathbf{f}} - \frac{C_n^{\ell,\mathbf{c}_1} + C_n^{\ell,\mathbf{c}_2}}{2} - \bar{C}_n^{\ell,\mathbf{f}} + \bar{C}_n^{\ell,\mathbf{c}} \right\|_p$$
$$\lesssim \left\| C_n^{\ell,\mathbf{f}} - \frac{C_n^{\ell,\mathbf{c}_1} + C_n^{\ell,\mathbf{c}_2}}{2} - \tilde{\bar{C}}_n^{\ell,\mathbf{f}} + \frac{\tilde{\bar{C}}_n^{\ell,\mathbf{c}_1} + \tilde{\bar{C}}_n^{\ell,\mathbf{c}_2}}{2} \right\|_p \tag{38}$$
$$+ \left\| \tilde{\bar{C}}_n^{\ell,\mathbf{f}} - \frac{\tilde{\bar{C}}_n^{\ell,\mathbf{c}_1} + \tilde{\bar{C}}_n^{\ell,\mathbf{c}_2}}{2} - \bar{C}_n^{\ell,\mathbf{f}} + \frac{\bar{C}_n^{\ell,\mathbf{c}} + \bar{C}_n^{\ell,\mathbf{c}}}{2} \right\|_p.$$

Using that

$$E_{P_{\ell-1}}\left[\sum_{j=1}^{2} \frac{v_n^{\ell,\mathbf{c}_j}(v_n^{\ell,\mathbf{c}_j})^{\mathbf{T}}}{2} \right] = E_{P_\ell}[v_n^{\ell,\mathbf{c}}(v_n^{\ell,\mathbf{c}})^{\mathbf{T}}]$$

$$\text{and} \quad E_{P_{\ell-1}}\left[\sum_{j=1}^{2} \frac{\bar{v}_n^{\ell,\mathbf{c}_j}(\bar{v}_n^{\ell,\mathbf{c}_j})^{\mathbf{T}}}{2} \right] = E_{P_\ell}[\bar{v}_n^{\ell,\mathbf{c}}(\bar{v}_n^{\ell,\mathbf{c}})^{\mathbf{T}}],$$

we obtain

$$\left\| C_n^{\ell,\mathbf{f}} - \frac{C_n^{\ell,\mathbf{c}_1} + C_n^{\ell,\mathbf{c}_2}}{2} - \tilde{\bar{C}}_n^{\ell,\mathbf{f}} + \frac{\tilde{\bar{C}}_n^{\ell,\mathbf{c}_1} + \tilde{\bar{C}}_n^{\ell,\mathbf{c}_2}}{2} \right\|_p$$
$$\leq \left\| E_{P_\ell}[v_n^{\ell,\mathbf{f}}(v_n^{\ell,\mathbf{f}})^{\mathbf{T}} - \bar{v}_n^{\ell,\mathbf{f}}(\bar{v}_n^{\ell,\mathbf{f}})^{\mathbf{T}}] - E_{P_\ell}[v_n^{\ell,\mathbf{c}}(v_n^{\ell,\mathbf{c}})^{\mathbf{T}} - \bar{v}_n^{\ell,\mathbf{c}}(\bar{v}_n^{\ell,\mathbf{c}})^{\mathbf{T}}] \right\|_p$$
$$+ \left\| \mathbb{E}_{P_\ell}[v_n^{\ell,\mathbf{f}}]E_{P_\ell}[(v_n^{\ell,\mathbf{f}})]^{\mathbf{T}} - E_{P_\ell}[\bar{v}_n^{\ell,\mathbf{f}}]E_{P_\ell}[(\bar{v}_n^{\ell,\mathbf{f}})]^{\mathbf{T}} \right.$$
$$- \frac{1}{2}\Big(E_{P_{\ell-1}}[v_n^{\ell,\mathbf{c}_1}]E_{P_{\ell-1}}[v_n^{\ell,\mathbf{c}_1}]^{\mathbf{T}} + E_{P_{\ell-1}}[v_n^{\ell,\mathbf{c}_2}]E_{P_{\ell-1}}[v_n^{\ell,\mathbf{c}_2}]^{\mathbf{T}}$$
$$\left. - E_{P_{\ell-1}}[\bar{v}_n^{\ell,\mathbf{c}_1}]E_{P_{\ell-1}}[\bar{v}_n^{\ell,\mathbf{c}_1}]^{\mathbf{T}} \quad - E_{P_{\ell-1}}[\bar{v}_n^{\ell,\mathbf{c}_2}]E_{P_{\ell-1}}[\bar{v}_n^{\ell,\mathbf{c}_2}]^{\mathbf{T}} \Big) \right\|_p$$
$$=: \mathcal{I}_{11} + \mathcal{I}_{12}.$$

For the first term, Lemma 6 and Corollary 4 yield

$$\mathcal{I}_{11} \lesssim \left\| E_{P_\ell}[(v_n^{\ell,\mathbf{f}} - v_n^{\ell,\mathbf{c}} - \bar{v}_n^{\ell,\mathbf{f}} + \bar{v}_n^{\ell,\mathbf{c}})(\bar{v}_n^{\ell,\mathbf{f}})^{\mathbf{T}}] \right\|_p$$
$$+ \left\| v_n^{\ell,\mathbf{f}} - v_n^{\ell,\mathbf{c}} - \bar{v}_n^{\ell,\mathbf{f}} + \bar{v}_n^{\ell,\mathbf{c}} \right\|_{2p} \left\| v_n^{\ell,\mathbf{f}} - \bar{v}_n^{\ell,\mathbf{f}} \right\|_{2p}$$
$$+ \left\| v_n^{\ell,\mathbf{f}} - v_n^{\ell,\mathbf{c}} \right\|_{2p} \left\| v_n^{\ell,\mathbf{c}} - \bar{v}_n^{\ell,\mathbf{c}} \right\|_{2p} + \left\| v_n^{\ell,\mathbf{f}} - \bar{v}_n^{\ell,\mathbf{f}} \right\|_{2p} \left\| \bar{v}_n^{\ell,\mathbf{f}} - \bar{v}_n^{\ell,\mathbf{c}} \right\|_{2p}$$
$$\lesssim \left\| E_{P_\ell}[(v_n^{\ell,\mathbf{f}} - v_n^{\ell,\mathbf{c}} - \bar{v}_n^{\ell,\mathbf{f}} + \bar{v}_n^{\ell,\mathbf{c}})(\bar{v}_n^{\ell,\mathbf{f}})^{\mathbf{T}}] \right\|_p + P_\ell^{-1/2}N_\ell^{-\beta/2} + P_\ell^{-1}.$$

For the second term, the identity $aa^{\mathbf{T}} + bb^{\mathbf{T}} = \frac{1}{2}\left[(a+b)(a+b)^{\mathbf{T}} + (a-b)(a-b)^{\mathbf{T}}\right]$ yields

$$\begin{aligned}
\mathcal{I}_{12} \leq{}& \left\| E_{P_\ell}[v_n^{\ell,\mathbf{f}}] E_{P_\ell}[(v_n^{\ell,\mathbf{f}})^{\mathbf{T}}] - E_{P_\ell}[\bar{v}_n^{\ell,\mathbf{f}}] E_{P_\ell}[(\bar{v}_n^{\ell,\mathbf{f}})^{\mathbf{T}}] \right. \\
&\left. - E_{P_\ell}[v_n^{\ell,\mathbf{c}}] E_{P_\ell}[(v_n^{\ell,\mathbf{c}})]^{\mathbf{T}} + E_{P_\ell}[\bar{v}_n^{\ell,\mathbf{c}}] E_{P_\ell}[(\bar{v}_n^{\ell,\mathbf{c}})]^{\mathbf{T}} \right\|_p \\
&+ \left\| \frac{1}{4} E_{P_{\ell-1}}[v_n^{\ell,\mathbf{c_1}} - v_n^{\ell,\mathbf{c_2}}] E_{P_{\ell-1}}[v_n^{\ell,\mathbf{c_1}} - v_n^{\ell,\mathbf{c_2}}]^{\mathbf{T}} \right. \\
&\left. + \frac{1}{4} E_{P_{\ell-1}}[\bar{v}_n^{\ell,\mathbf{c_1}} - \bar{v}_n^{\ell,\mathbf{c_2}}] E_{P_{\ell-1}}[\bar{v}_n^{\ell,\mathbf{c_1}} - \bar{v}_n^{\ell,\mathbf{c_2}}]^{\mathbf{T}} \right\|_p =: \mathcal{I}_{121} + \mathcal{I}_{122}
\end{aligned}$$

The term $\mathcal{I}_{121}$ can be bounded in a similar fashion as $\mathcal{I}_{11}$, and to bound the second term, we employ the identity $aa^{\mathbf{T}} - bb^{\mathbf{T}} = \frac{1}{2}[(a+b)(a-b)^{\mathbf{T}} + (a-b)(a+b)^{\mathbf{T}}]$, Jensen's inequality and Corollary 4:

$$\begin{aligned}
\mathcal{I}_{122} \lesssim{}& \left\| E_{P_{\ell-1}}[v_n^{\ell,\mathbf{c_1}} - v_n^{\ell,\mathbf{c_2}} + \bar{v}_n^{\ell,\mathbf{c_1}} - \bar{v}_n^{\ell,\mathbf{c_2}}] \right\|_{2p} \\
&\times \left\| E_{P_{\ell-1}}[v_n^{\ell,\mathbf{c_1}} - v_n^{\ell,\mathbf{c_2}} - \bar{v}_n^{\ell,\mathbf{c_1}} + \bar{v}_n^{\ell,\mathbf{c_2}}] \right\|_{2p} \\
&\lesssim \left\| v_n^{\ell,\mathbf{c}} - \bar{v}_n^{\ell,\mathbf{c}} \right\|_{2p}^2 \lesssim P_\ell^{-1}.
\end{aligned}$$

Consequently,

$$\mathcal{I}_{12} \lesssim \left\| E_{P_\ell}[v_n^{\ell,\mathbf{f}} - v_n^{\ell,\mathbf{c}} - \bar{v}_n^{\ell,\mathbf{f}} + \bar{v}_n^{\ell,\mathbf{c}}] \right\|_{2p} + P_\ell^{-1/2} N_\ell^{-\beta/2} + P_\ell^{-1}.$$

For the second term in (38), the equation $\mathbb{E}_{P_\ell}[\bar{v}_n^{\ell,\mathbf{c}}] = \mathbb{E}_{P_{\ell-1}}[(\bar{v}_n^{\ell,\mathbf{c_1}} + \bar{v}_n^{\ell,\mathbf{c_2}})/2]$ implies that

$$\begin{aligned}
\left\| \tilde{\bar{C}}_n^{\ell,\mathbf{f}} \right. &- \left. \frac{\tilde{\bar{C}}_n^{\ell,\mathbf{c_1}} + \tilde{\bar{C}}_n^{\ell,\mathbf{c_2}}}{2} - \bar{C}_n^{\ell,\mathbf{f}} + \bar{C}_n^{\ell,\mathbf{c}} \right\|_p \\
&\leq \left\| E_{P_\ell}\left[\bar{v}_n^{\ell,\mathbf{f}}(\bar{v}_n^{\ell,\mathbf{f}})^{\mathbf{T}} - \bar{v}_n^{\ell,\mathbf{c}}(\bar{v}_n^{\ell,\mathbf{c}})^{\mathbf{T}} - \mathbb{E}\left[\bar{v}_n^{\ell,\mathbf{f}}(\bar{v}_n^{\ell,\mathbf{f}})^{\mathbf{T}} - \bar{v}_n^{\ell,\mathbf{c}}(\bar{v}_n^{\ell,\mathbf{c}})^{\mathbf{T}} \right] \right] \right\|_p \\
&\quad + \left\| E_{P_\ell}[\bar{v}_n^{\ell,\mathbf{f}}] E_{P_\ell}[(\bar{v}_n^{\ell,\mathbf{f}})^{\mathbf{T}}] - \mathbb{E}\left[\bar{v}_n^{\ell,\mathbf{f}} \right] \mathbb{E}\left[(\bar{v}_n^{\ell,\mathbf{f}})^{\mathbf{T}} \right] \right. \\
&\qquad \left. - E_{P_\ell}[\bar{v}_n^{\ell,\mathbf{c}}] E_{P_\ell}[(\bar{v}_n^{\ell,\mathbf{c}})^{\mathbf{T}}] + \mathbb{E}\left[\bar{v}_n^{\ell,\mathbf{c}} \right] \mathbb{E}\left[(\bar{v}_n^{\ell,\mathbf{c}})^{\mathbf{T}} \right] \right\|_p \\
&\quad + \frac{1}{4} \left\| E_{P_{\ell-1}}[\bar{v}_n^{\ell,\mathbf{c_1}} - \bar{v}_n^{\ell,\mathbf{c_2}}] E_{P_{\ell-1}}[(\bar{v}_n^{\ell,\mathbf{c_1}} - \bar{v}_n^{\ell,\mathbf{c_2}})^{\mathbf{T}}] \right\|_p \\
&=: \mathcal{I}_{21} + \mathcal{I}_{22} + \mathcal{I}_{23}.
\end{aligned}$$

Hölder's inequality, Lemma 6 and the M-Z inequality imply that

$$\mathcal{I}_{21} \lesssim P_\ell^{-1/2} N_\ell^{-\beta/2}, \quad \mathcal{I}_{22} \lesssim P_\ell^{-1/2} N_\ell^{-\beta/2} + P_\ell^{-1}, \quad \text{and} \quad \mathcal{I}_{23} \lesssim P_\ell^{-1}$$

(where $\mathbb{E}\left[\bar{v}_n^{\ell,\mathbf{c_1}} - \bar{v}_n^{\ell,\mathbf{c_2}} \right] = 0$ was used in the last inequality). $\qquad \square$

Lemma 9. *For any $n \geq 0$, $p \geq 2$, denoting Jacobian of φ by $D\varphi$, it holds that*

$$\begin{aligned}
&\left\| E_{P_\ell}\left[D\varphi(\hat{\bar{v}}_n^{\ell,\mathbf{c}})(\hat{v}_n^{\ell,\mathbf{f}} - \hat{v}_n^{\ell,\mathbf{c}} - \hat{\bar{v}}_n^{\ell,\mathbf{f}} + \hat{\bar{v}}_n^{\ell,\mathbf{c}}) \right] \right\|_p \\
&\lesssim \left\| E_{P_\ell}\left[D\varphi(\hat{\bar{v}}_n^{\ell,\mathbf{c}})(I - K_n^{\ell,\mathbf{f}} H)(v_n^{\ell,\mathbf{f}} - v_n^{\ell,\mathbf{c}} - \bar{v}_n^{\ell,\mathbf{f}} + \bar{v}_n^{\ell,\mathbf{c}}) \right] \right\|_p \\
&\quad + \left\| E_{P_\ell}[(v_n^{\ell,\mathbf{f}} - v_n^{\ell,\mathbf{c}} - \bar{v}_n^{\ell,\mathbf{f}} + \bar{v}_n^{\ell,\mathbf{c}})(\bar{v}_n^{\ell,\mathbf{f}})^{\mathbf{T}}] \right\|_{2p} + \left\| E_{P_\ell}[v_n^{\ell,\mathbf{f}} - v_n^{\ell,\mathbf{c}} - \bar{v}_n^{\ell,\mathbf{f}} + \bar{v}_n^{\ell,\mathbf{c}}] \right\|_{2p} \\
&\quad + P_\ell^{-1/2} N_\ell^{-\beta/2} + P_\ell^{-1}.
\end{aligned}$$

Proof. Let $\tilde{y}_n^{\ell,j}$ denote the perturbed observation associated with $(v_n^{\ell,\mathbf{f}_j}, v_n^{\ell,\mathbf{c}_j})$. By the update equations (12) and (32), and $(\bar{v}_n^{\ell,\mathbf{f}_j}, \bar{v}_n^{\ell,\mathbf{c}_j})$, we obtain the representation

$$
\begin{aligned}
\hat{v}_n^{\ell,\mathbf{f}_j} - \hat{v}_n^{\ell,\mathbf{c}_j} - \hat{\bar{v}}_n^{\ell,\mathbf{f}_j} + \hat{\bar{v}}_n^{\ell,\mathbf{c}_j} = {} & \left(I - \bar{K}_n^{\ell,\mathbf{f}} H\right)(v_n^{\ell,\mathbf{f}_j} - v_n^{\ell,\mathbf{c}_j} - \bar{v}_n^{\ell,\mathbf{f}_j} + \bar{v}_n^{\ell,\mathbf{c}_j}) \\
& + (\bar{K}_n^{\ell,\mathbf{f}} - K_n^{\ell,\mathbf{f}}) H(\bar{v}_n^{\ell,\mathbf{f}_j} - \bar{v}_n^{\ell,\mathbf{c}_j}) - (\bar{K}_n^{\ell,\mathbf{f}} - \bar{K}_n^{\ell,\mathbf{c}}) H(v_n^{\ell,\mathbf{c}_j} - \bar{v}_n^{\ell,\mathbf{c}_j}) \\
& - (K_n^{\ell,\mathbf{f}} - K_n^{\ell,\mathbf{c}_j} - \bar{K}_n^{\ell,\mathbf{f}} + \bar{K}_n^{\ell,\mathbf{c}}) H(v_n^{\ell,\mathbf{c}_j} - \bar{v}_n^{\ell,\mathbf{c}_j}) \\
& - (K_n^{\ell,\mathbf{f}} - K_n^{\ell,\mathbf{c}_j} - \bar{K}_n^{\ell,\mathbf{f}} + \bar{K}_n^{\ell,\mathbf{c}})(H\bar{v}_n^{\ell,\mathbf{c}_j} - \tilde{y}_n^{\ell,j}),
\end{aligned}
\tag{39}
$$

By (39) and Hölder's inequality

$$
\begin{aligned}
& \left\| E_{P_\ell} \left[D\varphi(\hat{\bar{v}}_n^{\ell,\mathbf{c}})(\hat{v}_n^{\ell,\mathbf{f}} - \hat{v}_n^{\ell,\mathbf{c}} - \hat{\bar{v}}_n^{\ell,\mathbf{f}} + \hat{\bar{v}}_n^{\ell,\mathbf{c}}) \right] \right\|_p \\
&= \left\| E_{P_{\ell-1}} \left[\sum_{j=1}^2 D\varphi(\hat{\bar{v}}_n^{\ell,\mathbf{c}_j})(\hat{v}_n^{\ell,\mathbf{f}_j} - \hat{v}_n^{\ell,\mathbf{c}_j} - \hat{\bar{v}}_n^{\ell,\mathbf{f}_j} + \hat{\bar{v}}_n^{\ell,\mathbf{c}_j}) \right] \right\|_p \\
&\lesssim \left\| E_{P_{\ell-1}} \left[\sum_{j=1}^2 D\varphi(\hat{\bar{v}}_n^{\ell,\mathbf{c}_j})(I - K_n^{\ell,\mathbf{f}} H)(v_n^{\ell,\mathbf{f}_j} - v_n^{\ell,\mathbf{c}_j} - \bar{v}_n^{\ell,\mathbf{f}_j} + \bar{v}_n^{\ell,\mathbf{c}_j}) \right] \right\|_p \\
&\quad + \left\| D\varphi\left(\hat{\bar{v}}_n^{\ell,\mathbf{c}_j}\right) \right\|_{2p} \left\{ \left\| \bar{K}_n^{\ell,\mathbf{f}} - K_n^{\ell,\mathbf{f}} \right\|_{4p} \left\| \bar{v}_n^{\ell,\mathbf{f}_j} - \bar{v}_n^{\ell,\mathbf{c}_j} \right\|_{4p} \right. \\
&\quad + \left\| \bar{K}_n^{\ell,\mathbf{f}} - \bar{K}_n^{\ell,\mathbf{c}} \right\|_{4p} \left\| v_n^{\ell,\mathbf{c}_j} - \bar{v}_n^{\ell,\mathbf{c}_j} \right\|_{4p} \\
&\quad \left. + \left(\left\| K_n^{\ell,\mathbf{f}} - \bar{K}_n^{\ell,\mathbf{f}} \right\|_{4p} + \left\| K_n^{\ell,\mathbf{c}_j} - \bar{K}_n^{\ell,\mathbf{c}} \right\|_{4p} \right) \left\| v_n^{\ell,\mathbf{c}_j} - \bar{v}_n^{\ell,\mathbf{c}_j} \right\|_{4p} \right\} \\
&\quad + \left\| E_{P_{\ell-1}} \left[\sum_{j=1}^2 D\varphi\left(\hat{\bar{v}}_n^{\ell,\mathbf{c}_j}\right) \left(K_n^{\ell,\mathbf{f}} - K_n^{\ell,\mathbf{c}_j} - \bar{K}_n^{\ell,\mathbf{f}} + \bar{K}_n^{\ell,\mathbf{c}}\right) \left(H\bar{v}_n^{\ell,\mathbf{c}_j} - \tilde{y}_n^{\ell,j}\right) \right] \right\|_p \\
&=: \mathcal{J}_{1,\ell} + \mathcal{J}_{2,\ell} + \mathcal{J}_{3,\ell}.
\end{aligned}
\tag{40}
$$

The properties $\varphi \in \mathbb{F} \subset C_P^2(\mathbb{R}^d, \mathbb{R})$ and $\hat{\bar{v}}_n^{\ell,\mathbf{c}} \in \cap_{r \geq 2} L^r(\Omega, \mathbb{R}^d)$ imply $\|D\varphi(\hat{\bar{v}}_n^{\ell,\mathbf{c}_j})\|_{2p} < \infty$, and Lemmas 4, 5 and 6, and Corollaries 3 and 4 yield that

$$
\mathcal{J}_{2,\ell} \lesssim P_\ell^{-1/2} N_\ell^{-\beta/2} + P_\ell^{-1}.
$$

For the last term, we use $K_n^{\ell,\mathbf{c}_1} = (K_n^{\ell,\mathbf{c}_1} + K_n^{\ell,\mathbf{c}_2})/2 + (K_n^{\ell,\mathbf{c}_1} - K_n^{\ell,\mathbf{c}_2})/2$, the Frobenius scalar product $\langle \cdot, \cdot \rangle_F$ on $\mathbb{R}^{d \times d_O} \times \mathbb{R}^{d \times d_O}$ and the shorthand $\mathcal{D}_n^\ell := D\varphi\left(\hat{\bar{v}}_n^{\ell,\mathbf{c}_1}\right)^{\mathbf{T}} \left(H\bar{v}_n^{\ell,\mathbf{c}_1} - \tilde{y}_n^{\ell,1}\right)^{\mathbf{T}} - D\varphi\left(\hat{\bar{v}}_n^{\ell,\mathbf{c}_2}\right)^{\mathbf{T}} \left(H\bar{v}_n^{\ell,\mathbf{c}_2} - \tilde{y}_n^{\ell,2}\right)^{\mathbf{T}}$ to obtain

$$
\begin{aligned}
\mathcal{J}_{3,\ell} \leq {} & \left\| D\varphi\left(\hat{\bar{v}}_n^{\ell,\mathbf{c}_j}\right)\left(K_n^{\ell,\mathbf{f}} - \frac{K_n^{\ell,\mathbf{c}_1} + K_n^{\ell,\mathbf{c}_2}}{2} - \bar{K}_n^{\ell,\mathbf{f}} + \bar{K}_n^{\ell,\mathbf{c}}\right)\left(H\bar{v}_n^{\ell,\mathbf{c}_j} - \tilde{y}_n^{\ell,j}\right) \right\|_p \\
& + \left\| \left\langle \frac{K_n^{\ell,\mathbf{c}_2} - K_n^{\ell,\mathbf{c}_1}}{2}, E_{P_{\ell-1}}\left[\mathcal{D}_n^\ell\right] \right\rangle_{\mathbf{F}} \right\|_p \\
\lesssim {} & \left\| K_n^{\ell,\mathbf{f}} - \frac{K_n^{\ell,\mathbf{c}_1} + K_n^{\ell,\mathbf{c}_2}}{2} - \bar{K}_n^{\ell,\mathbf{f}} + \bar{K}_n^{\ell,\mathbf{c}} \right\|_{2p} + \left\| \frac{K_n^{\ell,\mathbf{c}_2} - K_n^{\ell,\mathbf{c}_1}}{2} \right\|_{2p} P_\ell^{-1/2} \\
\lesssim {} & \left\| E_{P_\ell}\left[(v_n^{\ell,\mathbf{f}} - v_n^{\ell,\mathbf{c}} - \bar{v}_n^{\ell,\mathbf{f}} + \bar{v}_n^{\ell,\mathbf{c}})\left(\bar{v}_n^{\ell,\mathbf{f}}\right)^{\mathbf{T}}\right] \right\|_{2p} \\
& + \left\| E_{P_\ell}[v_n^{\ell,\mathbf{f}} - v_n^{\ell,\mathbf{c}} - \bar{v}_n^{\ell,\mathbf{f}} + \bar{v}_n^{\ell,\mathbf{c}}] \right\|_{2p} + P_\ell^{-1/2} N_\ell^{-\beta/2} + P_\ell^{-1}.
\end{aligned}
$$

Here, the second last inequality follows from the M-Z inequality applied to the right argument in the scalar product (as it is a sample average of $P_{\ell-1}$ iid, mean-zero random matrices). The last inequality follows by Lemmas 7 and 8 and (37). $\qquad\square$

Lemma 11 shows that Theorem 2 is achieved through bounding $\|E_{P_\ell}[\varphi(\hat{v}_n^{\ell,\mathbf{f}}) - \varphi(\hat{v}_n^{\ell,\mathbf{c}}) - \varphi(\hat{\bar{v}}_n^{\ell,\mathbf{f}}) + \varphi(\hat{\bar{v}}_n^{\ell,\mathbf{c}})]\|_p$ from above. To obtain this bound, we introduce the following sequence of random $d \times d$ matrices: for $\ell \geq 0$, $r, s \in \{1, \ldots, n\}$ and $j \in \{1, 2\}$,

$$A_{r,s}^\ell = \begin{cases} \prod_{i=r}^s D\Psi_i^{N_\ell}(\hat{\bar{v}}_i^{\ell,\mathbf{c}})(I - K_i^{\ell,\mathbf{f}}H) & \text{if } r \leq s \\ I & \text{if } r > s. \end{cases}$$

Moreover, for $j = 1, 2$,

$$A_{r,s}^{\ell,j} = \begin{cases} \prod_{i=r}^s D\Psi_i^{N_\ell}(\hat{\bar{v}}_i^{\ell,\mathbf{c}_j})(I - K_i^{\ell,\mathbf{f}}H) & \text{if } r \leq s \\ I & \text{if } r > s. \end{cases}$$

We note that $|A_{r,s}^{\ell,j}| \in \cap_{q \geq 2} L^q(\Omega)$ for all index values.

Corollary 5. *For any $n \geq 2$, $k \leq n-1$ and $p \geq 2$ it holds that*

$$\big\| E_{P_\ell}\big[D\varphi(\hat{\bar{v}}_n^{\ell,\mathbf{c}})(I - K_n^{\ell,\mathbf{f}}H)A_{k+1,n-1}^\ell D\Psi_k^{N_\ell}(\hat{\bar{v}}_k^{\ell,\mathbf{c}})(\hat{v}_k^{\ell,\mathbf{f}} - \hat{\bar{v}}_k^{\ell,\mathbf{f}} - \hat{v}_k^{\ell,\mathbf{c}} + \hat{\bar{v}}_k^{\ell,\mathbf{c}})\big] \big\|_p$$

$$\lesssim \big\| E_{P_\ell}\big[D\varphi(\hat{\bar{v}}_n^{\ell,\mathbf{c}})(I - K_n^{\ell,\mathbf{f}}H)A_{k,n-1}^\ell(v_k^{\ell,\mathbf{f}} - v_k^{\ell,\mathbf{c}} - \bar{v}_k^{\ell,\mathbf{f}} + \bar{v}_k^{\ell,\mathbf{c}})\big] \big\|_p$$

$$+ \big\| E_{P_\ell}\big[(v_k^{\ell,\mathbf{f}} - v_k^{\ell,\mathbf{c}} - \bar{v}_k^{\ell,\mathbf{f}} + \bar{v}_k^{\ell,\mathbf{c}})(\bar{v}_k^{\ell,\mathbf{f}})^{\mathbf{T}} \big] \big\|_{2p} + \big\| E_{P_\ell}[v_k^{\ell,\mathbf{f}} - v_k^{\ell,\mathbf{c}} - \bar{v}_k^{\ell,\mathbf{f}} + \bar{v}_k^{\ell,\mathbf{c}}] \big\|_{2p}$$

$$+ P_\ell^{-1/2} N_\ell^{-\beta/2} + P_\ell^{-1}.$$

Sketch of proof. Proceeding as in the proof of Lemma 9, we obtain three terms that respectively are similar to $\mathcal{J}_{1,\ell}, \mathcal{J}_{2,\ell}$ and $\mathcal{J}_{3,\ell}$ in (40), but now with the prefactor $D\varphi(\hat{\bar{v}}_n^{\ell,\mathbf{c}_j})$ replaced by $D\varphi(\hat{\bar{v}}_n^{\ell,\mathbf{c}_j})(I - K_n^{\ell,\mathbf{f}}H)A_{k+1,n-1}^{\ell,j}D\Psi_k^{N_\ell}(\hat{\bar{v}}_k^{\ell,\mathbf{c}_j})$. The proof is obtained through the equality

$$A_{k+1,n-1}^{\ell,j}D\Psi_k^{N_\ell}(\hat{\bar{v}}_k^{\ell,\mathbf{c}_j})(I - K_k^{\ell,\mathbf{f}}H) = A_{k,n-1}^{\ell,j},$$

the boundedness of $A_{k,n-1}^{\ell,j}$ and $D\varphi(\hat{\bar{v}}_n^{\ell,\mathbf{c}_j})(I - K_n^{\ell,\mathbf{f}}H)$, and by bounding the terms corresponding to $\mathcal{J}_{2,\ell}$ and $\mathcal{J}_{3,\ell}$ in this corollary similarly as in said lemma. $\square$

Corollary 6. *For any $n \geq 1$, $k \leq s \leq n-1$ and $p \geq 2$ it holds that*

$$\big\| E_{P_\ell}\big[A_{k+1,s}^\ell D\Psi_k^{N_\ell}(\hat{\bar{v}}_k^{\ell,\mathbf{c}})(\hat{v}_k^{\ell,\mathbf{f}} - \hat{\bar{v}}_k^{\ell,\mathbf{f}} - \hat{v}_k^{\ell,\mathbf{c}} + \hat{\bar{v}}_k^{\ell,\mathbf{c}})\big] \big\|_p$$

$$\lesssim \big\| E_{P_\ell}\big[A_{k,s}^\ell(v_k^{\ell,\mathbf{f}} - v_k^{\ell,\mathbf{c}} - \bar{v}_k^{\ell,\mathbf{f}} + \bar{v}_k^{\ell,\mathbf{c}})\big] \big\|_p \qquad\qquad (41)$$

$$+ \big\| E_{P_\ell}\big[(v_k^{\ell,\mathbf{f}} - v_k^{\ell,\mathbf{c}} - \bar{v}_k^{\ell,\mathbf{f}} + \bar{v}_k^{\ell,\mathbf{c}})(\bar{v}_k^{\ell,\mathbf{f}})^{\mathbf{T}} \big] \big\|_{2p}$$

$$+ \big\| E_{P_\ell}[v_k^{\ell,\mathbf{f}} - v_k^{\ell,\mathbf{c}} - \bar{v}_k^{\ell,\mathbf{f}} + \bar{v}_k^{\ell,\mathbf{c}}] \big\|_{2p} + P_\ell^{-1/2} N_\ell^{-\beta/2} + P_\ell^{-1}.$$

Sketch of proof. The statement of this corollary is similar to Corollary 5, but with the $1 \times d$ prefactor $D\varphi(\hat{\bar{v}}_n^{\ell,\mathbf{c}_j})(I - K_n^{\ell,\mathbf{f}}H)A_{k+1,n-1}^{\ell,j}$ replaced by $A_{k+1,s}^{\ell,j}D\Psi_k^{N_\ell}(\hat{\bar{v}}_k^{\ell,\mathbf{c}_j})$. As the latter factor also is bounded (it is an element of $\cap_{\tilde{r} \geq 2} L^r(\Omega, \mathbb{R}^{d \times d})$), the proof follows by a similar argument as in Corollary 5. However, since the prefactor in this case is a $d \times d$ matrix rather than a vector, the term in (41) that corresponds to $\mathcal{J}_{3,\ell}$ in the proof of Lemma 9 with the shorthand $\mathcal{G}_k^j := K_k^{\ell,\mathbf{f}} - K_k^{\ell,\mathbf{c}_j} - \bar{K}_k^{\ell,\mathbf{f}} + \bar{K}_k^{\ell,\mathbf{c}}$ becomes

$$\Big\| E_{P_{\ell-1}}\Big[\sum_{j=1}^2 A_{k+1,s}^{\ell,j} D\Psi_k^{N_\ell}(\hat{\bar{v}}_k^{\ell,\mathbf{c}_j})\mathcal{G}_k^j(H\bar{v}_k^{\ell,\mathbf{c}_j} - \tilde{y}_k^{\ell,j}) \Big] \Big\|_p$$

$$\leq \sum_{i=1}^d \Big\| E_{P_{\ell-1}}\Big[\sum_{j=1}^2 \big(A_{k+1,s}^{\ell,j} D\Psi_k^{N_\ell}(\hat{\bar{v}}_k^{\ell,\mathbf{c}_j})\big)_i \mathcal{G}_k^j(H\bar{v}_k^{\ell,\mathbf{c}_j} - \tilde{y}_k^{\ell,j}) \Big] \Big\|_p.$$

Here $\big(A_{k+1,s}^{\ell,j} D\Psi_k^{N_\ell}(\hat{\bar{v}}_k^{\ell,\mathbf{c}_j})\big)_i$ denotes the i-th row vector of the matrix. Hence, we have d terms to bound that are similar to the $\mathcal{J}_{3,\ell}$-term in Corollary 5. $\qquad\square$

Corollary 7. *For any $n \geq 1$, $k \leq s \leq n-1$, $r \leq n$ and $p \geq 2$ it holds that*

$$
\big\| E_{P_\ell}\big[A_{k+1,s}^\ell D\Psi_k^{N_\ell}(\hat{\bar{v}}_k^{\ell,\mathbf{c}})(\hat{v}_k^{\ell,\mathbf{f}} - \hat{\bar{v}}_k^{\ell,\mathbf{f}} - \hat{v}_k^{\ell,\mathbf{c}} + \hat{\bar{v}}_k^{\ell,\mathbf{c}})(\bar{v}_r^{\ell,\mathbf{f}})^{\mathbf{T}}\big]\big\|_p
$$
$$
\lesssim \big\| E_{P_\ell}\big[A_{k,s}^\ell(v_k^{\ell,\mathbf{f}} - v_k^{\ell,\mathbf{c}} - \bar{v}_k^{\ell,\mathbf{f}} + \bar{v}_k^{\ell,\mathbf{c}})(\bar{v}_r^{\ell,\mathbf{f}})^{\mathbf{T}}\big]\big\|_p + \big\| E_{P_\ell}[v_k^{\ell,\mathbf{f}} - v_k^{\ell,\mathbf{c}} - \bar{v}_k^{\ell,\mathbf{f}} + \bar{v}_k^{\ell,\mathbf{c}}]\big\|_{2p}
$$
$$
+ \big\| E_{P_\ell}\big[(v_k^{\ell,\mathbf{f}} - v_k^{\ell,\mathbf{c}} - \bar{v}_k^{\ell,\mathbf{f}} + \bar{v}_k^{\ell,\mathbf{c}})(\bar{v}_k^{\ell,\mathbf{f}})^{\mathbf{T}}\big]\big\|_{2p} + P_\ell^{-1/2} N_\ell^{-\beta/2} + P_\ell^{-1}.
\tag{42}
$$

Sketch of proof. The statement of this corollary is similar to Corollary 5, but here with the additional $1 \times d$ postfactor $(\bar{v}_r^{\ell,\mathbf{f}})^{\mathbf{T}}$. The only technicality this introduces is in bounding the term in (42) that corresponds to $\mathcal{J}_{3,\ell}$ in the proof of Lemma 9. That is,

$$
\Big\| E_{P_{\ell-1}}\Big[\sum_{j=1}^2 A_{k+1,s}^{\ell,j} D\Psi_k^{N_\ell}(\hat{\bar{v}}_k^{\ell,\mathbf{c}_j})\mathcal{G}_k^j(H\bar{v}_k^{\ell,\mathbf{c}_j} - \tilde{y}_k^{\ell,j})(\bar{v}_r^{\ell,\mathbf{f}})^{\mathbf{T}}\Big]\Big\|_p
$$
$$
\leq \sum_{i=1}^d \Big\| E_{P_{\ell-1}}\Big[\sum_{j=1}^2 A_{k+1,s}^{\ell,j} D\Psi_k^{N_\ell}(\hat{\bar{v}}_k^{\ell,\mathbf{c}_j})\mathcal{G}_k^j(H\bar{v}_k^{\ell,\mathbf{c}_j} - \tilde{y}_k^{\ell,j})(\bar{v}_r^{\ell,\mathbf{f}_j})_i^{\mathbf{T}}\Big]\Big\|_p.
$$

Here $(\bar{v}_r^{\ell,\mathbf{f}_j})_i^{\mathbf{T}}$ denotes the i-th component of the row vector. We thus have d terms to bound which are similar to the $\mathcal{J}_{3,\ell}$-term in Corollary 5. $\qquad\square$

Lemma 10. *For any $n \geq 1$, $k \leq s \leq n-1$, $r \leq n$ and $p \geq 2$ it holds that*

$$
\big\| E_{P_\ell}\big[D\varphi(\hat{\bar{v}}_n^{\ell,\mathbf{c}})(I - K_n^{\ell,\mathbf{f}}H)A_{k+1,n-1}^\ell(v_{k+1}^{\ell,\mathbf{f}_j} - v_{k+1}^{\ell,\mathbf{c}} - \bar{v}_{k+1}^{\ell,\mathbf{f}} + \bar{v}_{k+1}^{\ell,\mathbf{c}})\big]\big\|_p
$$
$$
\lesssim \big\| E_{P_\ell}\big[D\varphi(\hat{\bar{v}}_n^{\ell,\mathbf{c}})(I - K_n^{\ell,\mathbf{f}}H)A_{k,n-1}^\ell(v_k^{\ell,\mathbf{f}} - v_k^{\ell,\mathbf{c}} - \bar{v}_k^{\ell,\mathbf{f}} + \bar{v}_k^{\ell,\mathbf{c}})\big]\big\|_{2p}
$$
$$
+ \big\| E_{P_\ell}\big[(v_k^{\ell,\mathbf{f}} - v_k^{\ell,\mathbf{c}} - \bar{v}_k^{\ell,\mathbf{f}} + \bar{v}_k^{\ell,\mathbf{c}})(\bar{v}_k^{\ell,\mathbf{f}})^{\mathbf{T}}\big]\big\|_{2p} + \big\| E_{P_\ell}[v_k^{\ell,\mathbf{f}} - v_k^{\ell,\mathbf{c}} - \bar{v}_k^{\ell,\mathbf{f}} + \bar{v}_k^{\ell,\mathbf{c}}]\big\|_{2p}
$$
$$
+ N_\ell^{-\beta/2} P_\ell^{-1/2} + P_\ell^{-1},
\tag{43}
$$

$$
\big\| E_{P_\ell}\big[A_{k+1,s}^\ell(v_{k+1}^{\ell,\mathbf{f}} - v_{k+1}^{\ell,\mathbf{c}} - \bar{v}_{k+1}^{\ell,\mathbf{f}} + \bar{v}_{k+1}^{\ell,\mathbf{c}})\big]\big\|_p
$$
$$
\lesssim \big\| E_{P_\ell}\big[A_{k,s}^\ell(v_k^{\ell,\mathbf{f}} - v_k^{\ell,\mathbf{c}} - \bar{v}_k^{\ell,\mathbf{f}} + \bar{v}_k^{\ell,\mathbf{c}})\big]\big\|_p + \big\| E_{P_\ell}[v_k^{\ell,\mathbf{f}} - v_k^{\ell,\mathbf{c}} - \bar{v}_k^{\ell,\mathbf{f}} + \bar{v}_k^{\ell,\mathbf{c}}]\big\|_{2p}
\tag{44}
$$
$$
+ \big\| E_{P_\ell}\big[(v_k^{\ell,\mathbf{f}} - v_k^{\ell,\mathbf{c}} - \bar{v}_k^{\ell,\mathbf{f}} + \bar{v}_k^{\ell,\mathbf{c}})(\bar{v}_k^{\ell,\mathbf{f}})^{\mathbf{T}}\big]\big\|_{2p} + P_\ell^{-1/2} N_\ell^{-\beta/2} + P_\ell^{-1},
$$

and

$$
\big\| E_{P_\ell}\big[A_{k+1,s}^\ell(v_{k+1}^{\ell,\mathbf{f}} - v_{k+1}^{\ell,\mathbf{c}} - \bar{v}_{k+1}^{\ell,\mathbf{f}} + \bar{v}_{k+1}^{\ell,\mathbf{c}})(\bar{v}_r^{\ell,\mathbf{f}})^{\mathbf{T}}\big]\big\|_p
$$
$$
\lesssim \big\| E_{P_\ell}\big[A_{k,s}^\ell(v_k^{\ell,\mathbf{f}} - v_k^{\ell,\mathbf{c}} - \bar{v}_k^{\ell,\mathbf{f}} + \bar{v}_k^{\ell,\mathbf{c}})(\bar{v}_r^{\ell,\mathbf{f}})^{\mathbf{T}}\big]\big\|_p + \big\| E_{P_\ell}[v_k^{\ell,\mathbf{f}} - v_k^{\ell,\mathbf{c}} - \bar{v}_k^{\ell,\mathbf{f}} + \bar{v}_k^{\ell,\mathbf{c}}]\big\|_{2p}
$$
$$
+ \big\| E_{P_\ell}\big[(v_k^{\ell,\mathbf{f}} - v_k^{\ell,\mathbf{c}} - \bar{v}_k^{\ell,\mathbf{f}} + \bar{v}_k^{\ell,\mathbf{c}})(\bar{v}_k^{\ell,\mathbf{f}})^{\mathbf{T}}\big]\big\|_{2p} + P_\ell^{-1/2} N_\ell^{-\beta/2} + P_\ell^{-1}.
\tag{45}
$$

Proof. The mean-value theorem yields the expansion

$$
\begin{aligned}
v_{k+1}^{\ell,\mathbf{f}} - v_{k+1}^{\ell,\mathbf{c}} - \bar{v}_{k+1}^{\ell,\mathbf{f}} + \bar{v}_{k+1}^{\ell,\mathbf{c}} &= \Psi_k^{N_\ell}(\hat{v}_k^{\ell,\mathbf{f}}) - \Psi_k^{N_{\ell-1}}(\hat{v}_k^{\ell,\mathbf{c}}) - \Psi_k^{N_\ell}(\hat{\bar{v}}_k^{\ell,\mathbf{f}}) + \Psi_k^{N_{\ell-1}}(\hat{\bar{v}}_k^{\ell,\mathbf{c}}) \\
&= D\Psi_k^{N_\ell}(\hat{\bar{v}}_k^{\ell,\mathbf{c}})(\hat{v}_k^{\ell,\mathbf{f}} - \hat{\bar{v}}_k^{\ell,\mathbf{f}} - \hat{v}_k^{\ell,\mathbf{c}} + \hat{\bar{v}}_k^{\ell,\mathbf{c}}) + (\hat{v}_k^{\ell,\mathbf{f}} - \hat{\bar{v}}_k^{\ell,\mathbf{f}})^{\mathbf{T}} D^2\Psi_k^{N_\ell}(\theta_k)(\hat{\bar{v}}_k^{\ell,\mathbf{f}} - \hat{\bar{v}}_k^{\ell,\mathbf{c}}) \\
&\quad + \left(D\Psi_k^{N_\ell}(\hat{\bar{v}}_k^{\ell,\mathbf{c}}) - D\Psi_k^{N_{\ell-1}}(\hat{\bar{v}}_k^{\ell,\mathbf{c}}) \right)(\hat{v}_k^{\ell,\mathbf{f}} - \hat{\bar{v}}_k^{\ell,\mathbf{f}}) \\
&\quad + \frac{1}{2}(\hat{v}_k^{\ell,\mathbf{f}} - \hat{\bar{v}}_k^{\ell,\mathbf{f}})^{\mathbf{T}} D^2\Psi_k^{N_\ell}(\tilde{\theta}_k)(\hat{v}_k^{\ell,\mathbf{f}} - \hat{\bar{v}}_k^{\ell,\mathbf{f}}) - \frac{1}{2}(\hat{v}_k^{\ell,\mathbf{c}} - \hat{\bar{v}}_k^{\ell,\mathbf{c}})^{\mathbf{T}} D^2\Psi_k^{N_\ell}(\check{\theta}_k)(\hat{v}_k^{\ell,\mathbf{c}} - \hat{\bar{v}}_k^{\ell,\mathbf{c}}),
\end{aligned}
\tag{46}
$$

where for $x, \check{x} \in \mathbb{R}^d$, $\theta_k \in \mathrm{Conv}(\hat{\bar{v}}_k^{\ell,\mathbf{f}}, \hat{\bar{v}}_k^{\ell,\mathbf{c}})$, $\tilde{\theta}_k \in \mathrm{Conv}(\hat{v}_k^{\ell,\mathbf{f}}, \hat{\bar{v}}_k^{\ell,\mathbf{f}})$, $\check{\theta}_k \in \mathrm{Conv}(\hat{v}_k^{\ell,\mathbf{c}}, \hat{\bar{v}}_k^{\ell,\mathbf{c}})$ and $\mathrm{Conv}(x, \check{x}) := \{xt + (1-t)\check{x} \mid t \in [0,1]\}$. The expansion, Assumption 2 and Corollary 5 yield (43). And, similarly, Corollaries 6 and 7 respectively yield (44) and (45). $\qquad\square$

Lemma 11. *For any $n \geq 0$ and $p \geq 2$, it holds that*

$$
\left\| E_{P_\ell}\left[\varphi(\hat{v}_n^{\ell,\mathbf{f}}) - \varphi(\hat{v}_n^{\ell,\mathbf{c}}) - \varphi(\hat{\bar{v}}_n^{\ell,\mathbf{f}}) + \varphi(\hat{\bar{v}}_n^{\ell,\mathbf{c}}) \right] \right\|_p \lesssim N_\ell^{-\beta/2} P_\ell^{-1/2} + P_\ell^{-1}.
$$

Proof. By assumption 1(ii) and a similar application of the mean-value theorem as in (46),

$$
\begin{aligned}
&\left\| E_{P_\ell}\left[\varphi(\hat{v}_n^{\ell,\mathbf{f}}) - \varphi(\hat{v}_n^{\ell,\mathbf{c}}) - \varphi(\hat{\bar{v}}_n^{\ell,\mathbf{f}}) + \varphi(\hat{\bar{v}}_n^{\ell,\mathbf{c}}) \right] \right\|_p \\
&\qquad \lesssim \left\| E_{P_\ell}\left[D\varphi(\hat{\bar{v}}_n^{\ell,\mathbf{c}})(\hat{v}_n^{\ell,\mathbf{f}} - \hat{v}_n^{\ell,\mathbf{c}} - \hat{\bar{v}}_n^{\ell,\mathbf{f}} + \hat{\bar{v}}_n^{\ell,\mathbf{c}}) \right] \right\|_p + N_\ell^{-\beta/2} P_\ell^{-1/2} + P_\ell^{-1}.
\end{aligned}
$$

By Lemma 9, and thereafter using that $A_{n,n-1}^\ell = A_{n-1,n-2}^\ell = I$ and applying Lemma 10, we obtain

$$
\begin{aligned}
\mathcal{K}_\ell &:= \left\| E_{P_\ell}\left[D\varphi(\hat{\bar{v}}_n^{\ell,\mathbf{c}})(\hat{v}_n^{\ell,\mathbf{f}} - \hat{v}_n^{\ell,\mathbf{c}} - \hat{\bar{v}}_n^{\ell,\mathbf{f}} + \hat{\bar{v}}_n^{\ell,\mathbf{c}}) \right] \right\|_p \\
&\lesssim \left\| E_{P_\ell}\left[D\varphi(\hat{\bar{v}}_n^{\ell,\mathbf{c}})(I - K_n^{\ell,\mathbf{f}} H) A_{n,n-1}^\ell (v_n^{\ell,\mathbf{f}} - v_n^{\ell,\mathbf{c}} - \bar{v}_n^{\ell,\mathbf{f}} + \bar{v}_n^{\ell,\mathbf{c}}) \right] \right\|_p \\
&\quad + \left\| E_{P_\ell}[A_{n,n-1}^\ell (v_n^{\ell,\mathbf{f}} - v_n^{\ell,\mathbf{c}} - \bar{v}_n^{\ell,\mathbf{f}} + \bar{v}_n^{\ell,\mathbf{c}})(\bar{v}_n^{\ell,\mathbf{f}})^{\mathbf{T}}] \right\|_{2p} \\
&\quad + \left\| E_{P_\ell}[A_{n,n-1}^\ell (v_n^{\ell,\mathbf{f}} - v_n^{\ell,\mathbf{c}} - \bar{v}_n^{\ell,\mathbf{f}} + \bar{v}_n^{\ell,\mathbf{c}})] \right\|_{2p} + P_\ell^{-1/2} N_\ell^{-\beta/2} + P_\ell^{-1} \\
&\lesssim \left\| E_{P_\ell}\left[D\varphi(\hat{\bar{v}}_n^{\ell,\mathbf{c}})(I - K_n^{\ell,\mathbf{f}} H) A_{n-1,n-1}^\ell (v_{n-1}^{\ell,\mathbf{f}} - v_{n-1}^{\ell,\mathbf{c}} - \bar{v}_{n-1}^{\ell,\mathbf{f}} + \bar{v}_{n-1}^{\ell,\mathbf{c}}) \right] \right\|_p \\
&\quad + \left\| E_{P_\ell}[A_{n-1,n-1}^\ell (v_{n-1}^{\ell,\mathbf{f}} - v_{n-1}^{\ell,\mathbf{c}} - \bar{v}_{n-1}^{\ell,\mathbf{f}} + \bar{v}_{n-1}^{\ell,\mathbf{c}})(\bar{v}_n^{\ell,\mathbf{f}})^{\mathbf{T}}] \right\|_{2p} \\
&\quad + \left\| E_{P_\ell}[A_{n-1,n-1}^\ell (v_{n-1}^{\ell,\mathbf{f}} - v_{n-1}^{\ell,\mathbf{c}} - \bar{v}_{n-1}^{\ell,\mathbf{f}} + \bar{v}_{n-1}^{\ell,\mathbf{c}})] \right\|_{2p} \\
&\quad + \left\| E_{P_\ell}\left[A_{n-1,n-2}^\ell (v_{n-1}^{\ell,\mathbf{f}} - v_{n-1}^{\ell,\mathbf{c}} - \bar{v}_{n-1}^{\ell,\mathbf{f}} + \bar{v}_{n-1}^{\ell,\mathbf{c}})(\bar{v}_{n-1}^{\ell,\mathbf{f}})^{\mathbf{T}} \right] \right\|_{4p} \\
&\quad + \left\| E_{P_\ell}\left[A_{n-1,n-2}^\ell (v_{n-1}^{\ell,\mathbf{f}} - v_{n-1}^{\ell,\mathbf{c}} - \bar{v}_{n-1}^{\ell,\mathbf{f}} + \bar{v}_{n-1}^{\ell,\mathbf{c}}) \right] \right\|_{4p} + P_\ell^{-1/2} N_\ell^{-\beta/2} + P_\ell^{-1}.
\end{aligned}
$$

Recalling that $\hat{v}_0^{\ell,\mathbf{f}} - \hat{v}_0^{\ell,\mathbf{c}} - \hat{\hat{v}}_0^{\ell,\mathbf{f}} + \hat{\hat{v}}_0^{\ell,\mathbf{c}} = 0$ and applying Lemma 10 iteratively $n-1$ times, we obtain that

$$\mathcal{K}_\ell \lesssim \left\| E_{P_\ell}[D\varphi(\hat{\hat{v}}_n^{\ell,\mathbf{c}})(I - K_n^{\ell,\mathbf{f}}H)A_{1,n-1}^\ell D\Psi_0^{N_\ell}(\hat{\hat{v}}_0^{\ell,\mathbf{c}})(\hat{v}_0^{\ell,\mathbf{f}} - \hat{v}_0^{\ell,\mathbf{c}} - \hat{\hat{v}}_0^{\ell,\mathbf{f}} + \hat{\hat{v}}_0^{\ell,\mathbf{c}})] \right\|_p$$

$$+ \sum_{k=1}^{n-1} \left\{ \left\| E_{P_\ell}[A_{1,n-k}^\ell D\Psi_0^{N_\ell}(\hat{\hat{v}}_0^{\ell,\mathbf{c}})(\hat{v}_0^{\ell,\mathbf{f}} - \hat{v}_0^{\ell,\mathbf{c}} - \hat{\hat{v}}_0^{\ell,\mathbf{f}} + \hat{\hat{v}}_0^{\ell,\mathbf{c}})(\bar{v}_n^{\ell,\mathbf{f}})^{\mathbf{T}}] \right\|_{2^k p} \right.$$

$$+ \left\| E_{P_\ell}[A_{1,n-k}^\ell D\Psi_0^{N_\ell}(\hat{\hat{v}}_0^{\ell,\mathbf{c}})(\hat{v}_0^{\ell,\mathbf{f}} - \hat{v}_0^{\ell,\mathbf{c}} - \hat{\hat{v}}_0^{\ell,\mathbf{f}} + \hat{\hat{v}}_0^{\ell,\mathbf{c}})] \right\|_{2^k p} \right\} + N_\ell^{-\beta/2}P_\ell^{-1/2} + P_\ell^{-1}$$

$$\lesssim N_\ell^{-\beta/2}P_\ell^{-1/2} + P_\ell^{-1}.$$

$\square$

B.3. **Proof of corollary 2.** By Theorem 2, we recall that

$$\|\mu_n^{\mathrm{ML}}[\varphi] - \bar{\mu}_n[\varphi]\|_p \lesssim N_L^{-\beta/2}P_L^{-1/2} + P_L^{-1} + N_L^{-\alpha} + \sum_{\ell=0}^{L} M_\ell^{-1/2}(N_\ell^{-\beta/2}P_\ell^{-1/2} + P_\ell^{-1}),$$

with $P_\ell \eqsim 2^\ell$, $N_\ell \eqsim 2^{s\ell}$ for some $s > 0$. Our objective is to prove that the parameter choices for s, L, $\{M_\ell\}$ stated in Corollary 2 ensure that the goal

$$\|\mu_n^{\mathrm{ML}}[\varphi] - \bar{\mu}_n[\varphi]\|_p \lesssim \epsilon$$

is reached at the asymptotic computational cost (17), where

$$\mathrm{Cost(MLEnKF)} := \sum_{\ell=0}^{L} M_\ell N_\ell P_\ell.$$

Fix the value of $s > 0$. It then follows straightforwardly that to control the "bias"

$$N_L^{-\beta/2}P_L^{-1/2} + P_L^{-1} + N_L^{-\alpha} \lesssim \epsilon,$$

one must have $L \eqsim \log_2(\epsilon^{-1})/\min(1, (1+\beta s)/2, \alpha s) + 1$. It remains to minimize $\sum_{\ell=0}^{L} M_\ell N_\ell P_\ell$ subject to the constraint

$$\sum_{\ell=0}^{L} M_\ell^{-1/2}(N_\ell^{-\beta/2}P_\ell^{-1/2} + P_\ell^{-1}) \lesssim \epsilon, \tag{47}$$

and also having in mind that M_ℓ must be a natural number for all $\ell \leq L$.

The method of Lagrange multipliers applied to

$$\mathcal{L}(\{M_\ell\}, \lambda) := \sum_{\ell=0}^{L} M_\ell N_\ell P_\ell + \lambda(\sum_{\ell=0}^{L} M_\ell^{-1/2}D_\ell - \epsilon)$$

with $D_\ell := (N_\ell^{-\beta/2}P_\ell^{-1/2} + P_\ell^{-1})$ yields

$$M_\ell \eqsim \lambda^{2/3}(N_\ell P_\ell)^{-2/3}D_\ell^{2/3} + 1 \quad \text{and} \quad \lambda = \epsilon^{-3}\left(\sum_{\ell=0}^{L}(N_\ell P_\ell)^{1/3}D_\ell^{2/3}\right)^3. \tag{48}$$

Since

$$P_\ell N_\ell \eqsim 2^{(1+s)\ell} \quad \text{and} \quad D_\ell \eqsim 2^{-(\min(\beta s, 1)+1)\ell/2},$$

we have

$$\sum_{\ell=0}^{L}(N_\ell P_\ell)^{1/3}D_\ell^{2/3} \eqsim \sum_{\ell=0}^{L} 2^{(s-\min(\beta s,1))\ell/3} \eqsim \begin{cases} 1 & \text{if } \min(\beta s, 1) > s, \\ L & \text{if } \min(\beta s, 1) = s, \\ 2^{(s-\min(\beta s,1))L/3} & \text{if } \min(\beta s, 1) < s, \end{cases}$$

and the optimal formula for $\{M_\ell\}$ for a fixed $s > 0$, cf. (14), follows from the last equality and (48).

By (47), the choice for $\{M_\ell\}$ leads to

$$\text{Cost(MLEnKF)} \approx \begin{cases} \epsilon^{-2} + \epsilon^{-f(s)} & \text{if } \min(\beta s, 1) > s, \\ \epsilon^{-2}L^3 + \epsilon^{-f(s)} & \text{if } \min(\beta s, 1) = s, \\ \epsilon^{-2 - \frac{s - \min(\beta s, 1)}{\min(1, (\beta s+1)/2, \alpha s)}} + \epsilon^{-f(s)} & \text{if } \min(\beta s, 1) < s, \end{cases}$$

where

$$f(s) := \frac{1+s}{\min(1, (\beta s + 1)/2, \alpha s)} = \begin{cases} \dfrac{1+s}{\alpha s} & \text{if } s \leq \frac{\min((1+\beta s)/2, 1)}{\alpha}, \\ \dfrac{2(1+s)}{\beta s + 1} & \text{if } \frac{(1+\beta s)}{2\alpha} < s < \alpha^{-1}, \\ 1+s & \text{if } \min((1 + \beta s)/2, \alpha s) \geq 1. \end{cases}$$

We next consider the problem of determining the value/inclusion set of s which minimizes the asymptotic growth rate of Cost(MLEnKF). We consider three cases separately:

1. If $\beta < 1$, then $\min(\beta s, 1) < s$ for all $s > 0$, cf. Figure 7(a), and

$$\min(1, (\beta s + 1)/2, \alpha s) \leq \frac{1 + \min(1, \beta s)}{2},$$

implies that

$$2 + \frac{s - \min(\beta s, 1)}{\min(1, (\beta s + 1)/2, \alpha s)} \leq f(s).$$
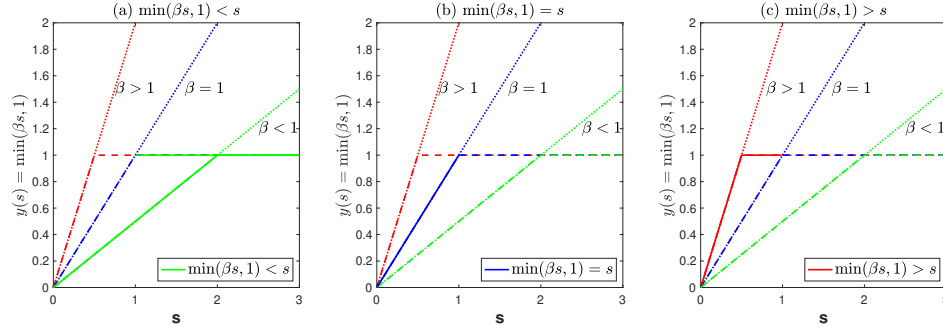
Consequently,

FIGURE 7. (a) The inequality $\min(\beta s, 1) < s$ (green line). (b) The equality $\min(\beta s, 1) = s$ (blue line). (c) The inequality $\min(\beta s, 1) > s$ (red line). The dash lines correspond to the function $y(s) = \min(\beta s, 1)$ and the dotted lines refer to the function $y(s) = \beta s$ varying by different cases of β value.

$$\text{Cost(s)} \approx \epsilon^{-\frac{1+s}{\min(1, (\beta s+1)/2, \alpha s)}}.$$

Observing that f is strictly decreasing on the set $(s < \frac{\min((1+\beta s)/2, 1)}{\alpha})$ and strictly increasing on $(s > \frac{\min((1+\beta s)/2, 1)}{\alpha})$, we obtain the unique minimizer

$$\arg \min_{s > 0} f(s) = \begin{cases} \alpha^{-1} & \text{if } \alpha \leq \beta, \\ (2\alpha - \beta)^{-1} & \text{otherwise.} \end{cases} \tag{49}$$

2. If $\beta = 1$, then $\min(\beta s, 1) = s$ for all $s \in (0, 1]$ and $\min(\beta s, 1) < s$ for $s > 1$, cf. Figure 7(a-b). Consequently,

$$\text{Cost(s)} \approx \begin{cases} \epsilon^{-2} L^3 + \epsilon^{-f(s)} & \text{if } s \in (0, 1], \\ \epsilon^{-f(s)} & \text{if } s > 1. \end{cases}$$

By (49), now simplifying to $f(s) = \max(1 + s, 2, (1 + s)/(\alpha s))$, we obtain that

$$\text{Cost(s)} \approx \begin{cases} \epsilon^{-2} L^3 & \text{if } \alpha \geq 1 \quad \& \quad s \in [\alpha^{-1}, 1] \\ \epsilon^{-(1+\alpha^{-1})} & \text{if } \alpha < 1 \quad \& \quad s = \alpha^{-1}. \end{cases}$$

3. If $\beta > 1$, then $\min(\beta s, 1) > s$ for all $s \in (0, 1)$ and $\min(\beta s, 1) = s$ for $s = 1$, and $\min(\beta s, 1) < s$ for $s > 1$. Thus

$$\text{Cost(s)} \approx \begin{cases} \epsilon^{-2} + \epsilon^{-f(s)} & \text{if } s \in (0, 1), \\ \epsilon^{-2} L^3 + \epsilon^{-f(s)} & \text{if } s = 1, \\ \epsilon^{-f(s)} & \text{if } s > 1. \end{cases}$$

If $\alpha > 1$, then $f(s) \leq 2 \iff s \in [\alpha^{-1}, 1]$. If $\alpha = 1$, then $f(s) \leq 2 \iff s = 1$. And if $\alpha < 1$, then

$$f(s) = \min_{s>0} f(s) = (1 + \alpha^{-1}) \iff s = \alpha^{-1}.$$

This yields

$$\text{Cost(s)} \approx \begin{cases} \epsilon^{-2} & \text{if } \alpha > 1 \quad \& \quad s \in [\alpha^{-1}, 1) \\ \epsilon^{-2} L^3 & \text{if } \alpha = 1 \quad \& \quad s = 1 \\ \epsilon^{-(1+\alpha^{-1})} & \text{if } \alpha < 1 \quad \& \quad s = \alpha^{-1}. \end{cases}$$

Appendix C. **DMFEnKF algorithm.** This section describes the algorithm for the density-based deterministic approximation of the MFEnKF, which iteratively computes the prediction density $\rho_{\bar{v}_n}$ and updated density $\rho_{\hat{v}_n}$ for $n = 1, 2, \ldots$. Each iteration cycle consists of two steps: one transition from $\rho_{\hat{v}_n}$ to $\rho_{\bar{v}_{n+1}}$ governed by the Fokker-Planck equation (FPE) and another from $\rho_{\bar{v}_{n+1}}$ to $\rho_{\hat{v}_{n+1}}$ via an affine transformation and a convolution with a Gaussian density. For simplicity, we show the algorithm for the one-dimensional state-space case, i.e, $d = 1$.

Let $\mathcal{S}^t \rho$ denote a solution at time t of the FPE

$$\partial_t p(x, t) = \partial_x (V'(x) p(x, t)) + \frac{\sigma^2}{2} \partial_x^2 p(x, t), \qquad (x, t) \in \mathbb{R} \times (0, \infty),$$

with the initial condition $p(\cdot, 0) = \rho$. Note that for the mean-field dynamics (22) with $N = \infty$ that satisfies the SDE (20), it holds that $\Psi(\hat{v}_n) \sim \mathcal{S}^1 \rho_{\hat{v}_n}$ for any $n \geq 0$.

Since the updated mean-field ensembles can be viewed as the sum of the following independent random variables

$$\hat{v}_n = \underbrace{(I - \bar{K}_n H)\bar{v}_n + \bar{K}_n y_n}_{X} + \underbrace{\bar{K}_n \tilde{\eta}_n}_{Y},$$

the updated density can be written

$$\rho_{\hat{v}_n}(x) = \rho_{X+Y}(x) = \int_{\mathbb{R}} \rho_X(z) \rho_Y(x - z) dz = \rho_X * \rho_Y(x),$$

where $*$ denotes the convolution operator.

Algorithm 1: DMFEnKF

Input: The initial updated density $\rho_{\hat{v}_0} = \rho_{u_0|Y_0}$, the number of time steps
N_t, the number of spatial steps N_x, the discretization interval
$[x_0, x_1]$, the simulation length $\mathcal{N}$.

Output: The prediction and updated density, $\rho_{\bar{v}_n}$ and $\rho_{\hat{v}_n}$, respectively.

1 $\Delta t = \frac{1}{N_t}$, $\Delta x = \frac{x_1 - x_0}{N_x}$.

2 for $n=1 : \mathcal{N}$ **do**

3   Compute the prediction density $\rho_{\bar{v}_n}(x) = \mathcal{S}^1 \rho_{\hat{v}_{n-1}}$ by Crank-Nicolson
        numerical method with the discretization steps $(\Delta t, \Delta x)$.

4   Compute the prediction covariance $\bar{C}_n = \int x^2 \rho_{\bar{v}_n}(x)dx - (\int x \rho_{\bar{v}_n}(x)dx)^2$
        using a quadrature rule.

5   Compute the Kalman gain $\bar{K}_n = \bar{C}_n H^{\mathbf{T}}(H\bar{C}_n H^{\mathbf{T}} + \Gamma)^{-1}$.

6   Compute the updated density $\rho_{\hat{v}_n} = \rho_X * \rho_Y$ by discrete convolution of
        the two functions represented on the spatial mesh.

Remark 8. The discretization interval $[x_0, x_1]$ must be chosen such that the truncation error pertaining to the integral on the complement of said interval is close to zero, i.e.,

$$\int_{[x_0,x_1]^c} \rho_{\hat{v}_n}(x)dx \approx 0.$$

For the problem in Section 3.2, we found by numerical experiments using the Crank-Nicolson method that $[x_0, x_1] = [-5, 5]$, $\Delta t = 10^{-3}$ and $\Delta x = 10^{-5}$ were suitable resolution parameters to obtain solutions with negligible approximation errors relative to the statistical and bias error introduced by EnKF and MLEnKF.

Appendix D. **Extension from MLEnKF to multi-index EnKF (MIEnKF).**
Following the construction of multi-index Monte Carlo mehods for mean-field dynamics approximations [19], we sketch an extension from MLEnKF to MIEnKF. On the particle level, this extension may be viewed as an extension from two-particle coupling to four-particle coupling. Let for simplicity $N_{\ell_1} \approx 2^{\ell_1}$ and $P_{\ell_2} \approx 2^{\ell_2}$ and introduce the following MIEnKF estimator

$$\mu_n^{MIEnKF}[\varphi] := \sum_{(\ell_1,\ell_2)\in\mathcal{I}} \sum_{m=1}^{M_{\ell_1,\ell_2}} \frac{\Delta_{\ell_1,\ell_2}\mu_n^m[\varphi]}{M_{\ell_1,\ell_2}},$$

where the index-set $\mathcal{I} \subset \mathbb{N}_0^2$ that the estimator is summing over is a function of the accuracy $\epsilon > 0$ and $\{M_{\ell_1,\ell_2}\}_{(\ell_1,\ell_2)\in\mathcal{I}}$ denotes the number of four-coupled iid samples of EnKF estimators on level (ℓ_1, ℓ_2):

$$\Delta_{\ell_1,\ell_2}\mu_n^m[\varphi] := \left(\mu_n^{N_{\ell_1},P_{\ell_2},m} - \left(\mu_n^{N_{\ell_1},P_{\ell_2-1},1,m} + \mu_n^{N_{\ell_1},P_{\ell_2-1},2,m} \right)/2 \right.$$
$$\left. - \mu_n^{N_{\ell_1-1},P_{\ell_2},m} + \left(\mu_n^{N_{\ell_1-1},P_{\ell_2-1},1,m} + \mu_n^{N_{\ell_1-1},P_{\ell_2-1},2,m} \right)/2 \right)[\varphi].$$

Tentative numerical tests for both filtering problems considered in Section 3 yield that

$$\left| \mathbb{E}[\boldsymbol{\Delta}_{(\ell_1, \ell_2)} \mu_n^m[\varphi]] \right| \lesssim N_{\ell_1}^{-1} P_{\ell_2}^{-1},$$
$$\left\| \boldsymbol{\Delta}_{(\ell_1, \ell_2)} \mu_n^m[\varphi]] \right\|_p \lesssim N_{\ell_1}^{-1} P_{\ell_2}^{-1}.$$

On the basis of these rates, one may optimally determine the index set $\mathcal{I}$ and $\{M_{\ell_1, \ell_2}\}_{(\ell_1, \ell_2) \in \mathcal{I}}$, cf. [18]. A detailed description and performance study of the MIEnKF method is left as future work.

REFERENCES

[1] S. I. Aanonsen, G. Nævdal, D. S. Oliver, A. C. Reynolds, B. Vallès, et al., The ensemble Kalman filter in reservoir engineering–a review, *Spe Journal*, **14** (2009), 393–412.

[2] A. Beskos, A. Jasra, K. J. H. Law, Y. Marzouk and Y. Zhou, Multilevel sequential Monte Carlo with dimension-independent likelihood-informed proposals, *SIAM/ASA Journal on Uncertainty Quantification*, **6** (2018), 762–786.

[3] A. Beskos, A. Jasra, K. Law, R. Tempone and Y. Zhou, Multilevel sequential monte carlo samplers, *Stochastic Processes and their Applications*, **127** (2017), 1417–1440.

[4] J. Bezanson, A. Edelman, S. Karpinski and V. B. Shah, Julia: A fresh approach to numerical computing, *SIAM Review*, **59** (2017), 65–98.

[5] D. Blömker, C. Schillings, P. Wacker and S. Weissmann, *Well Posedness and Convergence Analysis of the Ensemble Kalman Inversion*, Inverse Problems, IOP Publishing, (2019).

[6] G. Burgers, P. J. van Leeuwen and G. Evensen, Analysis scheme in the ensemble Kalman filter, *Monthly Weather Review*, **126** (1998), 1719–1724.

[7] A. Chernov, H. Hoel, K. J. H. Law, F. Nobile and R. Tempone, Multilevel ensemble Kalman filtering for spatio-temporal processes, preprint, `arXiv:1710.07282`.

[8] N. D. Conrad, L. Helfmann, J. Zonker, S. Winkelmann and C. Schütte, Human mobility and innovation spreading in ancient times: A stochastic agent-based simulation approach, in *EPJ Data Science*, Springer, **7** (2018), 24.

[9] J. de Wiljes, S. Reich and W. Stannat, Long-time stability and accuracy of the ensemble Kalman–Bucy filter for fully observed processes and small measurement noise, *SIAM Journal on Applied Dynamical Systems*, **17**, (2018) 1152–1181.

[10] T. J. Dodwell, C. Ketelsen, R. Scheichl and A. L. Teckentrup, A hierarchical multilevel Markov chain Monte Carlo algorithm with applications to uncertainty quantification in subsurface flow, *SIAM/ASA Journal on Uncertainty Quantification*, **3** (2015), 1075–1108.

[11] O. G. Ernst, B. Sprungk and H. Starkloff, Analysis of the ensemble and polynomial chaos Kalman filters in Bayesian inverse problems, *SIAM/ASA Journal on Uncertainty Quantification*, **3** (2015), 823–851.

[12] G. Evensen, Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *Journal of Geophysical Research: Oceans*, **99(C5)** (1998), 10143–10162.

[13] K. Fossum, T. Mannseth and A. S. Stordal, Assessment of multilevel ensemble-based data assimilation for reservoir history matching, *Computational Geosciences*, **17** (2019), 1–23.

[14] M. B. Giles, Multilevel Monte Carlo path simulation, *Oper. Res.*, **56** (2008), 607–617.

[15] C. Graham and D. Talay, Stochastic simulation and Monte Carlo methods: Mathematical foundations of stochastic simulation, *Springer Science & Business Media*, **68** (2013).

[16] A. Gregory and C. J. Cotter, A seamless multilevel ensemble transform particle filter, *SIAM Journal on Scientific Computing*, **39** (2017), A2684–A2701.

[17] A. Gregory, C. J. Cotter and S. Reich, Multilevel ensemble transform particle filtering, *SIAM Journal on Scientific Computing*, **38** (2016), A1317–A1338.

[18] A. Haji-Ali, F. Nobile and R. Tempone, Multi-index Monte Carlo: When sparsity meets sampling, *Numerische Mathematik*, **132** (2016), 767–806.

[19] A. Haji-Ali and R. Tempone, Multilevel and Multi-index Monte Carlo methods for the McKean–Vlasov equation, *Statistics and Computing*, **28** (2018), 923–935.

[20] S. Heinrich, Multilevel Monte Carlo methods, *Large-Scale Scientific Computing*, (2001), 58–67.

[21] H. Hoel, J. Häppölä and R. Tempone, Construction of a mean square error adaptive Euler–Maruyama method with applications in multilevel Monte Carlo, *Monte Carlo and Quasi-Monte Carlo Methods*, (2016), 29–86.

[22] H. Hoel, K. J. H. Law and R. Tempone, Multilevel ensemble Kalman filtering, *SIAM Journal on Numerical Analysis*, **54** (2016), 1813–1839.

[23] H. Hoel, E. von Schwerin, A. Szepessy and R. Tempone, Adaptive multilevel monte carlo simulation, *Numerical Analysis of Multiscale Computations*, (2012), 217–234.

[24] H. Hoel, E. Von Schwerin, A. Szepessy and R. Tempone, Implementation and analysis of an adaptive multilevel Monte Carlo algorithm, *Monte Carlo Methods and Applications*, **20** (2014), 1–41.

[25] P. L. Houtekamer and H. L. Mitchell, Data assimilation using an ensemble Kalman filter technique, *Monthly Weather Review*, **126** (1998), 796–811.

[26] P. L. Houtekamer, H. L. Mitchell, G. Pellerin, M. Buehner, M. Charron, L. Spacek and B. Hansen, Atmospheric data assimilation with an ensemble Kalman filter: Results with real observations, *Monthly Weather Review*, **133** (2005), 604–620.

[27] A. Jasra, K. Kamatani, K. J. H. Law and Y. Zhou, Multilevel particle filters, *SIAM Journal on Numerical Analysis*, **55** (2017), 3068–3096.

[28] R. E. Kalman, A new approach to linear filtering and prediction problems, *Journal of basic Engineering*, **82** (1960), 35–45.

[29] E. Kalnay, *Atmospheric Modeling, Data Assimilation and Predictability*, Cambridge university press, (2003).

[30] D. T. B. Kelly, K. J. H. Law and A. M. Stuart, Well-posedness and accuracy of the ensemble Kalman filter in discrete and continuous time, *Nonlinearity*, **27** (2014), 2579.

[31] P. E. Kloeden and E. Platen, Numerical solution of stochastic differential equations, in *Applications of Mathematics (New York)*, 82, Springer-Verlag, Berlin, 1992.

[32] E. Kwiatkowski and J. Mandel, Convergence of the square root ensemble Kalman filter in the large ensemble limit, *SIAM/ASA Journal on Uncertainty Quantification*, **3** (2015), 1–17.

[33] T. Lange and W. Stannat, On the continuous time limit of the ensemble Kalman filter, preprint, `arXiv:1901.05204`.

[34] J. Latz, I. Papaioannou and E. Ullmann, Multilevel sequential Monte Carlo for Bayesian inverse problems, *Journal of Computational Physics*, **368** (2018), 154–178.

[35] K. J. H. Law, H. Tembine and R. Tempone, Deterministic mean-field ensemble Kalman filtering, *SIAM Journal on Scientific Computing*, **38** (2016), A1251–A1279.

[36] F. Le Gland, V. Monbet, V. Tran, et al., Large sample asymptotics for the ensemble Kalman filter, in *Oxford University Press* (eds. D. Crisan, B. Rozovskii), 2011, 598–631.

[37] J. Mandel, L. Cobb and J. D. Beezley, On the convergence of the ensemble Kalman filter, *Applications of Mathematics*, **56** (2011), 533–541.

[38] P. D. Moral, A. Jasra, K. J. H. Law and Y. Zhou, Multilevel sequential Monte Carlo samplers for normalizing constants, *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, **27** (2017), 20.

[39] B. Peherstorfer, K. Willcox and M. Gunzburger, Optimal model management for multifidelity Monte Carlo estimation, *SIAM Journal on Scientific Computing*, **38** (2016), A3163–A3194.

[40] A. Popov, C. Mou, T. Iliescu and A. Sandu, A multifidelity ensemble Kalman filter with reduced order control variates, preprint, `arXiv:2007.00793`.

[41] B. V. Rosić, A. Kučerová, J. Sỳkora, O. Pajonk, A. Litvinenko and H. G. Matthies, Parameter identification in a probabilistic setting, *Engineering Structures*, **50** (2013), 179–196.

[42] C. Schillings and A. M. Stuart, Analysis of the ensemble Kalman filter for inverse problems, *SIAM Journal on Numerical Analysis*, **55** (2017), 1264–1290.

[43] C. Schillings and A. M. Stuart, Convergence analysis of ensemble Kalman inversion: The linear, noisy case, *Applicable Analysis*, **97** (2018), 107–123.

[44] C. Schütte and M. Sarich, Metastability and Markov State Models in Molecular Dynamics, American Mathematical Soc., **24** (2013).
[45] A. Szepessy, R. Tempone and G. E. Zouraris, Adaptive weak approximation of stochastic differential equations, Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences, **54** (2001), 1169–1214.
[46] X. T. Tong, A. J. Majda and D. Kelly, Nonlinear stability and ergodicity of ensemble based Kalman filters, Nonlinearity, **29** (2016), 657.

E-mail address: hoel@uq.rwth-aachen.de
E-mail address: gaukhar.shaimerdenova@kaust.edu.sa
E-mail address: tempone@uq.rwth-aachen.de, raul.tempone@kaust.edu.sa