SIAM/ASA J. Uncertainty Quantification
Vol. 12, No. 2, pp. 411–441

Ensemble Kalman Filters with Resampling*

Omar Al-Ghattas†, Jiajun Bao‡, and Daniel Sanz-Alonso†

Abstract. Filtering is concerned with online estimation of the state of a dynamical system from partial and noisy observations. In applications where the state of the system is high dimensional, ensemble Kalman filters are often the method of choice. These algorithms rely on an ensemble of interacting particles to sequentially estimate the state as new observations become available. Despite the practical success of ensemble Kalman filters, theoretical understanding is hindered by the intricate dependence structure of the interacting particles. This paper investigates ensemble Kalman filters that incorporate an additional resampling step to break the dependency between particles. The new algorithm is amenable to a theoretical analysis that extends and improves upon those available for filters without resampling, while also performing well in numerical examples.

Key words. ensemble Kalman filter, resampling, nonasymptotic error bounds

MSC codes. 62F15, 68Q25, 60G35, 62M05

DOI. 10.1137/23M1594935

1. Introduction.

The filtering problem of estimating a time-evolving state from partial and noisy observations arises in numerous applications, including numerical weather prediction, automatic control, robotics, signal processing, machine learning, and finance [45, 11, 42, 5, 29, 36, 43]. When the state is high dimensional and the dynamics governing its evolution are complex, the method of choice is often the ensemble Kalman filter (EnKF) [15, 18, 17, 23, 19]. In this filtering algorithm, a Kalman gain matrix defined via the first two moments of an ensemble of particles determines the relative importance given to the dynamics and the observations in estimating the state. The size of the ensemble controls both the accuracy and the computational cost of the algorithm. Operational implementations of EnKF give accurate state estimation with a moderate ensemble size, significantly smaller than the state dimension [23]. However, nonasymptotic theory that explains the successful performance of EnKF with moderate ensemble size is still not fully developed. An important impediment to such a theory is the presence of correlations between particles, since the Kalman gain used to update each particle depends on the entire ensemble. This paper investigates a modification of EnKF that incorporates a resampling step to break these correlations. The new algorithm is amenable to a theoretical analysis that extends and improves upon those available for filters without resampling, while also maintaining a similar empirical performance.

*Received by the editors August 16, 2023; accepted for publication (in revised form) March 4, 2024; published electronically May 23, 2024.
https://doi.org/10.1137/23M1594935
Funding: This research was supported by NSF DMS-2027056, DOE DE-SC0022232, and the BBVA Foundation.
†Department of Statistics, University of Chicago, Chicago, IL 60637 USA (ghattas@uchicago.edu, sanzalonso@uchicago.edu).
‡Committee on Computational and Applied Mathematics, University of Chicago, Chicago, IL 60637 USA (jiajunb@uchicago.edu).

1.1. Resampling in filtering algorithms. Resampling techniques are routinely employed to enhance particle filtering algorithms which assimilate observations by weighting particles according to their likelihood [12, 13]. For particle filters, resampling converts weighted particles into unweighted ones to alleviate weight degeneracy and achieve variance reduction at later times [10, Chapter 9]. In contrast, EnKF assimilates observations by using *unweighted* particles and relying on a Gaussian *ansatz* and Kalman-type formulae. EnKF avoids weight degeneracy by design, but remains vulnerable to filter divergence and ensemble collapse [22, 26]; several works have proposed using resampling to remedy these issues.

An early discussion of resampling for EnKF can be found in [4], which replaces the standard Gaussian *ansatz* with a more flexible sum of Gaussian kernels. The paper [52] introduced bootstrap methods for identifying and alleviating the impact of spurious correlations, thereby enhancing the robustness of the Kalman gain. The work [32] proposed a resampling scheme to improve the performance of deterministic filters in nonlinear settings. This method involves periodically resampling the ensemble based on a "bootstrapping" approach as suggested by [4], which is fundamentally based on a kernel density technique taken from the particle filtering literature. Closest to our work is the paper [39], which demonstrates that resampling the Kalman gain in the conditioning step of EnKF can help prevent the ensemble from collapsing over time, consequently enhancing ensemble stability and reliability. The numerical experiments in [39] suggest that, relative to the nonresampled setting, EnKF algorithms that employ resampling give more reliable prediction intervals with a slight trade-off in the accuracy of their point predictions.

Ensemble Kalman methods are also used for offline parameter estimation and, relatedly, as numerical solvers for inverse problems; see, e.g., [21, 1, 33, 24, 7]. While not the focus of this paper, we point out that resampling techniques have also been investigated in this context. For instance, [51] removes particles that significantly deviate from the posterior distribution via a resampling procedure, thus improving the performance of standard implementations. A similar idea is also considered in [50], which proposes adding an extra resampling step in each iterative cycle. This method improves the convergence of the iterative EnKF by perturbing the shrinking ensemble covariances to prevent early stopping while preserving the consistent Kalman update direction of standard implementations.

1.2. Our contributions. Whereas previous work investigates resampling from a methodological viewpoint [4, 52, 32, 39], the primary objective of this paper is to demonstrate that resampling strategies provide a promising approach to the design of ensemble Kalman algorithms with nonasymptotic theoretical guarantees. We consider a simple parametric resampling scheme: at the beginning of each filtering step, members of the ensemble are independently sampled from a Gaussian distribution whose mean and covariance match those of the ensemble at the previous time-step. Thereafter, the filtering step can be carried out using any of the numerous existing EnKF variants [17, 46]. For the resulting algorithm, which we term REnKF, we establish theoretical guarantees that extend and improve upon those available for filters without resampling.

Our theoretical guarantees hold in the linear-Gaussian setting in which we provide a detailed error analysis of the ensemble mean and covariance as estimators of the mean and covariance of the filtering distributions, given by the Kalman filter [25]. Our theory covers

both stochastic and deterministic dynamical systems; in addition, it covers both stochastic implementations based on *perturbed observations* [16] and deterministic implementations based on *square-root* filters [46, 3, 6]. Importantly, our error-bounds are *nonasymptotic* and *dimension-free*: they hold for any given ensemble size and are written in terms of the *effective-dimension* of the covariance of the initial distribution, and of the dynamics and observation models. The nonasymptotic and dimension-free analysis of ensemble Kalman updates has recently been considered in [2], which demonstrated rigorously the success of ensemble Kalman updates whenever the ensemble size scaled with the effective dimension of the state as opposed to its ambient dimension. Given that ensemble Kalman algorithms are often employed in problems where the state dimension is very large, our results also contribute to the theoretical understanding of why ensemble methods are able to perform well even when the ensemble size is taken to be much smaller than the state dimension. This paper extends the results in [2] by providing new bounds over multiple assimilation cycles. Our work may also be compared to [37], which puts forward a nonasymptotic and dimension-free analysis of a multistep EnKF that utilizes a different modification than the one used to define REnKF. Specifically, [37] employs an additional projection step that determines the effective dimension of the method.

Other multistep analyses were limited to square-root filters with deterministic dynamics [28, 2] and to *asymptotic* analysis of stochastic implementations [28], which, while ensuring consistency of the filters, do not explain their practical success when deployed with a small ensemble size. The key reason why existing nonasymptotic analyses [2] do not extend to stochastic implementations and dynamics is that these additional sources of randomness further complicate the correlations between particles, which we break via resampling.

We numerically illustrate the theory in a linear setting and also demonstrate the successful performance of REnKF on the Lorenz 96 equations [34], a simplified model for atmospheric dynamics widely used to test filtering algorithms [35, 36, 30, 44]. In our experiments, REnKF performs similarly to standard, nonresampled EnKF in fully and partially-observed settings. Moreover, the results are robust to the noise level in the dynamics and in the observations. Python code to reproduce all numerical experiments is publicly available at https://github.com/Jiajun-Bao/EnKF-with-Resampling.

1.3. Outline. The rest of this paper is organized as follows. Section 2 formalizes the problem setting and provides necessary background on EnKF. Section 3 introduces and analyzes the new REnKF algorithm. The main result, Theorem 3.2, gives nonasymptotic and dimension-free error bounds. We report numerical results that confirm and complement the theory in section 4. Proofs are collected in section 5. We close in section 6 with a discussion of our results and directions for future research.

1.4. Notation. For a vector $u = (u(1), \dots, u(d))^\top$ and $q \geq 1$, $|u|_q = (\sum_{i=1}^d |u(i)|^q)^{1/q}$ and $|u| = |u|_2$. For a random variable X and $q \geq 1$, we write $\|X\|_q = (\mathbb{E}|X|^q)^{1/q}$ and $\|X\| = \|X\|_2$. $X \sim \mathcal{N}(m, C)$ denotes that X is a Gaussian random vector with mean m and covariance C, and we denote its density at a point x by $\mathcal{N}(x; m, C)$. $\mathcal{S}_+^d$ denotes the set of $d \times d$ symmetric positive-semidefinite matrices, and $\mathcal{S}_{++}^d$ denotes the set of $d \times d$ symmetric positive-definite matrices. For two $d \times d$ matrices A, B, $A \succ B$ implies $A - B \in \mathcal{S}_{++}^d$ and $A \succeq B$ implies $A - B \in \mathcal{S}_+^d$, and similarly for $\prec, \preceq$. For a $n \times m$ matrix $A = (A_{ij})_{i=1,j=1}^{n,m}$, the operator norm is given by $|A| = \sup_{\|v\|_2 = 1} |Av|_2$. $\mathbf{1}\{S\}$ denotes the indicator of the set S. The identity matrix

will be denoted by I, and on occasion its dimension will be made explicit with a subscript. The $n \times m$ zero matrix will be denoted by $O_{n\times m}$.

2. Problem setting and ensemble Kalman filters. We consider a d-dimensional unobserved *state process* $\{u^{(j)}\}_{j\geq 0}$ and a k-dimensional *observation process* $\{y^{(j)}\}_{j\geq 1}$ whose relationship over discrete time j is governed by the following hidden Markov model:

$$u^{(0)} \sim \mathcal{N}(\mu^{(0)}, \Sigma^{(0)}), \tag{2.1}$$

$$u^{(j)} = \Psi(u^{(j-1)}) + \xi^{(j)}, \qquad \xi^{(j)} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \Xi), \qquad j = 1, 2, \dots \tag{2.2}$$

$$y^{(j)} = Hu^{(j)} + \eta^{(j)}, \qquad \eta^{(j)} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \Gamma), \qquad j = 1, 2, \dots \tag{2.3}$$

We assume that the initial distribution $\mathcal{N}(\mu^{(0)}, \Sigma^{(0)})$, where $\mu^{(0)} \in \mathbb{R}^d$, $\Sigma^{(0)} \in \mathcal{S}^d_{++}$, the model dynamics map $\Psi : \mathbb{R}^d \to \mathbb{R}^d$, the observation matrix $H \in \mathbb{R}^{k\times d}$, and the dynamics and observation noise covariance matrices $\Xi \in \mathcal{S}^d_+, \Gamma \in \mathcal{S}^k_{++}$ are known; otherwise, they may be estimated from the observations; see, e.g., [19, 8, 9]. We further assume that the random variables $u^{(0)}$, $\{\xi^{(j)}\}_{j\geq 1}$, and $\{\eta^{(j)}\}_{j\geq 1}$ are mutually independent. All methods and theory presented in this paper extend immediately to dynamics and/or observation models that are not time homogeneous at the expense of a more cumbersome notation. Additionally, nonlinear observations can be dealt with by augmenting the state; see, e.g., [3].

For a given time index $j \in \mathbb{N}$, the *filtering* goal is to compute the *filtering distribution* $p(u^{(j)}|Y^{(j)})$, where $Y^{(j)} := \{y^{(1)}, \dots, y^{(j)}\}$. The filtering distribution provides a probabilistic summary of the state $u^{(j)}$ conditional on observations up to time j. Given access to the filtering distribution at the preceding time-step $j-1$, $p(u^{(j)}|Y^{(j)})$ may be obtained by the following two-step procedure:

$$p(u^{(j)}|Y^{(j-1)}) = \int \mathcal{N}(u^{(j)}; \Psi(u^{(j-1)}), \Xi) p(u^{(j-1)}|Y^{(j-1)})\, du^{(j-1)}, \tag{2.4}$$

$$p(u^{(j)}|Y^{(j)}) \propto \mathcal{N}(y^{(j)}; Hu^{(j)}, \Gamma) p(u^{(j)}|Y^{(j-1)}). \tag{2.5}$$

The *forecast distribution* $p(u^{(j)}|Y^{(j-1)})$ represents our knowledge of the state at time j given observations up to time $j-1$, and its computation in (2.4) utilizes the dynamics model (2.2). In the analysis step (2.5), the new observation y_j is assimilated through an application of Bayes formula with prior given by the forecast distribution and likelihood determined by the observation model (2.3). Closed-form expressions for the filtering and forecast distributions are only available for a small class of hidden Markov models [41]. For problems outside this class, many algorithms have been developed to approximate the filtering distributions, or, if this is too costly, to find point estimates of the state [45, 43].

This paper is concerned with EnKF algorithms that belong to the larger family of Kalman methods. These methods invoke a Gaussian *ansatz* for the forecast distribution, so that Bayes formula in the analysis step can be readily applied using the conjugacy of the Gaussian forecast distribution and the Gaussian likelihood model (2.3). The distinctive feature of EnKF is that the Gaussian approximation is defined using the first two moments of an ensemble of particles. Then, in the analysis step each individual particle is updated with a Kalman gain matrix which incorporates the forecast covariance. Several stochastic and deterministic implementations for the analysis step have been proposed in the literature; see, e.g., [23, 46, 17]. In

Algorithm 2.1, an example of a stochastic implementation of EnKF—commonly referred to as the Perturbed Observation EnKF—is provided for reference, and will be our focus for this work. At time $j = 0$, an *initial ensemble* of N particles are independently drawn from the initial distribution in (2.1). These ensemble members are then sequentially passed through forecast and analysis steps: In the forecast step, the ensemble is propagated through the system dynamics yielding the j-th *forecast ensemble.* In the analysis step, the new observation $y^{(j)}$ is assimilated by updating each ensemble member according to a Kalman-type formula, yielding the jth *analysis ensemble.* Although the initial ensemble members are mutually independent, the dependence structure of the ensemble is highly nontrivial beginning at the analysis step at time $j = 1$. Indeed, note that the Kalman Gain $K^{(1)}$ is a nonlinear transformation of the entire forecast ensemble, and this matrix is used to update each of the ensemble members when constructing the analysis ensemble. The recursive nature of the algorithm further complicates the dependence structure of the ensemble, rendering a nonasymptotic analysis highly challenging.

The stochastic variant of EnKF in Algorithm 2.1 is arguably the most popular in applications [15, 48]. Unfortunately, as noted in [20, 2] and further discussed in section 3, it is harder to analyze from a nonasymptotic viewpoint than deterministic variants of the EnKF.

The output $\widehat{\mu}^{(j)}$ of EnKF gives a point estimate of the state $u^{(j)}$ at time j. For such a state-estimation task, EnKF is very effective [31]. Additionally, the output $\widehat{\Sigma}^{(j)}$ may be used to construct confidence intervals. However, as often noted in the literature [14, 31] and further discussed in section 4, caution should be exercised when using ensemble Kalman algorithms for such uncertainty quantification tasks. EnKF performance for state estimation and uncertainty

Algorithm 2.1. Ensemble Kalman filter (EnKF).

1: **Input**: $\Psi, H, \Xi, \Gamma, \mu^{(0)}, \Sigma^{(0)}, N$. Sequentially acquired data $\{y^{(j)}\}_{j\geq 1}$.
2: **Initialization**: $u_n^{(0)} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu^{(0)}, \Sigma^{(0)}), \quad 1 \leq n \leq N$.
3: For $j = 1, 2, \ldots$ do the following forecast and analysis steps:
4: **Forecast**:

$$\begin{aligned} &\widehat{u}_n^{(j)} = \Psi(u_n^{(j-1)}) + \xi_n^{(j)}, \quad \xi_n^{(j)} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \Xi), \quad 1 \leq n \leq N, \\ &\widehat{m}^{(j)} = \frac{1}{N}\sum_{n=1}^{N} \widehat{u}_n^{(j)}, \qquad \widehat{C}^{(j)} = \frac{1}{N-1}\sum_{n=1}^{N} (\widehat{u}_n^{(j)} - \widehat{m}_j)(\widehat{u}_n^{(j)} - \widehat{m}_j)^\top. \end{aligned} \tag{2.6}$$

5: **Analysis**:

$$\begin{aligned} &K^{(j)} = \widehat{C}^{(j)} H^\top (H \widehat{C}^{(j)} H^\top + \Gamma)^{-1}, \\ &y_n^{(j)} = y^{(j)} + \eta_n^{(j)}, \quad \eta_n^{(j)} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \Gamma), \quad 1 \leq n \leq N, \\ &u_n^{(j)} = (I - K^{(j)} H) \widehat{u}_n^{(j)} + K^{(j)} y_n^{(j)}, \quad 1 \leq n \leq N, \\ &\widehat{\mu}^{(j)} = \frac{1}{N}\sum_{n=1}^{N} u_n^{(j)}, \qquad \widehat{\Sigma}^{(j)} = \frac{1}{N-1}\sum_{n=1}^{N} (u_n^{(j)} - \widehat{\mu}^{(j)})(u_n^{(j)} - \widehat{\mu}^{(j)})^\top. \end{aligned} \tag{2.7}$$

6: **Output**: Analysis mean $\widehat{\mu}^{(j)}$ and covariance $\widehat{\Sigma}^{(j)}$ for $j = 1, 2, \ldots$

quantification tasks can be assessed by the error in approximating the mean and covariance of the filtering distributions; the theory in subsection 3.2 adopts such performance metrics. If the moments of the filtering distributions are not available, performance metrics such as root mean squared error and coverage of confidence intervals can be employed [31], and we do so in the numerical experiments in section 4.

3. Ensemble Kalman filters with resampling.

In this section, we first introduce and motivate our main algorithm, EnKF with resampling (REnKF). We then present the nonasymptotic theoretical analysis of REnKF in a linear model dynamics setting.

3.1. Main algorithm.

The idea underlying REnKF, which is outlined in Algorithm 3.1, is to employ a resampling step at each filtering cycle to break the correlations between ensemble members described in section 2. We consider here a particularly simple parametric resampling scheme in which at the beginning of each filtering cycle, ensembles are independently sampled from a Gaussian distribution whose mean and covariance match those of the analysis ensemble at the previous time step. Although the resampling mechanism can be made to be more sophisticated—for example, one may consider nonparametric resampling schemes in which the empirical distribution of the ensemble is used instead—we note that such complications may be difficult to justify given the simplicity, theoretical guarantees (subsection 3.2), as well as the computational scalability and empirical performance (section 4) of the proposed resampling strategy. Other than the resampling step, the forecast and analysis steps of REnKF agree with those of EnKF, and consequently any of the stochastic or deterministic implementations of EnKF can be adopted. Our focus here is on the stochastic implementation of EnKF in Algorithm 2.1. As discussed in the next subsection—see Remarks 3.1 and 3.3—nonasymptotic theory for deterministic implementations can be obtained as a by-product of the theory that we develop.

Notice from Algorithm 3.1 that correlations between particles could alternately be broken by resampling between the forecast and analysis steps. While such an approach would be amenable to a nonasymptotic analysis akin to the one we develop, we empirically found that resampling after the forecast step significantly deteriorates the performance of the filter in nonlinear settings. A heuristic explanation is that resampling tacitly introduces a Gaussian approximation, and the filtering distribution is better approximated by a Gaussian than the forecast distribution when the dynamics are nonlinear and the observations are Gaussian.

Algorithm 3.1. Ensemble Kalman filter with resampling (REnKF).

1: **Input**: $\Psi, H, \Xi, \Gamma, \mu^{(0)}, \Sigma^{(0)}, N$. Sequentially acquired data $\{y^{(j)}\}_{j\geq 1}$.
2: **Initialization**: Set $\widehat{\mu}^{(0)} = \mu^{(0)}$ and $\widehat{\Sigma}^{(0)} = \Sigma^{(0)}$.
3: For $j = 1, 2, \ldots$ do the following resampling, forecast, and analysis steps:
4: **Resampling**:

$$u_n^{(j-1)} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\widehat{\mu}^{(j-1)}, \widehat{\Sigma}^{(j-1)}), \quad 1 \leq n \leq N. \tag{3.1}$$

5: **Forecast**: Do (2.6).
6: **Analysis**: Do (2.7).
7: **Output**: Analysis mean $\widehat{\mu}^{(j)}$ and covariance $\widehat{\Sigma}^{(j)}$ for $j = 1, 2, \ldots$

3.2. Nonasymptotic error bounds. Here we present theoretical guarantees for REnKF in a linear dynamics setting. We introduce the setting and necessary background in subsection 3.2.1. Then, the main result is stated and discussed in subsection 3.2.2.

3.2.1. Setting and preliminaries. We consider REnKF in the following linear version of the hidden Markov model governing the relationship between the state and observation processes:

$$u^{(0)} \sim \mathcal{N}(\mu^{(0)}, \Sigma^{(0)}), \tag{3.2}$$

$$u^{(j)} = Au^{(j-1)} + \xi^{(j)}, \qquad \xi^{(j)} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \Xi), \qquad j = 1, 2, \dots \tag{3.3}$$

$$y^{(j)} = Hu^{(j)} + \eta^{(j)}, \qquad \eta^{(j)} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \Gamma), \qquad j = 1, 2, \dots \tag{3.4}$$

with $u^{(0)}$ independent of the i.i.d. sequences $\{\xi^{(j)}\}$ and $\{\eta^{(j)}\}$. Thus, we assume that the dynamics map Ψ in (2.2) is linear and represented by a given matrix $A \in \mathbb{R}^{d\times d}$. In this case, it is well known that the forecast distributions $p(u^{(j)}|Y^{(j-1)}) = \mathcal{N}(u^{(j)}; m^{(j)}, C^{(j-1)})$ and the filtering distributions $p(u^{(j)}|Y^{(j)}) = \mathcal{N}(u^{(j)}; \mu^{(j)}, \Sigma^{(j)})$ are both Gaussian, and the means and covariances of these distributions are given by the Kalman filter [43]. We aim to derive nonasymptotic bounds between the outputs $\widehat{\mu}^{(j)}$ and $\widehat{\Sigma}^{(j)}$ of REnKF and the output $\mu^{(j)}$ and $\Sigma^{(j)}$ of the Kalman filter.

We follow the exposition in [28] and introduce three operators that are central to the theory: the *Kalman gain* operator $\mathscr{K}$, the *mean-update* operator $\mathscr{M}$, and the *covariance-update* operator $\mathscr{C}$, defined, respectively, by

$$\mathscr{K} : \mathcal{S}_+^d \to \mathbb{R}^{d\times k}, \qquad \mathscr{K}(C) = CH^\top (HCH^\top + \Gamma)^{-1}, \tag{3.5}$$

$$\mathscr{M} : \mathbb{R}^d \times \mathcal{S}_+^d \to \mathbb{R}^d, \qquad \mathscr{M}(m, C; y) = m + \mathscr{K}(C)(y - Hm), \tag{3.6}$$

$$\mathscr{C} : \mathcal{S}_+^d \to \mathcal{S}_+^d, \qquad \mathscr{C}(C) = (I - \mathscr{K}(C)H)C. \tag{3.7}$$

With this notation, the mean and covariance updates from time $j-1$ to time j given by the Kalman filter are summarized in Table 1. The table also shows the corresponding updates for REnKF, where $\bar{u}^{(j-1)}$, $\bar{\xi}^{(j)}$, and $\bar{\eta}^{(j)}$, respectively, denote the sample means of $\{u_n^{(j-1)}\}_{n=1}^N$, $\{\xi_n^{(j)}\}_{n=1}^N$, and $\{\eta_n^{(j)}\}_{n=1}^N$; $S^{(j-1)}$ denotes the empirical covariance of $\{u_n^{(j-1)}\}_{n=1}^N$; and $\widehat{C}_{u\xi}^{(j)} = (\widehat{C}_{\xi u}^{(j)})^\top$ denotes the empirical cross-covariance of $\{u_n^{(j-1)}\}_{n=1}^N$ and $\{\xi_n^{(j)}\}_{n=1}^N$. Finally, following [20, 2], we refer to

$$\begin{aligned}\widehat{O}^{(j)} := {} & \mathscr{K}(\widehat{C}^{(j)})(\widehat{\Gamma}^{(j)} - \Gamma)\mathscr{K}^\top(\widehat{C}^{(j)}) \\ & + (I - \mathscr{K}(\widehat{C}^{(j)})H)\widehat{C}_{u\eta}^{(j)}\mathscr{K}^\top(\widehat{C}^{(j)}) + \mathscr{K}(\widehat{C}^{(j)})(\widehat{C}_{u\eta}^{(j)})^\top(I - H^\top\mathscr{K}^\top(\widehat{C}^{(j)}))\end{aligned}$$

as the *offset,* where $\widehat{\Gamma}^{(j)}$ denotes the empirical covariance of $\{\eta_n^{(j)}\}_{n=1}^N$, and $\widehat{C}_{u\eta}^{(j)}$ denotes the empirical cross-covariance of $\{u_n^{(j-1)}\}_{n=1}^N$ and $\{\eta_n^{(j)}\}_{n=1}^N$.

Remark 3.1 (deterministic implementations). As noted earlier, our presentation and analysis will focus on the stochastic (perturbed observation) implementation of EnKF described in Algorithm 2.1, and which is used within REnKF; see Algorithm 3.1. We claim that this approach is sufficient to cover both deterministic and stochastic updates. Indeed, [2] shows that deterministic and stochastic updates at time j can be succinctly written as

Table 1
Kalman filter and REnKF *updates in terms of the operators* (3.5), (3.6), *and* (3.7).

	Kalman filter	REnKF
Forecast Mean	$m^{(j)} = A\mu^{(j-1)}$	$\widehat{m}^{(j)} = A\bar{u}^{(j-1)} + \bar{\xi}^{(j)}$
Forecast Cov.	$C^{(j)} = A\Sigma^{(j-1)}A^\top + \Xi$	$\widehat{C}^{(j)} = AS^{(j-1)}A^\top + \widehat{\Xi}^{(j)} + A\widehat{C}^{(j)}_{u\xi} + \widehat{C}^{(j)}_{\xi u}A^\top$
Analysis Mean	$\mu^{(j)} = \mathscr{M}(m^{(j)}, C^{(j)}; y^{(j)})$	$\widehat{\mu}^{(j)} = \mathscr{M}(\widehat{m}^{(j)}, \widehat{C}^{(j)}; y^{(j)}) + \mathscr{K}(\widehat{C}^{(j)})\bar{\eta}^{(j)}$
Analysis Cov.	$\Sigma^{(j)} = \mathscr{C}(C^{(j)})$	$\widehat{\Sigma}^{(j)} = \mathscr{C}(\widehat{C}^{(j)}) + \widehat{O}^{(j)}$

$$\begin{aligned} \widehat{\mu}^{(j)} &= \mathscr{M}(\widehat{m}^{(j)}, \widehat{C}^{(j)}) + \varphi\mathscr{K}(\widehat{C}^{(j)})\bar{\eta}^{(j)}, \\ \widehat{\Sigma}^{(j)} &= \mathscr{C}(\widehat{C}^{(j)}) + \varphi\widehat{O}^{(j)}, \end{aligned} \tag{3.8}$$

where $\varphi = 1$ for the stochastic update and $\varphi = 0$ for the deterministic update. Therefore, relative to the deterministic update, theory for the stochastic update is additionally complicated by the need to consider the term $\mathscr{K}(\widehat{C}^{(j)})\bar{\eta}^{(j)}$ in the mean update and the offset term $\widehat{O}^{(j)}$ in the covariance update. Accordingly, we are able to provide a result for the resampled version of the deterministic (square-root) EnKF as a by-product of our more general theory, and we refer to Remark 3.3 for further discussion.

3.2.2. Main result. We define the *effective dimension* [47] of a matrix $Q \in \mathcal{S}^d_+$ by

$$r_2(Q) := \frac{\mathrm{Tr}(Q)}{|Q|}, \tag{3.9}$$

where $\mathrm{Tr}(Q)$ and $|Q|$ denote the trace and operator norm of Q. The effective dimension quantifies the number of directions where Q has significant spectral content and may be significantly smaller than the ambient dimension d when the eigenvalues of Q decay quickly. As such, it is a more refined measure of complexity in high-dimensional problems with underlying low-dimensional structure. The monographs [47, 49] refer to $r_2(Q)$ as the intrinsic dimension, while [27] uses the term effective rank. This terminology is motivated by the observation that $1 \le r_2(Q) \le \mathrm{rank}(Q) \le d$ and that $r_2(Q)$ is insensitive to changes in the scale of Q; see [47]. We now state our main result, Theorem 3.2, which provides nonasymptotic bounds on the deviation of REnKF from the Kalman filter for any time j.

Theorem 3.2. *Consider REnKF, Algorithm 3.1, with linear dynamics $\Psi(\cdot) = A\cdot$. Suppose that $N \ge r_2(\Sigma^{(0)}) \vee r_2(\Gamma) \vee r_2(\Xi)$. For any $j = 1, 2, \ldots,$ and $q \ge 1$,*

$$\| |\widehat{\mu}^{(j)} - \mu^{(j)}| \|_q \le c_1 \left(\sqrt{\frac{r_2(\Sigma^{(0)})}{N}} \vee \sqrt{\frac{r_2(\Xi)}{N}} \vee \sqrt{\frac{r_2(\Gamma)}{N}} \right), \tag{3.10}$$

$$\| |\widehat{\Sigma}^{(j)} - \Sigma^{(j)}| \|_q \le c_2 \left(\sqrt{\frac{r_2(\Sigma^{(0)})}{N}} \vee \sqrt{\frac{r_2(\Xi)}{N}} \vee \sqrt{\frac{r_2(\Gamma)}{N}} \right), \tag{3.11}$$

where $\mu^{(j)}$ and $\Sigma^{(j)}$ are the mean and covariance of the filtering distributions, and c_1, c_2 are potentially different universal constants depending on

$$|\Sigma^{(0)}|, |A|, |H|, |\Gamma^{-1}|, |\Gamma|, |\Xi|, q, j,$$

and c_1 additionally depends on $\{|y^{(\ell)} - Hm^{(\ell)}|\}_{\ell \le j}$.

With the exception of [37, Theorem 3.4], which relies on covariance inflation and an additional projection step, Theorem 3.2 seems to be the first result in the literature that provides nonasymptotic guarantees on the performance of a stochastic EnKF over multiple assimilation cycles. We note that the assumption $N \geq r_2(\Sigma^{(0)}) \vee r_2(\Gamma) \vee r_2(\Xi)$ is merely for convenience and can be removed at the expense of a more cumbersome statement of the result. Importantly, the bounds (3.10) and (3.11) are nonasymptotic, in that they hold for a fixed ensemble size N. Further, the bounds are dimension-free as they do not exhibit any dependence on the state-space dimension d, implying that the ensemble need not scale with d in order for the algorithm to perform well, as has been observed empirically in the literature and confirmed in our numerical results in section 4. Finally, similar to previous accuracy analyses for square-root ensemble Kalman filters [38, 2], variational data assimilation algorithms [44, 30], and particle filters [43, Chapters 11 and 12], our proof relies on induction over the discrete time index j and does not account for potential dissipation of errors due to filter ergodicity. As a result, the constants c_1 and c_2 grow with j and our bounds (3.10) and (3.11) do not hold uniformly in time without, for instance, stability requirements on A.

Remark 3.3 (resampled square-root filter). While the result in Theorem 3.2 is specific to the stochastic REnKF in Algorithm 3.1, using the observation made in Remark 3.1 it is possible to show that for a deterministic variant, namely the square-root REnKF, and under the same assumptions on the ensemble size made in Theorem 3.2, we have that

$$
\begin{aligned}
\||\widehat{\mu}^{(j)} - \mu^{(j)}|\|_q &\leq c_1 \left(\sqrt{\frac{r_2(\Sigma^{(0)})}{N}} \vee \sqrt{\frac{r_2(\Xi)}{N}} \right), \\
\||\widehat{\Sigma}^{(j)} - \Sigma^{(j)}|\|_q &\leq c_2 \left(\sqrt{\frac{r_2(\Sigma^{(0)})}{N}} \vee \sqrt{\frac{r_2(\Xi)}{N}} \right),
\end{aligned}
\tag{3.12}
$$

where c_1, c_2 are potentially different universal constants depending on $|\Sigma^{(0)}|$, $|A|$, $|H|$, $|\Gamma^{-1}|$, $|\Xi|$, q, j, and c_1 additionally depends on $\{|y^{(\ell)} - Hm^{(\ell)}|\}_{\ell \leq j}$. In contrast to (3.10) and (3.11), the bounds in (3.12) do not depend on the effective dimension of the noise covariance, $r_2(\Gamma)$, nor do the associated constants depend on $|\Gamma|$. The statistical price to pay for utilizing stochastic rather than deterministic updates is captured by these terms. We further note that [2, Corollary A.12], gives a nonasymptotic and multistep analysis of a simplified version of the square-root filter (without resampling) with deterministic dynamics (that is, $\Xi = O_{d\times d}$). In such a setting, [2, Corollary A.12] implies the following bounds:

$$
\||\widehat{\mu}^{(j)} - \mu^{(j)}|\|_q \leq c_3 \sqrt{\frac{r_2(\Sigma^{(0)})}{N}}, \qquad \||\widehat{\Sigma}^{(j)} - \Sigma^{(j)}|\|_q \leq c_4 \sqrt{\frac{r_2(\Sigma^{(0)})}{N}},
$$

where c_3, c_4 are potentially different universal constants depending on $|\Sigma^{(0)}|$, $|A|$, $|H|$, $|\Gamma^{-1}|$, q, j, and c_3 additionally depends on $\{|y^{(\ell)} - Hm^{(\ell)}|\}_{\ell \leq j}$. Theorem 3.2 should further be compared to [28, Theorem 6.1], which is also limited to the case $\Xi = O_{d\times d}$ and shows that $\||\widehat{\mu}^{(j)} - \mu^{(j)}|\|_q \leq c_3' N^{-1/2}$ and $\||\widehat{\Sigma}^{(j)} - \Sigma^{(j)}|\|_q \leq c_4' N^{-1/2}$, where c_3', c_4' are universal constants with the same dependencies as c_3 and c_4. Importantly, the bounds in [28, Theorem 6.1] do not capture the dependence of the algorithm on the prior covariance and also cannot be easily extended to handle stochastic dynamics $\Xi \succ 0$ as accomplished in Theorem 3.2.

4. Numerical results. In this section, we investigate the empirical performance of REnKF (Algorithm 3.1) and provide detailed comparisons to the stochastic EnKF (Algorithm 2.1). In subsection 4.1, we consider a linear dynamics map, $\Psi(\cdot) = A\cdot$, with the primary goal of demonstrating the bounds of Theorem 3.2 in simulated settings. In subsection 4.2, we study a nonlinear setting where Ψ represents the Δt-flow of the Lorenz 96 system, and Δt is the (constant) time-span between observations. The aim of this subsection is to show that REnKF achieves comparable performance to EnKF even in challenging nonlinear regimes, further motivating the study of resampling in the context of ensemble algorithms. In both subsections 4.1 and 4.2, we examine the performance of REnKF and EnKF under varying noise levels, ensemble sizes, and state dimensions. Additionally, in subsection 4.2, we consider cases in which we have access to either fully observed or partially observed dynamics. These scenarios offer a comprehensive perspective on the adaptability of REnKF to varying observational conditions, thereby highlighting its potential for wide applicability in real-world situations where data are often limited or incomplete. For all experiments, we generate a ground-truth state process $\{u^{(j)}\}_{j=0}^{J}$ for a time-window of length $J = 200$ using the initialization (2.1) and dynamics model (2.2). For each set of system parameters we examine, a unique set of observations $\{y^{(j)}\}_{j=1}^{J}$ is generated from the ground-truth state process utilizing the observation model (2.3). Python code to reproduce all numerical experiments is publicly available at https://github.com/Jiajun-Bao/EnKF-with-Resampling.

4.1. Linear dynamics. In this subsection, we numerically investigate the performance of REnKF for the linear-Gaussian hidden Markov model (3.2)–(3.4) analyzed in subsection 3.2. We will consider a variety of choices for the initial distribution, the dynamics noise covariance, and the observation noise covariance. Throughout, we take identity dynamics $A = I_d$ and full observations $H = I_d$. To compare the performance of EnKF and REnKF, we will consider the following metrics:

$$\text{(4.1)} \qquad \text{(Mean error)} \qquad \mathsf{E}_{\text{Linear}} = \frac{1}{J}\sum_{j=1}^{J} |\widehat{\mu}^{(j)} - \mu^{(j)}|_2,$$

$$\text{(4.2)} \qquad \text{(CI width)} \qquad \mathsf{W} = \frac{1}{J}\sum_{j=1}^{J}\frac{1}{d}\sum_{i=1}^{d} 2 \times 1.96\sqrt{\widehat{\Sigma}_{ii}^{(j)}},$$

$$\text{(4.3)} \qquad \text{(CI coverage)} \qquad \mathsf{V} = \frac{1}{J}\sum_{j=1}^{J}\frac{1}{d}\sum_{i=1}^{d} \mathbf{1}\left\{ u^{(j)}(i) \in \left(\widehat{\mu}^{(j)}(i) \pm 1.96\sqrt{\widehat{\Sigma}_{ii}^{(j)}}\right)\right\}.$$

The mean error (4.1) quantifies the approximation of the EnKF/REnKF analysis mean to the mean $\mu^{(j)}$ of the KF. Our theory for REnKF provides nonasymptotic bounds for this error, and our numerical results will show that this error is similar to that of EnKF in a variety of settings. The confidence interval (CI) width and coverage in (4.2)–(4.3) assess the ability of the filter to provide reliable uncertainty quantification: a short interval with high coverage would be preferable, but an overconfident short width interval with low coverage can lead to a misleading and potentially dangerous assessment of uncertainty. We illustrate these three metrics in Figure 1, which corresponds to a setup outlined in Table 2. This setup will be further explored in subsection 4.1.1. As depicted in the plot, the indicator in (4.3)

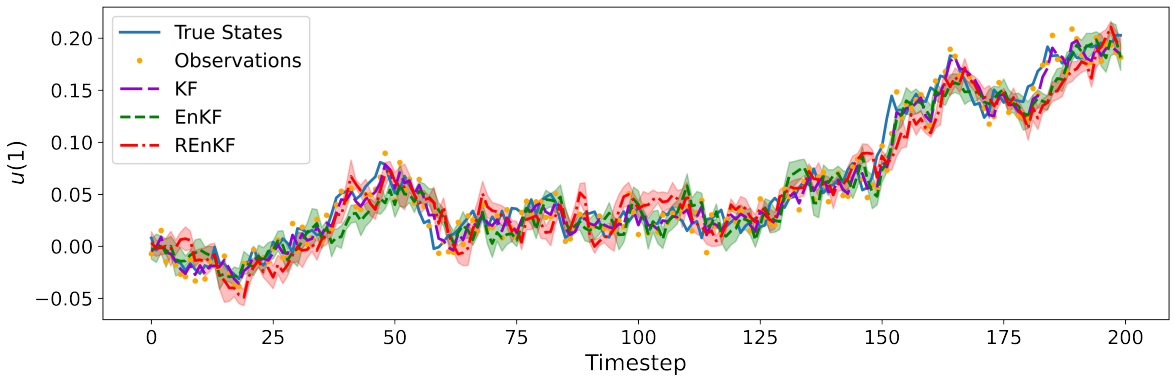

Figure 1. *State estimation and uncertainty quantification for coordinate $u(1)$ in the linear setting with ensemble size $N = 10$ and small noise $\alpha = 10^{-4}$. Note that the Kalman filter (KF) is optimal in the linear setting.*

Table 2
Performance metrics in the linear setting with $d = 20$.

Ensemble	Metric	Small noise $\alpha = 10^{-4}$	Moderate noise $\alpha = 10^{-2}$	Large noise $\alpha = 10^{-1}$
$N = 10$	EnKF Mean error	0.0608	0.6133	1.9931
	REnKFMean error	0.0616	0.6199	2.0310
	EnKF CI width	0.0194	0.1940	0.6134
	REnKFCI width	0.0188	0.1875	0.5930
	EnKF CI coverage (%)	39.57	38.90	38.35
	REnKF CI coverage (%)	37.83	37.14	36.58
$N = 40$	EnKF Mean error	0.0193	0.1930	0.6243
	REnKF Mean error	0.0209	0.2091	0.6739
	EnKF CI width	0.0278	0.2780	0.8790
	REnKF CI width	0.0274	0.2739	0.8663
	EnKF CI coverage (%)	69.94	69.26	68.90
	REnKF CI coverage (%)	68.65	67.76	67.43

corresponds to whether the solid blue line (representing the true states) fall within the shaded confidence intervals. We point out that the ability of ensemble Kalman methods to provide reliable uncertainty quantification, especially in nonlinear settings, has often been questioned [14, 31]. Our results will show that the CIs obtained with REnKF have similar width and coverage as those obtained by EnKF, but that coverage for both algorithms is not reliable when the ensemble size is small (subsections 4.1 and 4.2) or the dynamics are highly nonlinear (subsection 4.2).

Since the outputs $\{\widehat{\mu}^{(j)}, \widehat{\Sigma}^{(j)}\}_{j=1}^{J}$ of EnKF and REnKF are random, for each experiment we run both algorithms M times and we report the average value of the metrics (4.1), (4.2), and (4.3) as well as the value of M. More details can be found in Appendix A.

4.1.1. Effects of noise level and ensemble size. We perform two distinct analyses to assess the impact of different variables on the performance of EnKF and REnKF. The first,

which we term the *noise-level* analysis, investigates the relationship between mean error, $\mathsf{E}_{\text{Linear}}$, and the noise level, α. The second analysis, referred to as the *ensemble-size analysis*, explores how the mean error varies with the ensemble size, N. Both analyses are carried out using a fixed state dimension $d = 20$.

In the noise-level analysis, α is varied over a grid of 15 evenly spaced values between 10^{-16} and 1, allowing us to investigate a range of scenarios beginning with those with virtually no noise to those with substantial noise. In order to isolate the influence of α, we maintain the initial distribution with a fixed zero mean and covariance $\Sigma^{(0)} = 10^{-8} \times I_{20}$, as well as a fixed ensemble size of $N = 20$. In the ensemble-size analysis, N is varied between 10 and 100, in increments of 10. To isolate the effects of N, we fix $\alpha = 10^{-1}$ and maintain the initial distribution to have a fixed zero mean and covariance $\Sigma^{(0)} = 1.1\alpha \times I_{20}$. The covariance is adjusted to represent a higher initial uncertainty level compared to the noise-level analysis. The factor 1.1 was introduced to ensure that the initial states possess a slightly different level of uncertainty relative to the noise in the dynamics and observations. Both analyses are averaged over $M = 10$ runs of the algorithms. The results of both analyses are depicted in Figure 2. In addition to $\mathsf{E}_{\text{Linear}}$, in Table 2 we consider the effect of varying α and N on CI widths, W, and CI coverage, V. Here, we categorize the levels of noise as being either small, moderate, or large, which correspond to α values of 10^{-4}, 10^{-2}, or 10^{-1} respectively, as described under Case A in Table 3. Further, we repeat the experiments with ensembles of size $N = 10$ and $N = 40$. For the experimental settings summarized in Table 2, the state dimension and initial distribution are taken as in the ensemble-size analysis described earlier. These metrics are calculated based on averages over $M = 100$ runs of the algorithms.

The results in Figure 2 and in Table 2 confirm that across a wide variety of linear experimental settings, REnKF exhibits similar performance to EnKF as measured by the mean error, CI width, and CI coverage.

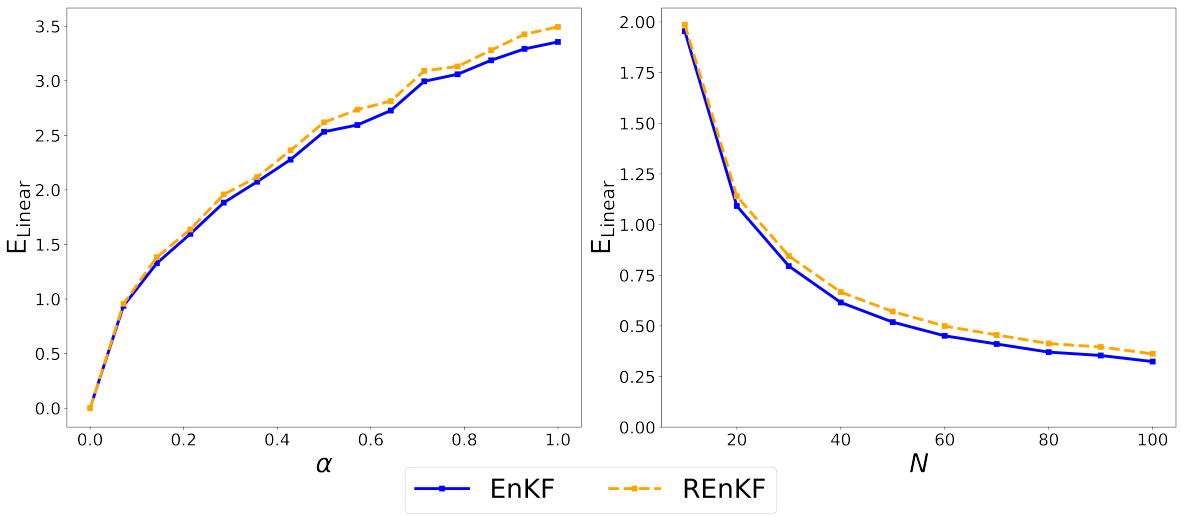

Figure 2. *Effects of α and N in the linear setting with $d = 20$.*

4.1.2. Effects of state dimension and spectrum decay. We now study the sensitivity of EnKF and REnKF to changes in the state dimension, d. Recall that our main result, Theorem 3.2, implies that REnKF performs well whenever the ensemble size scales with the largest of the effective dimensions of the noise covariances: $\Sigma^{(0)}$, Γ, and Ξ. This motivates our study of covariance matrices with structure summarized in Cases A and B of Table 3. In Case A, the effective dimension of the covariance matrix is proportional to the state dimension, d, and so the theory suggests that REnKF will do well only if the ensemble size also scales with d. In Case B, we consider covariance matrices that are diagonal, with ith diagonal element proportional to $i^{-\beta}$ where $\beta > 0$ is a rate parameter controlling the speed of decay. Table 4 demonstrates that two matrices of this form that are equal in dimension may differ drastically in their effective dimension for different choices of β. Here, then, the theory suggests that REnKF will do well so long as the ensemble size scales with the effective dimension, which may be much smaller than d. To test our theory, we run REnKF under both Cases A and B in Table 3 where d is varied over the set $\{2^1, 2^2, \ldots, 2^8\}$ and where the ensemble size is fixed at $N = 10$ throughout. For both cases we fix $\alpha = 10^{-4}$ and for Case B we consider $\beta \in \{0.1, 1, 1.5\}$. Figure 3 presents the results of averaging $\mathsf{E}_{\text{Linear}}$ over $M = 10$ runs of the algorithm in each of the experimental set-ups. We see that for all choices of β, EnKF and REnKF exhibit near-identical performance. For Case A, the performance deteriorates as d increases and this behavior is identical across all three displays. For Case B, when $\beta = 0.1$ (first display) so that the effective dimension increases significantly with dimension as described in the first row of Table 4, the performance deteriorates significantly as d increases. As β is increased to 1 in the second display, so that the effective dimension grows slowly with d, performance deteriorates at a much slower rate. This is further pronounced in the final display with $\beta = 1.5$.

Table 3

Covariance matrix settings explored numerically in subsections 4.1.2 and 4.2.

Noise	Case A	Case B ($i = 1, \ldots, d$)	Case C
Dynamics (Ξ)	$\Xi^A = \alpha \times I_d$	$\Xi^B_{ii} = \alpha \times i^{-\beta}$	$\Xi^C = \alpha \times I_d$
Observation (Γ)	$\Gamma^A = \alpha \times I_d$	$\Gamma^B_{ii} = \alpha \times i^{-\beta}$	$\Gamma^C = \alpha \times I_{\frac{2d}{3}}$
Prior ($\Sigma^{(0)}$)	$(\Sigma^{(0)})^A = 1.1 \times \Xi^A$	$(\Sigma^{(0)})^B_{ii} = 1.1 \times \Xi^B_{ii}$	$(\Sigma^{(0)})^C = 1.1 \times \Xi^C$

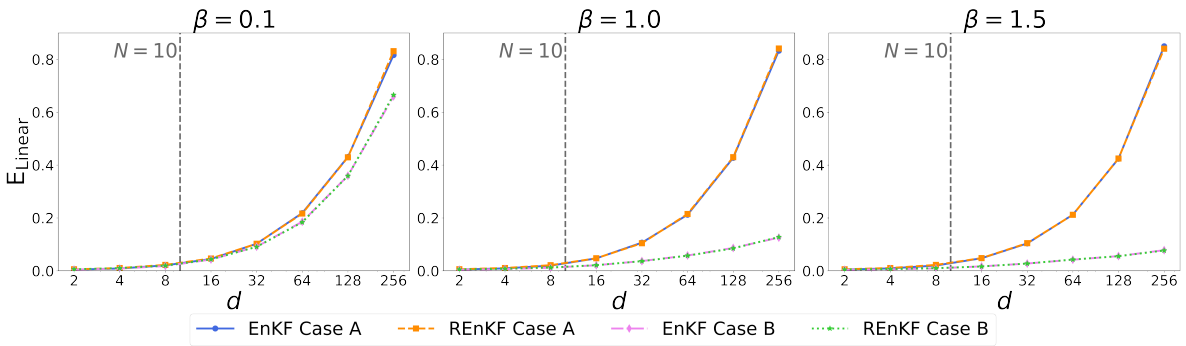

Figure 3. *Effect of spectrum decay in the linear setting.*

These numerical results demonstrate the key role played by the effective dimension in determining the performance of EnKF and REnKF, and are in agreement with Theorem 3.2 for REnKF.

4.2. Lorenz 96 dynamics. In this subsection, we extend our numerical investigation of REnKF to the nonlinear setting by taking Ψ in (2.2) to be the Δt-flow of the Lorenz 96 equations. Here Δt represents the time-span between observations, which is assumed to be constant. Assuming the following cyclic boundary conditions $u(-1) = u(d-1)$, $u(0) = u(d)$, and $u(d+1) = u(1)$ with $d \geq 4$, the system is governed by:

$$\frac{du(i)}{dt} = (u(i+1) - u(i-2))u(i-1) - u(i) + F, \quad i = 1, \ldots, d. \tag{4.4}$$

In our experiments, we set $\Delta t = 0.01$, $F = 8$, and the state dimension d is subject to variation. The choice $F = 8$ leads to strongly chaotic turbulence, which hinders predictability in the absence of observations [36]. For the observation process (2.3), we consider both full observations in which $H = I_d$, and partial observations in which only two out of every three state components are observed. The latter setting results in a modified $H \in \mathbb{R}^{\frac{2d}{3} \times d}$ which corresponds to I_d with every third row removed. This observation set-up is motivated by [44, 29], which prove that observing two-out-of-three coordinates of the Lorenz 96 system suffices in order to tame the unpredictability of the system and achieve long-time filter accuracy in a small noise regime. As in subsection 4.1, we examine various choices of initial distribution, dynamics noise covariance, and observation noise covariance. To compare EnKF and REnKF, we make use of the same CI width (4.2) and CI coverage (4.3) metrics as in subsection 4.1. However, since in the nonlinear setting the mean of the filtering distribution is not available in closed form, we replace the metric $\mathsf{E}_{\text{Linear}}$ with

$$\mathsf{E}_{\text{L96}} = \frac{1}{J} \sum_{j=1}^{J} |\widehat{\mu}^{(j)} - u^{(j)}|_2, \tag{4.5}$$

which quantifies the accuracy of the filter as an estimator of the ground-truth state process $\{u^{(j)}\}_{j=1}^{J}$. As before, the metrics we report are averaged over M runs of the algorithms.

In Table 5, we compare the performance of REnKF and EnKF. In the case of full observations, the covariance configuration is outlined in Case A of Table 3, and in the case of partial observations it is outlined in Case C of Table 3. We repeat the experiments with ensembles of size $N = 21$ and $N = 84$, and the metrics are computed over $M = 100$ runs of the algorithms. In Figure 4, we present a single representative simulation of the first component $u(1)$—which is observed—and the third component $u(3)$—which is unobserved—corresponding to a particular choice of parameters in Table 5. Additional experiments in the accompanying Github

Table 4

Effective dimension of initialization and noise covariances used in Figure 3.

State dimension (d)	2	4	8	16	32	64	128	256
$\beta = 0.1$	1.93	3.70	7.02	13.25	24.89	46.64	87.25	163.05
$\beta = 1.0$	1.50	2.08	2.72	3.38	4.06	4.74	5.43	6.12
$\beta = 1.5$	1.35	1.67	1.93	2.12	2.26	2.36	2.44	2.49

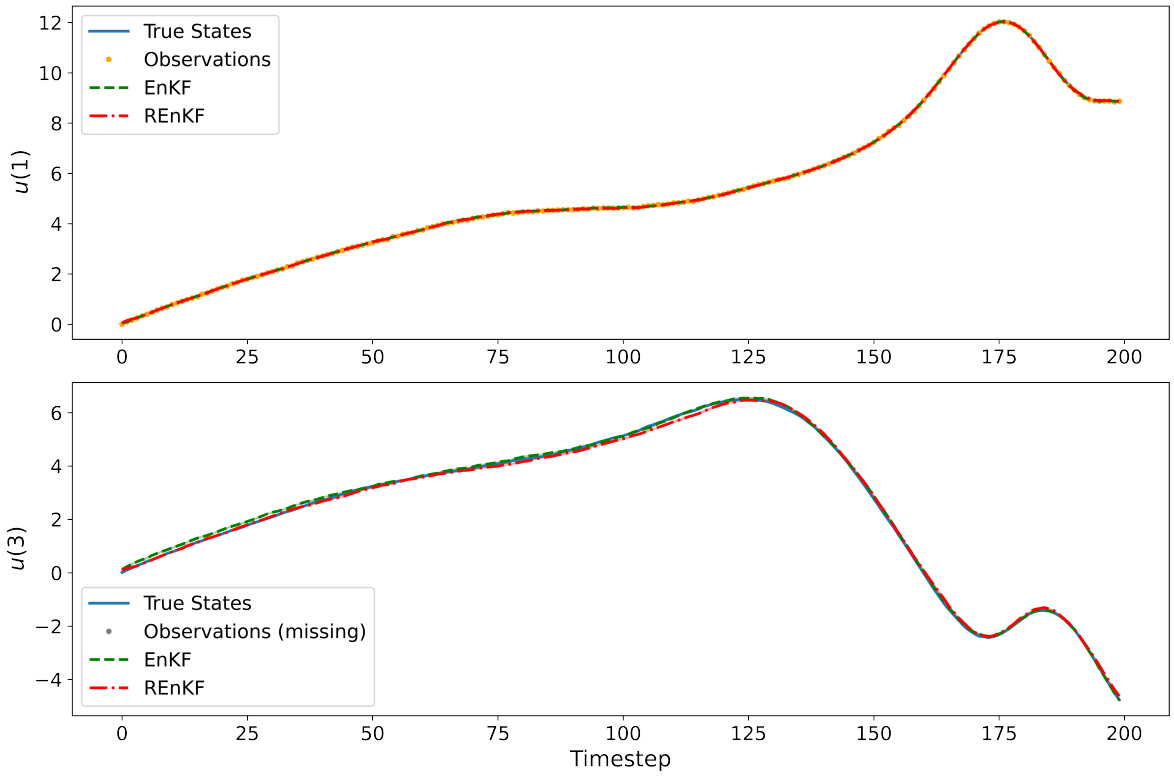

Figure 4. *State estimation of coordinates* $u(1)$ *(observed) and* $u(3)$ *(unobserved) in a partially observed Lorenz* 96 *system with ensemble size* $N = 21$ *and small noise* $\alpha = 10^{-4}$. REnKF *accurately recovers observed and unobserved coordinates of the state.*

repository show that, as the noise level α increases, state estimation remains effective for observed variables but deteriorates for unobserved ones. This behavior explains the larger error for moderate and large noise levels in the partial observation set-up in Table 5.

In Figure 5, we further analyze the effects of varying α (column 1), N (column 2), and d (column 3) on EL96 in both the full observation (row 1) and partial observation (row 2) settings. More precisely, in the first column of Figure 5, α is varied over a grid of 15 evenly spaced values between 10^{-16} and 1 while holding fixed $N = 20$ and $d = 42$. In both full and partial observation settings, we take the initial distribution to have zero mean and covariance $\Sigma^{(0)} = 10^{-8} \times I_{42}$. In the second column of Figure 5, the ensemble size N ranges from 10 to 100, increasing in steps of 10, while fixing $\alpha = 10^{-4}$ and $d = 42$. In both full and partial observation settings, we take the initial distribution to have zero mean and covariance $\Sigma^{(0)} = 1.1\alpha \times I_{42}$. In the third column of Figure 5, the dimension d is varied over the values in $\{6, 18, 30, 42, 54, 66, 78, 90, 102\}$ which are all multiples of 3 to facilitate convenient calculations in the partially observed setting. We fix $N = 20$ and $\alpha = 10^{-4}$ and in both full and partial observation settings, we take the initial distribution to have zero mean and covariance $\Sigma^{(0)} = 1.1\alpha \times I_d$, respectively.

Table 5

Performance metrics for the Lorenz 96 *model with* $d = 42$.

		Full observation			Partial observation		
Ensemble	Metric	Small noise $\alpha = 10^{-4}$	Moderate noise $\alpha = 10^{-2}$	Large noise $\alpha = 10^{-1}$	Small noise $\alpha = 10^{-4}$	Moderate noise $\alpha = 10^{-2}$	Large noise $\alpha = 10^{-1}$
$N = 21$	EnKF Mean error	0.1011	0.9573	3.0231	0.4064	3.3882	10.5921
	REnKFMean error	0.1016	0.9616	3.0335	0.4071	3.3565	10.6379
	EnKF CI width	0.0208	0.2083	0.6586	0.0266	0.2660	0.8412
	REnKFCI width	0.0205	0.2047	0.6475	0.0258	0.2584	0.8167
	EnKF CI coverage (%)	50.24	51.55	51.61	39.62	43.25	43.26
	REnKF CI coverage (%)	49.07	50.34	50.44	38.25	42.04	41.87
$N = 84$	EnKF Mean error	0.0582	0.5682	1.7971	0.2919	2.4181	7.6282
	REnKF Mean error	0.0590	0.5760	1.8218	0.2977	2.5004	7.9011
	EnKF CI width	0.0281	0.2813	0.8895	0.0438	0.4383	1.3861
	REnKF CI width	0.0279	0.2785	0.8806	0.0412	0.4120	1.3033
	EnKF CI coverage (%)	87.96	88.61	88.61	71.47	75.31	75.30
	REnKF CI coverage (%)	86.80	87.52	87.52	69.25	72.54	72.61

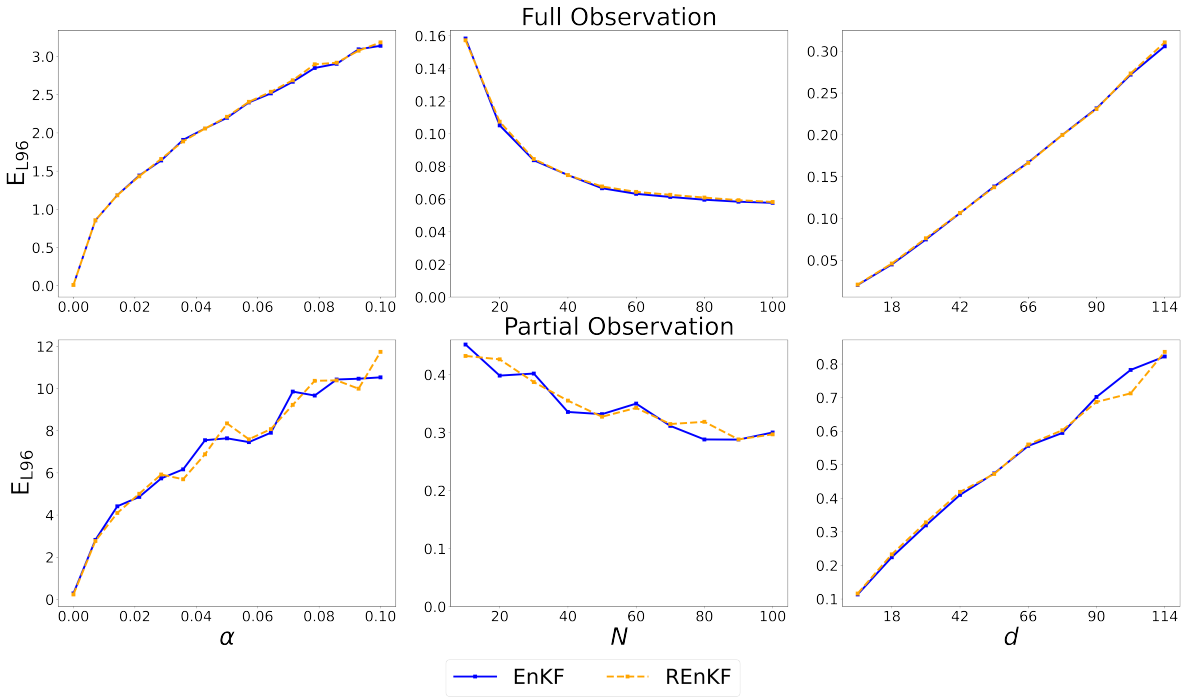

Figure 5. *Effects of* α, N, *and* d *in the Lorenz* 96 *example.*

Our findings, as illustrated in Table 5 and Figure 5, demonstrate that REnKF achieves performance comparable to that of EnKF, even in challenging nonlinear regimes. Notably, for both algorithms we observe a slightly inferior performance with partial observations compared to full observations under identical conditions. Moreover, a consistent trend is noticed in the

dependency of EL96 on the noise level, state dimension, and ensemble size. Notice, however, that the performance of REnKF deteriorates further in non-Gaussian settings with partial observations, large N, and large noise. Such worsened performance may be partly explained by the additional Gaussian assumption tacitly imposed in the resampling step, which further destroys the non-Gaussian structure of the problem for nonlinear forward models. Table 5 further demonstrates that REnKF is as effective as EnKF in the task of uncertainty quantification. Nevertheless, both EnKF and REnKF encounter difficulties in delivering reliable uncertainty quantification, especially in scenarios with partial observation and small ensemble size.

5. Proof of Theorem 3.2. The result will be established by strong induction on the *mean bound* (3.10) and the *covariance bound* (3.11) along with induction on two additional bounds: for any $j = 1, 2, \dots$ and $q \geq 1$

$$\| |\mathrm{Tr}(\widehat{\Sigma}^{(j-1)})| \|_q \leq c_3 r_2(\Sigma^{(0)}), \tag{5.1}$$

$$\| |\widehat{C}^{(j)} - C^{(j)}| \|_q \leq c_4 \left(\sqrt{\frac{r_2(\Sigma^{(0)})}{N}} \vee \sqrt{\frac{r_2(\Xi)}{N}} \vee \sqrt{\frac{r_2(\Gamma)}{N}} \right), \tag{5.2}$$

where c_3 and c_4 are again potentially different universal constants that depend on the same parameters as c_2 in the statement of Theorem 3.2. We will refer to (5.1) as the *covariance trace bound* and to (5.2) as the *forecast covariance bound.* In this section, we require the following additional notation: given two positive sequences $\{a_n\}$ and $\{b_n\}$, the relation $a_n \lesssim b_n$ denotes that $a_n \leq c b_n$ for some constant $c > 0$. If the constant c depends on some quantity τ, then we write $a \lesssim_\tau b$. Throughout, we denote positive universal constants by c, c_1, c_2, c_3, c_4, and the value of a universal constant may differ from line to line. In some cases, the explicit dependence of a universal constant on the parameter τ is indicated by writing $c(\tau)$.

This section is organized as follows. Subsection 5.1 contains preliminary results. We then prove the base case $j = 1$ in subsection 5.2. Finally, in subsection 5.3 we show that the bounds (3.10), (3.11), (5.1), and (5.2) hold for j assuming they hold for all $\ell \leq j - 1$.

5.1. Preliminary results.

Lemma 5.1 (operator norm of covariance). *For any $j \geq 0$, let $\Sigma^{(j)}$ be the analysis covariance at iteration j. Then,*

$$|\Sigma^{(j)}| \leq |A|^{2j} |\Sigma^{(0)}| + |\Xi| \sum_{\ell=0}^{j-1} |A|^{2\ell} \leq c(|A|, |\Xi|, |\Sigma^{(0)}|, j).$$

Proof. By Lemma B.6, $|\mathscr{C}(C)| \leq |C|$, and so

$$\begin{aligned} |\Sigma^{(j)}| = |\mathscr{C}(C^{(j)})| \leq |C^{(j)}| &= |A\Sigma^{(j-1)}A^\top + \Xi| \leq |A|^2 |\Sigma^{(j-1)}| + |\Xi| \\ &\leq |A|^4 |\Sigma^{(j-2)}| + |A|^2 |\Xi| + |\Xi| \leq \cdots \leq |A|^{2j} |\Sigma^{(0)}| + |\Xi| \sum_{\ell=0}^{j-1} |A|^{2\ell}. \end{aligned}$$

■

Lemma 5.2 (trace of offset). *For any $j \geq 1$, we have that*

$$\begin{aligned} Tr(\widehat{O}^{(j)}) \leq |H|^2 |\widehat{\Gamma}^{(j)} - \Gamma| |\Gamma^{-1}|^2 |\widehat{C}^{(j)}| \, Tr(\widehat{C}^{(j)}) \\ + 2(1 + |\widehat{C}^{(j)}| |H|^2 |\Gamma^{-1}|) |\Gamma^{-1}| |\widehat{C}^{(j)}_{u\eta}| |H| \, Tr(\widehat{C}^{(j)}). \end{aligned}$$

Proof. Write $\widehat{O}^{(j)} = \sum_{\ell=1}^{3} \widehat{O}_\ell^{(j)}$ with

$$\begin{aligned}
\widehat{O}_1^{(j)} &:= \mathscr{K}(\widehat{C}^{(j)})(\widehat{\Gamma}^{(j)} - \Gamma)\mathscr{K}^\top(\widehat{C}^{(j)}),\\
\widehat{O}_2^{(j)} &:= (I - \mathscr{K}(\widehat{C}^{(j)})H)\widehat{C}_{u\eta}^{(j)}\mathscr{K}^\top(\widehat{C}^{(j)}),\\
\widehat{O}_3^{(j)} &:= \mathscr{K}(\widehat{C}^{(j)})(\widehat{C}_{u\eta}^{(j)})^\top(I - H^\top\mathscr{K}^\top(\widehat{C}^{(j)})).
\end{aligned}$$

By linearity of the trace, $\mathrm{Tr}(\widehat{O}^{(j)}) = \sum_{\ell=1}^{3}\mathrm{Tr}(\widehat{O}_\ell^{(j)})$. Note first that by Lemma B.6 applied four times

$$\begin{aligned}
\mathrm{Tr}(\widehat{O}_1^{(j)}) &\le |\widehat{\Gamma}^{(j)} - \Gamma|\mathrm{Tr}(\mathscr{K}^\top(\widehat{C}^{(j)})\mathscr{K}(\widehat{C}^{(j)}))\\
&= |\widehat{\Gamma}^{(j)} - \Gamma|\mathrm{Tr}((H\widehat{C}^{(j)}H^\top + \Gamma)^{-1}H(\widehat{C}^{(j)})^\top\widehat{C}^{(j)}H^\top(H\widehat{C}^{(j)}H^\top + \Gamma)^{-1})\\
&\le |\widehat{\Gamma}^{(j)} - \Gamma||(H\widehat{C}^{(j)}H^\top + \Gamma)^{-1}|^2\mathrm{Tr}((\widehat{C}^{(j)})^\top\widehat{C}^{(j)}H^\top H)\\
&\le |H|^2|\widehat{\Gamma}^{(j)} - \Gamma||(H\widehat{C}^{(j)}H^\top + \Gamma)^{-1}|^2|\widehat{C}^{(j)}|\mathrm{Tr}(\widehat{C}^{(j)})\\
&\le |H|^2|\widehat{\Gamma}^{(j)} - \Gamma||\Gamma^{-1}|^2|\widehat{C}^{(j)}|\mathrm{Tr}(\widehat{C}^{(j)}),
\end{aligned}$$

where the final inequality holds since $H\widehat{C}^{(j)}H^\top + \Gamma \succeq \Gamma$ implies that $\Gamma^{-1} \succeq (H\widehat{C}^{(j)}H^\top + \Gamma)^{-1}$. Invoking once more Lemma B.6 repeatedly, we get that

$$\begin{aligned}
\mathrm{Tr}(\widehat{O}_2^{(j)}) &= \mathrm{Tr}((I - \mathscr{K}(\widehat{C}^{(j)})H)\widehat{C}_{u\eta}^{(j)}\mathscr{K}^\top(\widehat{C}^{(j)}))\\
&\le |(I - \mathscr{K}(\widehat{C}^{(j)})H)|\mathrm{Tr}(\widehat{C}_{u\eta}^{(j)}\mathscr{K}^\top(\widehat{C}^{(j)}))\\
&\le |(I - \mathscr{K}(\widehat{C}^{(j)})H)||(H\widehat{C}^{(j)}H^\top + \Gamma)^{-1}|\mathrm{Tr}(H\widehat{C}^{(j)}\widehat{C}_{u\eta}^{(j)})\\
&\le |(I - \mathscr{K}(\widehat{C}^{(j)})H)||(H\widehat{C}^{(j)}H^\top + \Gamma)^{-1}||\widehat{C}_{u\eta}^{(j)}||H|\mathrm{Tr}(\widehat{C}^{(j)})\\
&\le (1 + |\widehat{C}^{(j)}||H|^2|\Gamma^{-1}|)|\Gamma^{-1}||\widehat{C}_{u\eta}^{(j)}||H|\mathrm{Tr}(\widehat{C}^{(j)}),
\end{aligned}$$

where the last inequality holds since, by Lemma B.1,

$$|(I - \mathscr{K}(\widehat{C}^{(j)})H)| \le 1 + |\mathscr{K}(\widehat{C}^{(j)})||H| \le 1 + |\widehat{C}^{(j)}||H|^2|\Gamma^{-1}|.$$

Finally, note that since $\widehat{O}_3^{(j)} = (\widehat{O}_2^{(j)})^\top$, the analysis of $\widehat{O}_3^{(j)}$ follows in similar fashion. ∎

5.2. Base case. In the next four subsections we establish the covariance trace bound (5.1), the forecast covariance bound (5.2), the mean bound (3.10), and the covariance bound (3.11) in the base case $j = 1$.

5.2.1. Covariance trace bound. Since $\widehat{\Sigma}^{(0)} = \Sigma^{(0)}$, we directly obtain that

$$\||\mathrm{Tr}(\widehat{\Sigma}^{(0)})|\|_q = \mathrm{Tr}(\Sigma^{(0)}) = |\Sigma^{(0)}|r_2(\Sigma^{(0)}).$$

5.2.2. Forecast covariance bound. Let $\mathfrak{q} \in \{q, 2q, 4q\}$. It follows by the triangle inequality, Theorem B.5, and Lemma B.7 that

$$\begin{aligned}
\|\|\widehat{C}^{(1)} - C^{(1)}\|\|_{\mathfrak{q}} &\le |A|^2 \|\|S^{(0)} - \Sigma^{(0)}\|\|_{\mathfrak{q}} + \|\|\widehat{\Xi}^{(1)} - \Xi\|\|_{\mathfrak{q}} + 2|A| \|\|\widehat{C}^{(1)}_{u\xi}\|\|_{\mathfrak{q}} \\
&\lesssim_q |A|^2 |\Sigma^{(0)}| \sqrt{\frac{r_2(\Sigma^{(0)})}{N}} + |\Xi| \sqrt{\frac{r_2(\Xi)}{N}} \\
&\quad + 2|A|(|\Sigma^{(0)}| \vee |\Xi|) \left(\sqrt{\frac{r_2(\Sigma^{(0)})}{N}} \vee \sqrt{\frac{r_2(\Xi)}{N}} \right) \\
&\le c(|A|, |\Sigma^{(0)}|, |\Xi|, q) \left(\sqrt{\frac{r_2(\Sigma^{(0)})}{N}} \vee \sqrt{\frac{r_2(\Xi)}{N}} \right). \qquad (5.3)
\end{aligned}$$

5.2.3. Mean bound. By Lemma B.2,

$$\begin{aligned}
\|\|\widehat{\mu}^{(1)} - \mu^{(1)}\|\|_q &= \|\|\mathscr{M}(\widehat{m}^{(1)}, \widehat{C}^{(1)}; y^{(1)}) - \mathscr{M}(m^{(1)}, C^{(1)}; y^{(1)})\|\|_q + \|\|\mathscr{K}(\widehat{C}^{(1)})\bar{\eta}^{(1)}\|\|_q \\
&\le \left\| |\widehat{m}^{(1)} - m^{(1)}| \right\|_q + |H|^2 |\Gamma^{-1}| \|\|\widehat{C}^{(1)}\|\|_{2q} \|\|\widehat{m}^{(1)} - m^{(1)}\|\|_{2q} \\
&\quad + \|\|\widehat{C}^{(1)} - C^{(1)}\|\|_q |H| |\Gamma^{-1}| (1 + |H|^2 |\Gamma^{-1}| |C^{(1)}|) |y^{(1)} - H m^{(1)}| \\
&\quad + \|\|\mathscr{K}(\widehat{C}^{(1)})\bar{\eta}^{(1)}\|\|_q.
\end{aligned}$$

The mean bound (3.10) with $j = 1$ is then a direct consequence of the bounds that we now establish on $\|\|\widehat{m}^{(1)} - m^{(1)}\|\|_{\mathfrak{q}}$ for $\mathfrak{q} \in \{q, 2q\}$ and on $\|\|\mathscr{K}(\widehat{C}^{(1)})\bar{\eta}^{(1)}\|\|_q$.

Controlling $\|\|\widehat{m}^{(1)} - m^{(1)}\|\|_{\mathfrak{q}}$ for $\mathfrak{q} \in \{q, 2q\}$. It follows by the triangle inequality and Lemma B.9 applied twice that

$$\begin{aligned}
\|\|\widehat{m}^{(1)} - m^{(1)}\|\|_{\mathfrak{q}} &\le |A| \|\|\bar{u}^{(0)} - \mu^{(0)}\|\|_{\mathfrak{q}} + \|\|\bar{\xi}^{(1)}\|\|_{\mathfrak{q}} \lesssim_q |A| |\Sigma^{(0)}| \sqrt{\frac{r_2(\Sigma^{(0)})}{N}} + |\Xi| \sqrt{\frac{r_2(\Xi)}{N}} \\
&\le c(|A|, |\Sigma^{(0)}|, |\Xi|, q) \left(\sqrt{\frac{r_2(\Sigma^{(0)})}{N}} \vee \sqrt{\frac{r_2(\Xi)}{N}} \right).
\end{aligned}$$

Controlling $\|\|\mathscr{K}(\widehat{C}^{(1)})\bar{\eta}^{(1)}\|\|_q$. By Cauchy–Schwarz

$$\|\|\mathscr{K}(\widehat{C}^{(1)})\bar{\eta}^{(1)}\|\|_q \le \|\|\mathscr{K}(\widehat{C}^{(1)})\|\|_{2q} \|\|\bar{\eta}^{(1)}\|\|_{2q}.$$

We bound each term in turn. By Lemma B.9, $\|\|\bar{\eta}^{(1)}\|\|_{2q} \lesssim_q \sqrt{\mathrm{Tr}(\Gamma)/N} = \sqrt{|\Gamma| r_2(\Gamma)/N}$, and by Lemma B.1 and the forecast covariance bound (5.3),

$$\begin{aligned}
\|\|\mathscr{K}(\widehat{C}^{(1)})\|\|_{2q} &\le |H| |\Gamma^{-1}| \|\|\widehat{C}^{(1)}\|\|_{2q} \le |H| |\Gamma^{-1}| \left(\|\|\widehat{C}^{(1)} - C^{(1)}\|\|_{2q} + |C^{(1)}| \right) \\
&\lesssim |H| |\Gamma^{-1}| |C^{(1)}| c(|A|, |\Sigma^{(0)}|, |\Xi|, q) \left(1 \vee \sqrt{\frac{r_2(\Sigma^{(0)})}{N}} \vee \sqrt{\frac{r_2(\Xi)}{N}} \right).
\end{aligned}$$

Therefore,

$$\|\|\mathscr{K}(\widehat{C}^{(1)})\bar{\eta}^{(1)}\|\|_q \le c(|H|, |\Sigma^{(0)}|, |\Xi|, |H|, |\Gamma^{-1}|, |\Gamma|, q) \left(\sqrt{\frac{r_2(\Sigma^{(0)})}{N}} \vee \sqrt{\frac{r_2(\Xi)}{N}} \vee \sqrt{\frac{r_2(\Gamma)}{N}} \right).$$

5.2.4. Covariance bound. From Lemma B.3 and the forecast covariance bound derived in subsection 5.2.2, we have

$$
\begin{aligned}
|||\widehat{\Sigma}^{(1)} - \Sigma^{(1)}|||_q &\le |||\mathscr{C}(\widehat{C}^{(1)}) - \mathscr{C}(C^{(1)})|||_q + |||\widehat{O}|||_q \\
&\le |||\widehat{C}^{(1)} - C^{(1)}|||_q (1 + |H|^2|\Gamma^{-1}||C^{(1)}|) \\
&\quad + (|H|^2|\Gamma^{-1}| + |H|^4|\Gamma^{-1}|^2|C^{(1)}|)|||\widehat{C}^{(1)}|||_{2q}|||\widehat{C}^{(1)} - C^{(1)}|||_{2q} + |||\widehat{O}^{(1)}|||_q \\
&\le c(|A|, |H|, |\Gamma^{-1}|, |\Sigma^{(0)}|, |\Xi|, q) \left(\sqrt{\frac{r_2(\Sigma^{(0)})}{N}} \vee \sqrt{\frac{r_2(\Xi)}{N}} \right) + |||\widehat{O}^{(1)}|||_q.
\end{aligned}
$$

To derive the covariance bound (3.11), we need to control the offset term $|||\widehat{O}^{(1)}|||_q$. First, using the triangle inequality we write

$$
\begin{aligned}
|||\widehat{O}^{(1)}|||_q &\le |||\mathscr{K}(\widehat{C}^{(1)})(\widehat{\Gamma}^{(1)} - \Gamma)\mathscr{K}^\top(\widehat{C}^{(1)})|||_q + |||(I - \mathscr{K}(\widehat{C}^{(1)})H)\widehat{C}^{(1)}_{u\eta}\mathscr{K}^\top(\widehat{C}^{(1)})|||_q \\
&\quad + |||\mathscr{K}(\widehat{C}^{(1)})(\widehat{C}^{(1)}_{u\eta})^\top(I - H^\top\mathscr{K}^\top(\widehat{C}^{(1)}))|||_q \\
&=: |||\widehat{O}^{(1)}_1|||_q + |||\widehat{O}^{(1)}_2|||_q + |||\widehat{O}^{(1)}_3|||_q.
\end{aligned}
$$

We next bound each term in turn.

Controlling $|||\widehat{O}^{(1)}_1|||_q$. By Lemma B.1, Theorem B.5, and the forecast covariance bound (5.3), it holds that

$$
\begin{aligned}
&|||\mathscr{K}(\widehat{C}^{(1)})(\widehat{\Gamma}^{(1)} - \Gamma)\mathscr{K}^\top(\widehat{C}^{(1)})|||_q \\
&\le |||\mathscr{K}(\widehat{C}^{(1)})|||^2_{4q}|||\widehat{\Gamma}^{(1)} - \Gamma|||_{2q} \\
&\le |H|^2|\Gamma^{-1}|^2|||\widehat{C}^{(1)}|||^2_{4q}|||\widehat{\Gamma}^{(1)} - \Gamma|||_{2q} \\
&\lesssim_q |H|^2|\Gamma^{-1}|^2|\Gamma|\sqrt{\frac{r_2(\Gamma)}{N}} \left(1 \vee \sqrt{\frac{r_2(\Sigma^{(0)})}{N}} \vee \sqrt{\frac{r_2(\Xi)}{N}} \vee \sqrt{\frac{r_2(\Gamma)}{N}}\right)^2 \\
&\le c(|H|, |\Gamma|, |\Gamma^{-1}|, |\Xi|, |\Sigma^{(0)}|, q)\sqrt{\frac{r_2(\Gamma)}{N}},
\end{aligned}
$$

where the last inequality uses the fact that $N \ge r_2(\Sigma^{(0)}) \vee r_2(\Xi) \vee r_2(\Gamma)$.

Controlling $|||\widehat{O}^{(1)}_2|||_q$. By Lemmas B.1 and B.7 and the forecast covariance bound (5.3), we get

$$
\begin{aligned}
&|||(I - \mathscr{K}(\widehat{C}^{(1)})H)\widehat{C}^{(1)}_{u\eta}\mathscr{K}^\top(\widehat{C}^{(1)})|||_q \le |||\mathscr{K}(\widehat{C}^{(1)})||I - \mathscr{K}(\widehat{C}^{(1)})H||\widehat{C}^{(1)}_{u\eta}|||_q \\
&\le |||\mathscr{K}(\widehat{C}^{(1)})|(1 + |\mathscr{K}(\widehat{C}^{(1)})||H|)|\widehat{C}^{(1)}_{u\eta}|||_q \\
&\le |||\widehat{C}^{(1)}_{u\eta}|||_{2q} \left(|H||\Gamma^{-1}|||\widehat{C}^{(1)}|||_{2q} + |H|^3|\Gamma^{-1}|^2|||\widehat{C}^{(1)}|^2\|_{2q}\right) \\
&\le c(|A|, |\Sigma^{(0)}|, |\Xi|, q) \left(1 \vee \sqrt{\frac{r_2(\Sigma^{(0)})}{N}} \vee \sqrt{\frac{r_2(\Xi)}{N}} \vee \sqrt{\frac{r_2(\Gamma)}{N}}\right) \\
&\quad \times \left(|H|^3|\Gamma^{-1}| + |H|^7|\Gamma^{-1}|^2\right)(|\Sigma^{(0)}| \vee |\Gamma|) \left(\sqrt{\frac{r_2(\Sigma^{(0)})}{N}} \vee \sqrt{\frac{r_2(\Xi)}{N}} \vee \sqrt{\frac{r_2(\Gamma)}{N}}\right) \\
&\le c(|A|, |H|, |\Gamma^{-1}|, |\Gamma|, |\Sigma^{(0)}|, |\Xi|, q) \left(\sqrt{\frac{r_2(\Sigma^{(0)})}{N}} \vee \sqrt{\frac{r_2(\Xi)}{N}} \vee \sqrt{\frac{r_2(\Gamma)}{N}}\right).
\end{aligned}
$$

Controlling $\||\widehat{O}_3^{(1)}|\|_q$. Note that $\||\widehat{O}_3^{(1)}|\|_q = \||\widehat{O}_2^{(1)}|\|_q$.

5.3. Induction step. In this subsection, to reduce notation we write $\Omega = \sqrt{\frac{r_2(\Sigma^{(0)})}{N}} \vee \sqrt{\frac{r_2(\Xi)}{N}} \vee \sqrt{\frac{r_2(\Gamma)}{N}}$. Throughout, we work under the *inductive hypothesis* that, for all $\ell \le j-1$, it holds that

$$\begin{aligned} &\||\mathrm{Tr}(\widehat{\Sigma}^{(\ell-1)})|\|_q \le c_1 r_2(\Sigma^{(0)}), \qquad &&\||\widehat{C}^{(\ell)} - C^{(\ell)}|\|_{\mathfrak{q}} \le c_2\Omega, \\ &\||\widehat{\mu}^{(\ell)} - \mu^{(\ell)}|\|_q \le c_3\Omega, &&\||\widehat{\Sigma}^{(\ell)} - \Sigma^{(\ell)}|\|_q \le c_4\Omega, \end{aligned} \tag{5.4}$$

where c_1, c_2, c_3, and c_4 are constants depending on $|\Sigma^{(0)}|, |A|, |H|, |\Gamma^{-1}|, |\Gamma|, |\Xi|, q$, and j, and c_3 additionally depends on $\{|y^{(i)} - Hm^{(i)}|\}_{i\le \ell-1}$. For the remainder of the proof, c and c' denote constants that depend on $|\Sigma^{(0)}|, |A|, |H|, |\Gamma^{-1}|, |\Gamma|, |\Xi|, q$, and j, and c' additionally depends on $\{|y^{(i)} - Hm^{(i)}|\}_{i\le \ell-1}$ and are potentially different from line to line. In the next four subsections we show that, under the inductive hypothesis, the four bounds in (5.4) also hold for $\ell = j$. Throughout, we use without further notice that $|\Sigma^{(\ell)}| \lesssim c(|A|, |\Xi|, |\Sigma^{(0)}|, \ell)$, which was proved in Lemma 5.1.

5.3.1. Covariance trace bound. By Lemma B.3, $\mathrm{Tr}(\mathscr{C}(\widehat{C}^{(j-1)})) \le \mathrm{Tr}(\widehat{C}^{(j-1)})$ follows from the fact that $\mathscr{C}(\widehat{C}^{(j-1)}) \preceq \widehat{C}^{(j-1)}$, and so

$$\|\mathrm{Tr}(\widehat{\Sigma}^{(j-1)})\|_q \le \|\mathrm{Tr}(\mathscr{C}(\widehat{C}^{(j-1)}))\|_q + \|\mathrm{Tr}(\widehat{O}^{(j-1)})\|_q \le \|\mathrm{Tr}(\widehat{C}^{(j-1)})\|_q + \|\mathrm{Tr}(\widehat{O}^{(j-1)})\|_q.$$

We will show that both of the terms on the right-hand side are bounded above by a constant times $r_2(\Sigma^{(0)})$.

Controlling $\|\,Tr(\widehat{C}^{(j-1)})\|_q$. Noting first that

$$\begin{aligned} \mathbb{E}\left[\widehat{C}^{(j-1)} \middle| \widehat{\mu}^{(j-2)}, \widehat{\Sigma}^{(j-2)}\right] &= \mathbb{E}\left[AS^{(j-2)}A^\top + \widehat{\Xi}^{(j-1)} + A\widehat{C}_{u\xi}^{(j-1)} + \widehat{C}_{\xi u}^{(j-1)}A^\top \middle| \widehat{\mu}^{(j-2)}, \widehat{\Sigma}^{(j-2)}\right] \\ &= \mathbb{E}\left[AS^{(j-2)}A^\top \middle| \widehat{\mu}^{(j-2)}, \widehat{\Sigma}^{(j-2)}\right] = A\widehat{\Sigma}^{(j-2)}A^\top, \end{aligned}$$

and by Lemma B.6, it holds almost surely that

$$\mathrm{Tr}(A\widehat{\Sigma}^{(j-2)}A^\top) \le |A|^2 \mathrm{Tr}(\widehat{\Sigma}^{(j-2)}) = |A|^2 |\widehat{\Sigma}^{(j-2)}| r_2(\widehat{\Sigma}^{(j-2)}).$$

Then, by iterated expectations and Lemma B.8, we have

$$\begin{aligned} \mathbb{E}\left[\mathrm{Tr}(\widehat{C}^{(j-1)})\right]^q &= \mathbb{E}\left[\mathbb{E}\left[\left(\mathrm{Tr}(\widehat{C}^{(j-1)})\right)^q \middle| \widehat{\mu}^{(j-2)}, \widehat{\Sigma}^{(j-2)}\right]\right] \\ &\lesssim_q \mathbb{E}\left[\mathbb{E}\left[\left(\mathrm{Tr}\left(\widehat{C}^{(j-1)} - \mathbb{E}[\widehat{C}^{(j-1)}|\widehat{\mu}^{(j-2)}, \widehat{\Sigma}^{(j-2)}]\right)\right)^q \middle| \widehat{\mu}^{(j-2)}, \widehat{\Sigma}^{(j-2)}\right]\right] \\ &\quad + \mathbb{E}\left[\left(\mathrm{Tr}\left(\mathbb{E}[\widehat{C}^{(j-1)}|\widehat{\mu}^{(j-2)}, \widehat{\Sigma}^{(j-2)}]\right)\right)^q\right] \\ &\lesssim \mathbb{E}\left[\left(\frac{\mathrm{Tr}(\mathbb{E}[\widehat{C}^{(j-1)}|\widehat{\mu}^{(j-2)}, \widehat{\Sigma}^{(j-2)}])}{\sqrt{N}}\right)^q\right] + \mathbb{E}\left[\left(\mathrm{Tr}\left(\mathbb{E}\left[\widehat{C}^{(j-1)}|\widehat{\mu}^{(j-2)}, \widehat{\Sigma}^{(j-2)}\right]\right)\right)^q\right] \\ &\le \frac{|A|^{2q}}{N^{q/2}}\mathbb{E}\left[\left(\mathrm{Tr}(\widehat{\Sigma}^{(j-2)})\right)^q\right] + |A|^{2q}\mathbb{E}\left[\left(\mathrm{Tr}(\widehat{\Sigma}^{(j-2)})\right)^q\right] \\ &\lesssim \frac{|A|^{2q}}{N^{q/2}} c r_2(\Sigma^{(0)})^q + |A|^{2q} c r_2(\Sigma^{(0)})^q \le c r_2(\Sigma^{(0)})^q, \end{aligned}$$

where the second to last inequality holds by the inductive hypothesis (5.4).

Controlling $\|\widehat{O}^{(j-1)}\|_q$. By definition, we have

$$\begin{aligned}\widehat{O}^{(j-1)} &= \mathscr{K}(\widehat{C}^{(j-1)})(\widehat{\Gamma}^{(j-1)}-\Gamma)\mathscr{K}^\top(\widehat{C}^{(j-1)}) + (I-\mathscr{K}(\widehat{C}^{(j-1)})H)\widehat{C}^{(j-1)}_{u\eta}\mathscr{K}^\top(\widehat{C}^{(j-1)}) \\ &\quad + \mathscr{K}(\widehat{C}^{(j-1)})(\widehat{C}^{(j-1)}_{u\eta})^\top(I-H^\top\mathscr{K}^\top(\widehat{C}^{(j-1)})) \\ &=: \widehat{O}_1^{(j-1)} + \widehat{O}_2^{(j-1)} + \widehat{O}_3^{(j-1)}.\end{aligned}$$

Therefore, $\|\mathrm{Tr}(\widehat{O}^{(j-1)})\|_q \le \|\mathrm{Tr}(\widehat{O}_1^{(j-1)})\|_q + \|\mathrm{Tr}(\widehat{O}_2^{(j-1)})\|_q + \|\mathrm{Tr}(\widehat{O}_3^{(j-1)})\|_q$.

Controlling $\|\,Tr(\widehat{O}_1^{(j-1)})\|_q$. By Lemma 5.2,

$$\mathrm{Tr}(\widehat{O}_1^{(j-1)}) \le |H|^2|\widehat{\Gamma}^{(j-1)}-\Gamma||\Gamma^{-1}|^2|\widehat{C}^{(j-1)}|\mathrm{Tr}(\widehat{C}^{(j-1)}),$$

and so

$$\begin{aligned}\|\mathrm{Tr}(\widehat{O}_1^{(j-1)})\|_q &\le |H|^2|\Gamma^{-1}|^2\||\widehat{\Gamma}^{(j-1)}-\Gamma||\widehat{C}^{(j-1)}|\mathrm{Tr}(\widehat{C}^{(j-1)})\|_q \\ &\le |H|^2|\Gamma^{-1}|^2\||\widehat{\Gamma}^{(j-1)}-\Gamma|\|_{2q}\||\widehat{C}^{(j-1)}|\|_{4q}\|\mathrm{Tr}(\widehat{C}^{(j-1)})\|_{4q}.\end{aligned}$$

By Theorem B.5, $\||\widehat{\Gamma}^{(j-1)}-\Gamma|\|_{2q} \lesssim_q |\Gamma|\sqrt{\frac{r_2(\Gamma)}{N}}$, and by the inductive hypothesis (5.4) and the fact that $|C^{(j-1)}| \le |A|^2|\Sigma^{(j-2)}| + |\Xi|$, we have

$$\||\widehat{C}^{(j-1)}|\|_{4q} \le |C^{(j-1)}| + \||\widehat{C}^{(j-1)} - C^{(j-1)}|\|_{4q} \le c\,(1\vee\Omega)\,.$$

We have also previously shown that $\|\mathrm{Tr}(\widehat{C}^{(j-1)})\|_{4q} \lesssim r_2(\Sigma^{(0)})$. Noting that $(1\vee\Omega)\,r_2(\Sigma^{(0)}) \lesssim r_2(\Sigma^{(0)})$, we get that $\|\mathrm{Tr}(\widehat{O}_1^{(j-1)})\|_q \le c r_2(\Sigma^{(0)})$.

Controlling $\||\,Tr(\widehat{O}_2^{(j-1)})|\|_q$. By Lemma 5.2,

$$\mathrm{Tr}(\widehat{O}_2^{(j-1)}) \le (1+|\widehat{C}^{(j-1)}||H|^2|\Gamma^{-1}|)|\Gamma^{-1}||H||\widehat{C}^{(j-1)}_{u\eta}|\mathrm{Tr}(\widehat{C}^{(j-1)}).$$

Therefore,

$$\||\mathrm{Tr}(\widehat{O}_2^{(j-1)})|\|_q \le (1+\||\widehat{C}^{(j-1)}|\|_{2q}|H|^2|\Gamma^{-1}|)|\Gamma^{-1}||H|\||\widehat{C}^{(j-1)}_{u\eta}|\|_{4q}\|\mathrm{Tr}(\widehat{C}^{(j-1)})\|_{4q}.$$

By iterated expectation and Lemmas B.7 and B.8 we have, for $\mathfrak{q} \in \{q, 2q, 4q\}$,

$$\begin{aligned}\||\widehat{C}^{(j-1)}_{u\eta}|\|_{\mathfrak{q}} &= \mathbb{E}\left[\mathbb{E}\left[|\widehat{C}^{(j-1)}_{u\eta}|^{\mathfrak{q}}|\widehat{\mu}^{(j-2)},\widehat{\Sigma}^{(j-2)}\right]\right]^{1/\mathfrak{q}} \\ &\lesssim_q \left\|(|\widehat{\Sigma}^{(j-2)}|\vee|\Gamma|)\left(\sqrt{\frac{r_2(\widehat{\Sigma}^{(j-2)})}{N}}\vee\sqrt{\frac{r_2(\Gamma)}{N}}\right)\right\|_{\mathfrak{q}} \\ &\le (\||\widehat{\Sigma}^{(j-2)}|\|_{2\mathfrak{q}}\vee|\Gamma|)\left(\left\|\sqrt{\frac{r_2(\widehat{\Sigma}^{(j-2)})}{N}}\right\|_{2\mathfrak{q}}\vee\sqrt{\frac{r_2(\Gamma)}{N}}\right). \qquad (5.5)\end{aligned}$$

By the triangle inequality and the inductive hypothesis (5.4), it follows that

$$\||\widehat{\Sigma}^{(j-2)}|\|_{2\mathfrak{q}} \le \|\widehat{\Sigma}^{(j-2)} - \Sigma^{(j-2)}\|_{2\mathfrak{q}} + |\Sigma^{(j-2)}| \le c\,(1\vee\Omega)\,,$$

and also that

$$\left\| \sqrt{\frac{r_2(\widehat{\Sigma}^{(j-2)})}{N}} \right\|_{2\mathfrak{q}}^{2\mathfrak{q}} = \mathbb{E}\left[\left(\frac{r_2(\widehat{\Sigma}^{(j-2)})}{N} \right)^{\mathfrak{q}} \right] \lesssim \mathbb{E}\left[\left(\frac{\mathrm{Tr}(\widehat{\Sigma}^{(j-2)})}{N} \right)^{\mathfrak{q}} \right] \leq cN^{-\mathfrak{q}} r_2(\Sigma^{(0)})^{\mathfrak{q}}.$$

Using identical arguments to those used to control $\|\|\mathrm{Tr}(\widehat{O}_1^{(j-1)})\|\|_q$, we have that

$$\|\mathrm{Tr}(\widehat{O}_2^{(j-1)})\|_q \leq c r_2(\Sigma^{(0)}).$$

Controlling $\|Tr(\widehat{O}_3^{(j-1)})\|_q$. Note that $\|\mathrm{Tr}(\widehat{O}_2^{(j-1)})\|_q = \|\mathrm{Tr}(\widehat{O}_2^{(j-1)})\|_q$.

5.3.2. Forecast covariance bound. Let $\mathfrak{q} \in \{q, 2q, 4q\}$. By the triangle inequality, the inductive hypothesis (5.4), and Theorem B.5, we have

$$\begin{aligned}
\|\|\widehat{C}^{(j)} - C^{(j)}\|\|_{\mathfrak{q}} &\leq |A|^2 \|\|S^{(j-1)} - \Sigma^{(j-1)}\|\|_{\mathfrak{q}} + \|\|\widehat{\Xi}^{(j)} - \Xi\|\|_{\mathfrak{q}} + 2|A|\|\|\widehat{C}_{u\xi}^{(j)}\|\|_{\mathfrak{q}} \\
&\leq |A|^2 \left(\|\|S^{(j-1)} - \widehat{\Sigma}^{(j-1)}\|\|_{\mathfrak{q}} + \|\|\widehat{\Sigma}^{(j-1)} - \Sigma^{(j-1)}\|\|_{\mathfrak{q}} \right) + \|\|\widehat{\Xi}^{(j)} - \Xi\|\|_{\mathfrak{q}} + 2|A|\|\|\widehat{C}_{u\xi}^{(j)}\|\|_{\mathfrak{q}} \\
&\leq c|A|^2 \left(\|\|S^{(j-1)} - \widehat{\Sigma}^{(j-1)}\|\|_{\mathfrak{q}} + \Omega \right) + |\Xi| \left(1 \vee \sqrt{\frac{r_2(\Xi)}{N}} \right) + 2|A|\|\|\widehat{C}_{u\xi}^{(j)}\|\|_{\mathfrak{q}}.
\end{aligned}$$

The forecast covariance bound (5.2) is then a direct consequence of the bounds that we now establish on $\|\|\widehat{C}_{u\xi}^{(j)}\|\|_{\mathfrak{q}}$ and $\|\|S^{(j-1)} - \widehat{\Sigma}^{(j-1)}\|\|_{\mathfrak{q}}$.

Controlling $\|\|\widehat{C}_{u\xi}^{(j)}\|\|_{\mathfrak{q}}$. By an identical analysis to the one used in bounding $\|\|\widehat{C}_{u\eta}^{(j)}\|\|_{\mathfrak{q}}$ in (5.5), we have that

$$\|\|\widehat{C}_{u\xi}^{(j)}\|\|_{\mathfrak{q}} \lesssim c\Omega. \tag{5.6}$$

Controlling $\|\|S^{(j-1)} - \widehat{\Sigma}^{(j-1)}\|\|_{\mathfrak{q}}$. By iterated expectations and Theorem B.5, we have

$$\begin{aligned}
\|\|S^{(j-1)} - \widehat{\Sigma}^{(j-1)}\|\|_{\mathfrak{q}}^{\mathfrak{q}} &= \mathbb{E}\left[|S^{(j-1)} - \widehat{\Sigma}^{(j-1)}|^{\mathfrak{q}} \right] = \mathbb{E}\left[\mathbb{E}\left[|S^{(j-1)} - \widehat{\Sigma}^{(j-1)}|^{\mathfrak{q}} \middle| \widehat{\mu}^{(j-1)}, \widehat{\Sigma}^{(j-1)} \right] \right] \\
&\lesssim_q \mathbb{E}\left[|\widehat{\Sigma}^{(j-1)}|^{\mathfrak{q}} \left(\frac{r_2(\widehat{\Sigma}^{(j-1)})}{N} \right)^{\mathfrak{q}/2} \right] = \mathbb{E}\left[|\widehat{\Sigma}^{(j-1)}|^{\mathfrak{q}/2} \left(\frac{\mathrm{Tr}(\widehat{\Sigma}^{(j-1)})}{N} \right)^{\mathfrak{q}/2} \right] \\
&\leq \sqrt{\mathbb{E}\,|\widehat{\Sigma}^{(j-1)}|^{\mathfrak{q}}} \sqrt{\mathbb{E}\left[\frac{\mathrm{Tr}(\widehat{\Sigma}^{(j-1)})}{N} \right]^{\mathfrak{q}}}.
\end{aligned}$$

By the covariance trace bound proved in subsection 5.3.1 and the inductive hypothesis (5.4), we then have that $\|\|S^{(j-1)} - \widehat{\Sigma}^{(j-1)}\|\|_{\mathfrak{q}} \lesssim c\Omega$.

5.3.3. Mean bound. By Lemma B.2, we have

$$\begin{aligned}
\|\|\widehat{\mu}^{(j)} - \mu^{(j)}\|\|_q &= \|\|\mathscr{M}(\widehat{m}^{(j)}, \widehat{C}^{(j)}; y^{(j)}) - \mathscr{M}(m^{(j)}, C^{(j)}; y^{(j)})\|\|_q + \|\|\mathscr{K}(\widehat{C}^{(j)})\bar{\eta}^{(j)}\|\|_q \\
&\leq \left\| |\widehat{m}^{(j)} - m^{(j)}| \right\|_q + |H|^2 |\Gamma^{-1}| \|\|\widehat{C}^{(j)}\|\|_{2q} \|\|\widehat{m}^{(j)} - m^{(j)}\|\|_{2q} \\
&\quad + \|\|\widehat{C}^{(j)} - C^{(j)}\|\|_q |H| |\Gamma^{-1}| (1 + |H|^2 |\Gamma^{-1}| |C^{(j)}|) |y^{(j)} - Hm^{(j)}| \\
&\quad + \|\|\mathscr{K}(\widehat{C}^{(j)})\bar{\eta}^{(j)}\|\|_q.
\end{aligned}$$

The induction step for the mean bound (3.10) is then a direct consequence of the bounds that we now establish on $|||\widehat{m}^{(j)} - m^{(j)}|||_{\mathfrak{q}}$ for $\mathfrak{q} \in \{q, 2q\}$ and on $|||\mathscr{K}(\widehat{C}^{(j)})\bar{\eta}^{(j)}|||_q$.

Controlling $|||\widehat{m}^{(j)} - m^{(j)}|||_{\mathfrak{q}}$ for $\mathfrak{q} \in \{q, 2q\}$. It follows by the triangle inequality and the inductive hypothesis (5.4) that

$$\begin{aligned}|||\widehat{m}^{(j)} - m^{(j)}|||_{\mathfrak{q}} &\le |A|\,|||\bar{u}^{(j-1)} - \mu^{(j-1)}|||_{\mathfrak{q}} + |||\bar{\xi}^{(j)}|||_{\mathfrak{q}} \\ &\le |A|\left(|||\bar{u}^{(j-1)} - \widehat{\mu}^{(j-1)}|||_{\mathfrak{q}} + |||\widehat{\mu}^{(j-1)} - \mu^{(j-1)}|||_{\mathfrak{q}}\right) + |||\bar{\xi}^{(j)}|||_{\mathfrak{q}} \\ &\le c'|A|\left(|||\bar{u}^{(j-1)} - \widehat{\mu}^{(j-1)}|||_{\mathfrak{q}} + \Omega\right) + c(q)|\Xi|\sqrt{\frac{r_2(\Xi)}{N}}.\end{aligned}$$

By iterated expectations, Lemma B.9, and the covariance trace bound proved in subsection 5.3.1, it follows that

$$\begin{aligned}|||\bar{u}^{(j-1)} - \widehat{\mu}^{(j-1)}|||_{\mathfrak{q}}^{\mathfrak{q}} = \mathbb{E}[|\bar{u}^{(j-1)} - \widehat{\mu}^{(j-1)}|^{\mathfrak{q}}] &= \mathbb{E}\left[\mathbb{E}\left[|\bar{u}^{(j-1)} - \widehat{\mu}^{(j-1)}|^{\mathfrak{q}} \,\middle|\, \widehat{\mu}^{(j-1)}, \widehat{\Sigma}^{(j-1)}\right]\right] \\ &\lesssim_q \mathbb{E}\left[\left(\frac{\mathrm{Tr}(\widehat{\Sigma}^{(j-1)})}{N}\right)^{\mathfrak{q}/2}\right] \le c\left(\frac{r_2(\Sigma^{(0)})}{N}\right)^{\mathfrak{q}/2},\end{aligned}$$

and so $|||\widehat{m}^{(j)} - m^{(j)}|||_{\mathfrak{q}} \lesssim c'\Omega$.

Controlling $|||\mathscr{K}(\widehat{C}^{(j)})\bar{\eta}^{(j)}|||_q$. Note first that $|||\mathscr{K}(\widehat{C}^{(j)})\bar{\eta}^{(j)}|||_q \le |||\mathscr{K}(\widehat{C}^{(j)})|||_{2q}|||\bar{\eta}^{(j)}|||_{2q}$. We then have by Lemma B.9, $|||\bar{\eta}^{(j)}|||_{2q} \lesssim_q \sqrt{\frac{\mathrm{Tr}(\Gamma)}{N}} = \sqrt{|\Gamma|\frac{r_2(\Gamma)}{N}}$, and by Lemma B.1 and the forecast covariance bound established in subsection 5.3.2, we have

$$\begin{aligned}|||\mathscr{K}(\widehat{C}^{(j)})|||_{2q} \le |H||\Gamma^{-1}|\,|||\widehat{C}^{(j)}|||_{2q} &\le |H||\Gamma^{-1}|\left(|||\widehat{C}^{(j)} - C^{(j)}|||_{2q} + |C^{(j)}|\right) \\ &\le |H||\Gamma^{-1}|c\,(1 \vee \Omega)\,.\end{aligned}$$

Therefore, $|||\mathscr{K}(\widehat{C}^{(j)})\bar{\eta}^{(j)}|||_q \le c\Omega$.

5.3.4. Covariance bound.

By Lemma B.3, we have

$$\begin{aligned}|||\widehat{\Sigma}^{(j)} - \Sigma^{(j)}|||_q &\le |||\mathscr{C}(\widehat{C}^{(j)}) - \mathscr{C}(C^{(j)})|||_q + |||\widehat{O}^{(j)}|||_q \\ &\le |||\widehat{C}^{(j)} - C^{(j)}|||_q(1 + |A|^2|\Gamma^{-1}||C^{(j)}|) \\ &\quad + (|A|^2|\Gamma^{-1}| + |A|^4|\Gamma^{-1}|^2|C^{(j)}|)|||\widehat{C}^{(j)}|||_{2q}|||\widehat{C}^{(j)} - C^{(j)}|||_{2q} + |||\widehat{O}^{(j)}|||_q.\end{aligned}$$

The induction step for the forecast covariance has been proved in subsection 5.3.2, and so in order to show the induction step for the covariance bound we need only control the offset term. First, using the triangle inequality, we write

$$\begin{aligned}|||\widehat{O}^{(j)}|||_q \le |||\mathscr{K}(\widehat{C}^{(j)})(\widehat{\Gamma}^{(j)} - \Gamma)\mathscr{K}^\top(\widehat{C}^{(j)})|||_q + |||(I - \mathscr{K}(\widehat{C}^{(j)})H)\widehat{C}^{(j)}_{u\eta}\mathscr{K}^\top(\widehat{C}^{(j)})|||_q \\ + |||\mathscr{K}(\widehat{C}^{(j)})(\widehat{C}^{(j)}_{u\eta})^\top(I - H^\top\mathscr{K}^\top(\widehat{C}^{(j)}))|||_q = \|\widehat{O}_1^{(j)}\|_q + \|\widehat{O}_2^{(j)}\|_q + \|\widehat{O}_3^{(j)}\|_q.\end{aligned}$$

We next bound each term in turn.

Controlling $\|\widehat{O}_1^{(j)}\|_q$. By Lemma B.1, the forecast covariance bound established in subsection 5.3.2, and Theorem B.5, it holds that

$$\begin{aligned}\|\widehat{O}_1^{(j)}\|_q &= \||\mathscr{K}(\widehat{C}^{(j)})(\widehat{\Gamma}^{(j)}-\Gamma)\mathscr{K}^\top(\widehat{C}^{(j)})|\|_q \\ &\le \||\mathscr{K}(\widehat{C}^{(j)})|\|_{4q}^2\|\widehat{\Gamma}^{(j)}-\Gamma\|_{2q} \le |H|^2|\Gamma^{-1}|^2\||\widehat{C}^{(j)}|\|_{4q}^2\|\widehat{\Gamma}^{(j)}-\Gamma\|_{2q} \\ &\le c(1\vee\Omega)^2\sqrt{\frac{r_2(\Gamma)}{N}} \le c\sqrt{\frac{r_2(\Gamma)}{N}},\end{aligned}$$

where the last inequality uses that by assumption $N \ge r_2(\Sigma^{(0)}) \vee r_2(\Xi) \vee r_2(\Gamma)$.

Controlling $\|\widehat{O}_2^{(j)}\|_q$. By Lemma B.1, inequality (5.6), and the forecast covariance bound established in subsection 5.3.2, we get

$$\begin{aligned}\|\widehat{O}_2^{(j)}\|_q &= \||(I-\mathscr{K}(\widehat{C}^{(j)})H)\widehat{C}_{u\eta}^{(j)}\mathscr{K}^\top(\widehat{C}^{(j)})|\|_q \le \||\mathscr{K}(\widehat{C}^{(j)})||I-\mathscr{K}(\widehat{C}^{(j)})H||\widehat{C}_{u\eta}^{(j)}|\|_q \\ &\le \||\mathscr{K}(\widehat{C}^{(j)})|(1+|\mathscr{K}(\widehat{C}^{(j)})||H|)|\widehat{C}_{u\eta}^{(j)}|\|_q \\ &\le \||\widehat{C}_{u\eta}^{(j)}|\|_{2q}\left(|H||\Gamma^{-1}|\||\widehat{C}^{(j)}|\|_{2q}+|H|^3|\Gamma^{-1}|^2\||\widehat{C}^{(j)}|^2\|_{2q}\right) \le c\Omega.\end{aligned}$$

Controlling $\|\widehat{O}_3^{(j)}\|_q$. Note that $\|\widehat{O}_3^{(j)}\|_q = \|\widehat{O}_2^{(j)}\|_q$.

6. Conclusions. This paper has investigated REnKF, a modification of EnKF with improved theoretical guarantees. Theorem 3.2 gives nonasymptotic error bounds for a stochastic EnKF over multiple assimilation cycles. Numerical experiments demonstrate that the benefits of introducing resampling for theory purposes do not come at the price of a deterioration in state estimation or uncertainty quantification tasks.

Resampling techniques for ensemble Kalman algorithms deserve further research. From a theory viewpoint, resampling offers a promising path to develop long-time filter accuracy theory, blending our inductive analysis with existing results that ensure long-time stability of the filtering distributions [44]. From a methodological viewpoint, other resampling schemes can be considered [40]. Finally, while our numerical investigation has focused on settings where the standard EnKF algorithm is effective, an important open problem is to identify dynamical systems and/or observation models for which resampling may offer an empirical advantage.

Appendix A. Metrics for numerical results. In this appendix, we give a more extensive description of the Monte Carlo procedure utilized to calculate the metrics referred to in section 4. We summarize the approach in Algorithm A.1. We require the following additional notation: We write $\mathrm{diag}(A) = (A_{11}, A_{22}, \dots, A_{dd})^\top$. For a function $g : \mathbb{R} \to \mathbb{R}$, $g(u) = (g(u(1)), \dots, g(u(d)))^\top$ is the elementwise application of g to u.

Appendix B. Technical results.

B.1. Additional notation. Given a nondecreasing, nonzero convex function $\psi : [0,\infty] \to [0,\infty]$ with $\psi(0) = 0$, the Orlicz norm of a real random variable X is $\|X\|_\psi = \inf\{t > 0 : \mathbb{E}[\psi(t^{-1}|X|)] \le 1\}$. In particular, for the choice $\psi_p(x) = e^{x^p} - 1$ for $p \ge 1$, real random variables that satisfy $\|X\|_{\psi_2} < \infty$ are referred to as sub-Gaussian, and those that satisfy

Algorithm A.1. Metrics calculation for numerical results.

1: **Fixed quantities**: Ground-truth state $\{u^{(j)}\}_{j=0}^{J}$, observations $\{y^{(j)}\}_{j=1}^{J}$, and KF means $\{\mu^{(j)}\}_{j=1}^{J}$.

2: **Monte Carlo trials**: For $m=1,2,\dots,M$ run algorithm $\mathsf{A}\in\{\mathsf{EnKF}(2.1),\mathsf{REnKF}(3.1)\}$ and obtain $\{\widehat{\mu}_m^{(j),\mathsf{A}},\widehat{\Sigma}_m^{(j),\mathsf{A}}\}_{j,m}^{J,M}$.

Mean error:

$$\mathsf{E}^{\mathsf{A}}_{m,\mathrm{Linear}}=\frac{1}{J}\sum_{j=1}^{J}|\widehat{\mu}_m^{(j),\mathsf{A}}-\mu^{(j)}|_2,\qquad \mathsf{E}^{\mathsf{A}}_{m,\mathrm{L96}}=\frac{1}{J}\sum_{j=1}^{J}|\widehat{\mu}_m^{(j),\mathsf{A}}-u^{(j)}|_2. \tag{A.1}$$

Confidence interval: Let $\widehat{\sigma}_m^{(j),\mathsf{A}}=\sqrt{\mathrm{diag}(\widehat{\Sigma}_m^{(j),\mathsf{A}})}$, then compute

$$\begin{aligned}
\mathsf{I}_m^{(j),\mathsf{A}}&=\widehat{\mu}_m^{(j),\mathsf{A}}\pm 1.96\times\widehat{\sigma}_m^{(j),\mathsf{A}} && \text{(Interval)},\\
\mathsf{W}_m^{\mathsf{A}}&=\frac{2\times 1.96}{dJ}\sum_{j=1}^{J}|\widehat{\sigma}_m^{(j),\mathsf{A}}|_1 && \text{(Average width)},\\
\mathsf{V}_m^{\mathsf{A}}&=\frac{1}{dJ}\sum_{j=1}^{J}\sum_{i=1}^{d}\mathbf{1}\{u^{(j)}(i)\in\mathsf{I}_m^{(j),\mathsf{A}}(i)\} && \text{(Average coverage)}.
\end{aligned} \tag{A.2}$$

3: **Output**:

$$\begin{aligned}
\mathsf{E}^{\mathsf{A}}_{\mathrm{Linear}}&=\frac{1}{M}\sum_{m=1}^{M}\mathsf{E}^{\mathsf{A}}_{m,\mathrm{Linear}},\qquad \mathsf{E}^{\mathsf{A}}_{\mathrm{L96}}=\frac{1}{M}\sum_{m=1}^{M}\mathsf{E}^{\mathsf{A}}_{m,\mathrm{L96}},\\
\mathsf{W}^{\mathsf{A}}&=\frac{1}{M}\sum_{m=1}^{M}\mathsf{W}^{\mathsf{A}}_{m},\qquad \mathsf{V}^{\mathsf{A}}=\frac{1}{M}\sum_{m=1}^{M}\mathsf{V}^{\mathsf{A}}_{m}.
\end{aligned} \tag{A.3}$$

$\|X\|_{\psi_1}<\infty$ are subexponential. The random vector Y is sub-Gaussian (subexponential) if $\|v^\top Y\|_{\psi_2}<\infty$ ($\|v^\top Y\|_{\psi_1}<\infty$) for any vector v satisfying $|v|_2=1$.

B.2. Background results.

Lemma B.1 (properties of the Kalman gain operator [28, Lemma 4.1 and Corollary 4.2]). *Let $\mathscr{K}$ be the Kalman gain operator defined in (3.5). Let $P,Q\in\mathcal{S}_+^d$, $\Gamma\in\mathcal{S}_{++}^k$, and $H\in\mathbb{R}^{k\times d}$. The following hold:*

$$\begin{aligned}
|\mathscr{K}(Q)-\mathscr{K}(P)|&\le|Q-P||H||\Gamma^{-1}|\Big(1+\min\left(|P|,|Q|\right)|H|^2|\Gamma^{-1}|\Big),\\
|\mathscr{K}(Q)|&\le|Q||H||\Gamma^{-1}|,\\
|I-\mathscr{K}(Q)H|&\le 1+|Q||H|^2|\Gamma^{-1}|.
\end{aligned}$$

Lemma B.2 (properties of the mean-update operator [28, Lemma 4.10]). *Let $\mathscr{M}$ be the mean-update operator defined in (3.6). Let $m\in\mathbb{R}^d$ be a random vector, and let Q be a random matrix such that $Q\in\mathcal{S}_+^d$ almost surely. Let $P\in\mathcal{S}_+^d$, $\Gamma\in\mathcal{S}_{++}^k$, $H\in\mathbb{R}^{k\times d}$, $y\in\mathbb{R}^k$, and $m'\in\mathbb{R}^d$ be deterministic. Then, for any $1\le q<\infty$ and $y\in\mathbb{R}^k$ it holds that*

$$\begin{aligned}\|\|\mathscr{M}(m,Q;y)-\mathscr{M}(m',P;y)\|\|_q \le \big\|\|m-m'\|\big\|_q + |H|^2|\Gamma^{-1}|\|\|Q\|\|_{2q}\big\|\|m-m'\|\big\|_{2q}\\ + \|\|Q-P\|\|_q|H||\Gamma^{-1}|(1+|H|^2|\Gamma^{-1}||P|)|y-Hm'|.\end{aligned}$$

Lemma B.3 (properties of the covariance-update operator [28, Lemmas 4.6 and 4.8]). *Let $\mathscr{C}$ be the covariance-update operator defined in (3.7). Let P,Q,m,m',y,H, and Γ all be defined as in Lemma B.2. Then, for any $1\le q<\infty$, it holds that*

$$\begin{aligned}0\preceq\mathscr{C}(Q)\preceq Q,\qquad |\mathscr{C}(Q)|\le|Q|,\\ \|\|\mathscr{C}(Q)-\mathscr{C}(P)\|\|_q \le \|\|Q-P\|\|_q(1+|H|^2|\Gamma^{-1}||P|)\\ +(|H|^2|\Gamma^{-1}|+|H|^4|\Gamma^{-1}|^2|P|)\|\|Q\|\|_{2q}\|\|Q-P\|\|_{2q}.\end{aligned}$$

Theorem B.4 (Gaussian norm concentration, [49, Exercise 6.3.5]). *Let $X\in\mathbb{R}^d$ be a Gaussian random vector with $\mathbb{E}[X]=\mu^X$, $var[X]=\Sigma^X$. Then, for any $t\ge1$, with probability at least $1-ce^{-t}$ it holds that*

$$|X-\mu^X|_2 \lesssim \sqrt{Tr(\Sigma^X)}+\sqrt{t|\Sigma^X|} \lesssim \sqrt{|\Sigma^X|(r_2(\Sigma^X)\vee t)}\,.$$

Theorem B.5 (covariance bound, [27, Corollary 2]). *Let $X_1,\dots,X_n$ be i.i.d. copies of a d-dimensional Gaussian vector X with $\mathbb{E}[X]=0$ and $var[X]=\Sigma$. Let $\widehat{\Sigma}=\frac{1}{n}\sum_{i=1}^n X_iX_i^\top$ be the sample covariance estimator. For any $q\ge1$, it holds that*

$$\|\|\widehat{\Sigma}-\Sigma\|\|_q \lesssim_q |\Sigma|\left(\sqrt{\frac{r_2(\Sigma)}{n}}\vee\frac{r_2(\Sigma)}{n}\right).$$

Lemma B.6. *Let $A,B\in\mathcal{S}_+^d$. It holds that*

$$Tr(AB)\le|A|\,Tr(B).$$

Lemma B.7 (cross-covariance estimation—unstructured case). *Let $u_1,\dots,u_N\in\mathbb{R}^d$ be i.i.d. Gaussian random vectors with $\mathbb{E}[u_1]=m$, and let $var[u_1]=C$. Let $\eta_1,\dots,\eta_N\in\mathbb{R}^k$ be i.i.d. Gaussian random vectors with $\mathbb{E}[\eta_1]=0$ and $var[\eta_1]=\Gamma$, and assume that the two sequences are independent. Let*

$$\widehat{C}_{u\eta}=\frac{1}{N-1}\sum_{n=1}^N(u_n-\widehat{m})(\eta_n-\bar{\eta})^\top,$$

and assume that $N\ge r_2(C)\vee r_2(\Gamma)$. Then,

$$\|\|\widehat{C}_{u\eta}\|\|_q \lesssim_q (|C|\vee|\Gamma|)\left(\sqrt{\frac{r_2(C)}{N}}\vee\sqrt{\frac{r_2(\Gamma)}{N}}\right).$$

Proof. By [2, Lemma A.3], there exists a constant c such that, for all $t\ge1$, it holds with probability at least $1-ce^{-t}$ that

$$|\widehat{C}_{u\eta}| \lesssim (|C|\vee|\Gamma|)\left(\sqrt{\frac{r_2(C)}{N}}\vee\sqrt{\frac{r_2(\Gamma)}{N}}\vee\sqrt{\frac{t}{N}}\vee\frac{t}{N}\right).$$

Integrating the tail bound then yields the result. ∎

Lemma B.8. *Let $X_1, \dots, X_n$ be i.i.d. copies of a d-dimensional Gaussian vector X with $\mathbb{E}[X] = 0$ and $var[X] = \Sigma$. Let $\widehat{\Sigma} = \frac{1}{n}\sum_{i=1}^n X_i X_i^\top$ be the sample covariance estimator. Then, for any $\delta \geq 1$, it holds with probability at least $1 - 2e^{-\delta}$ that*

$$|Tr(\widehat{\Sigma}) - Tr(\Sigma)| \leq c\, Tr(\Sigma) \left(\sqrt{\frac{\delta}{n}} \vee \frac{\delta}{n} \right).$$

Further, for any $q \geq 1$,

$$\| |Tr(\widehat{\Sigma}) - Tr(\Sigma)| \|_q \lesssim_q \frac{Tr(\Sigma)}{\sqrt{n}}.$$

Proof. Let $Z_{ij} = \Sigma_{jj}^{-1/2} X_{ij}$ and note that, for any $t > 0$,

$$\begin{aligned}\mathbb{P}(|\mathrm{Tr}(\widehat{\Sigma}) - \mathrm{Tr}(\Sigma)| > t) = \mathbb{P}(|\mathrm{Tr}(\widehat{\Sigma} - \Sigma)| > t) &= \mathbb{P}\left(\left| \sum_{i=1}^n \left(\sum_{j=1}^d (X_{ij}^2 - \mathbb{E}\, X_{ij}^2) \right) \right| > nt \right) \\ &= \mathbb{P}\left(\left| \sum_{i=1}^n \left(\sum_{j=1}^d \Sigma_{jj} (Z_{ij}^2 - \mathbb{E}\, Z_{ij}^2) \right) \right| > nt \right).\end{aligned}$$

Note that the random variables $\sum_{j=1}^d \Sigma_{jj}(Z_{ij}^2 - \mathbb{E}\, Z_{ij}^2)$ for $i = 1, \dots, n$ are independent, mean-zero, and subexponential with ψ_1 norm at most $C\mathrm{Tr}(\Sigma)$, since

$$\begin{aligned}\left\| \sum_{j=1}^d \Sigma_{jj} (Z_{ij}^2 - \mathbb{E}\, Z_{ij}^2) \right\|_{\psi_1} &\leq \sum_{j=1}^d \Sigma_{jj} \left\| Z_{ij}^2 - \mathbb{E}\, Z_{ij}^2 \right\|_{\psi_1} \leq C \sum_{j=1}^d \Sigma_{jj} \left\| Z_{ij}^2 \right\|_{\psi_1} \\ &= C \sum_{j=1}^d \Sigma_{jj} \| Z_{ij} \|_{\psi_2}^2 \leq C \sum_{j=1}^d \Sigma_{jj} = C\mathrm{Tr}(\Sigma).\end{aligned}$$

The second inequality holds due to the Centering Lemma, [49, Lemma 2.6.8]. Therefore, by Bernstein's inequality we have

$$\mathbb{P}\left(\left| \sum_{i=1}^n \left(\sum_{j=1}^d \Sigma_{jj} (Z_{ij}^2 - \mathbb{E}\, Z_{ij}^2) \right) \right| > nt \right) \leq 2 \exp\left(-c \min\left(\frac{nt^2}{(\mathrm{Tr}(\Sigma))^2}, \frac{nt}{\mathrm{Tr}(\Sigma)} \right) \right).$$

For the expectation bound, we note that

$$\begin{aligned}\| |\mathrm{Tr}(\widehat{\Sigma}) - \mathrm{Tr}(\Sigma)| \|_q^q &= \int_0^\infty \mathbb{P}(|\mathrm{Tr}(\widehat{\Sigma}) - \mathrm{Tr}(\Sigma)|^q > t)\, dt \\ &\leq \zeta^q + q \int_C^\infty t^{q-1} \mathbb{P}(|\mathrm{Tr}(\widehat{\Sigma}) - \mathrm{Tr}(\Sigma)| > t)\, dt \\ &\leq \zeta^q + 2q \int_0^\infty t^{q-1} \exp\left(-c \min\left(\frac{nt^2}{(\mathrm{Tr}(\Sigma))^2}, \frac{nt}{\mathrm{Tr}(\Sigma)} \right) \right) dt \\ &= \zeta^q + 2qc \max\left(\frac{\Gamma(q/2)(\mathrm{Tr}(\Sigma))^q}{n^{q/2}}, \frac{\Gamma(q)(\mathrm{Tr}(\Sigma))^q}{n^q} \right).\end{aligned}$$

Taking $\zeta = \mathrm{Tr}(\Sigma)/n$, it then follows that

$$\left\|\left|\mathrm{Tr}(\widehat{\Sigma}) - \mathrm{Tr}(\Sigma)\right|\right\|_q \lesssim \zeta + c\mathrm{Tr}(\Sigma)\max\left(\frac{1}{\sqrt{n}}, \frac{1}{n}\right) \lesssim \frac{\mathrm{Tr}(\Sigma)}{\sqrt{n}}.$$

Lemma B.9. *Let $X_1, \ldots, X_n$ be i.i.d. copies of a d-dimensional Gaussian vector X with $\mathbb{E}[X] = \mu$, and let $var[X] = \Sigma$. Let $\overline{X} = \frac{1}{N}\sum_{n=1}^{N} X_n$. Then, for any $q \geq 1$,*

$$\left\|\left|\bar{X} - \mu\right|\right\|_q \lesssim_q \sqrt{\frac{Tr(\Sigma)}{N}}.$$

Proof. Let $c_2 := c_1\sqrt{\mathrm{Tr}(\Sigma)/N}$ where c_1 is a sufficiently large positive constant, then

$$\begin{aligned}
\mathbb{E}\left[|\bar{X} - \mu|^q\right] &= \int_0^\infty \mathbb{P}(|\bar{X} - \mu|^q > y)\, dy \leq c_2^q + \int_{c_2}^\infty \mathbb{P}(|\bar{X} - \mu|^q > y)\, dy \\
&= c_2^q + \int_{c_2}^\infty q y^{q-1} \mathbb{P}(|\bar{X} - \mu| > y)\, dy \\
&= c_2^q + \int_{c_2 - c\mathrm{Tr}(\Sigma/N)}^\infty q\left(c\sqrt{\frac{\mathrm{Tr}(\Sigma)}{N}} + t\right)^{q-1} \mathbb{P}\left(|\bar{X} - \mu| > c\sqrt{\frac{\mathrm{Tr}(\Sigma)}{N}} + t\right) dt,
\end{aligned}$$

where the last equality holds by a change of variable. By Theorem B.4 it follows that $\mathbb{P}(|\overline{X} - \mu| \geq c\sqrt{\mathrm{Tr}(\Sigma)} + t) \leq \exp(-ct^2/|\Sigma|)$, and so the expression in the above display is bounded above by

$$\begin{aligned}
&c_2^q + \int_{c_2 - c\mathrm{Tr}(\Sigma/N)}^\infty q\left(c\sqrt{\frac{\mathrm{Tr}(\Sigma)}{N}} + t\right)^{q-1} \exp\left(-\frac{c_2 n t^2}{|\Sigma|}\right) dt \\
&\lesssim c_2^q + \int_0^\infty q\left(\left(\frac{c\mathrm{Tr}(\Sigma)}{N}\right)^{(q-1)/2} + t^{q-1}\right) \exp\left(-\frac{c_2 n t^2}{|\Sigma|}\right) dt \\
&= c_2^q + q\left(\frac{1}{2}\Gamma(q/2)\left(\frac{|\Sigma|}{N}\right)^{q/2} + \frac{1}{2}\left(\frac{c\mathrm{Tr}(\Sigma)}{N}\right)^{(q-1)/2}\sqrt{\frac{\pi|\Sigma|}{N}}\right) \\
&\lesssim c_2^q + q\left(\frac{1}{2}\Gamma(q/2)\left(\frac{|\Sigma|}{N}\right)^{q/2} + \frac{1}{2}\left(\frac{c\mathrm{Tr}(\Sigma)}{N}\right)^{q/2}\right) \\
&\lesssim \left(\frac{\mathrm{Tr}(\Sigma)}{N}\right)^{q/2}.
\end{aligned}$$

Therefore,

$$\left\|\left|\bar{X} - \mu\right|\right\| \lesssim_q c_2 + \sqrt{\frac{|\Sigma|}{N}} + c\sqrt{\frac{\mathrm{Tr}(\Sigma)}{N}} \lesssim \sqrt{\frac{\mathrm{Tr}(\Sigma)}{N}},$$

where the last inequality holds since $\mathrm{Tr}(\Sigma) \geq |\Sigma|$ and the choice of c_2.

REFERENCES

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

[2] O. Al-Ghattas and D. Sanz-Alonso, *Non-asymptotic Analysis of Ensemble Kalman Updates: Effective Dimension and Localization*, preprint, https://arxiv.org/abs/2208.03246, 2022.

[3] J. L. Anderson, *An ensemble adjustment Kalman filter for data assimilation*, Mon. Weather Rev., 129 (2001), pp. 2884–2903.
[4] J. L. Anderson and S. L. Anderson, *A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts*, Mon. Weather Rev., 127 (1999), pp. 2741–2758.
[5] M. Asch, M. Bocquet, and M. Nodet, *Data Assimilation: Methods, Algorithms, and Applications*, SIAM, Philadelphia, 2016, https://doi.org/10.1137/1.9781611974546.
[6] C. H. Bishop, B. J. Etherton, and S. J. Majumdar, *Adaptive sampling with the ensemble transform Kalman filter. Part* I: *Theoretical aspects*, Monthly Weather Rev., 129 (2001), pp. 420–436.
[7] N. K. Chada, Y. Chen, and D. Sanz-Alonso, *Iterative ensemble Kalman methods: A unified perspective with some new variants*, Found. Data Sci., 3 (2021), pp. 331–369.
[8] Y. Chen, D. Sanz-Alonso, and R. Willett, *Autodifferentiable ensemble Kalman filters*, SIAM J. Math. Data Sci., 4 (2022), pp. 801–833, https://doi.org/10.1137/21M1434477.
[9] Y. Chen, D. Sanz-Alonso, and R. Willett, *Reduced-Order Autodifferentiable Ensemble Kalman Filters*, preprint, https://arxiv.org/abs/2301.11961, 2023.
[10] N. Chopin and O. Papaspiliopoulos, *An Introduction to Sequential Monte Carlo*, Vol. 4, Springer, 2020.
[11] D. Crisan and B. Rozovskii, *The Oxford Handbook of Nonlinear Filtering*, Oxford University Press, 2011.
[12] P. Del Moral, *Feynman-Kac Formulae*, Springer, 2004.
[13] A. Doucet and A. M. Johansen, *A tutorial on particle filtering and smoothing: Fifteen years later*, in Oxford Handbooks in Mathematics, Oxford University Press, Oxford, New York, pp. 656–705.
[14] O. G. Ernst, B. Sprungk, and H.-J. Starkloff, *Analysis of the ensemble and polynomial chaos Kalman filters in Bayesian inverse problems*, SIAM/ASA J. Uncertain. Quantif., 3 (2015), pp. 823–851, https://doi.org/10.1137/140981319.
[15] G. Evensen, *Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics*, J. Geophys. Res.: Oceans, 99 (1995), pp. 10143–10162.
[16] G. Evensen, *The ensemble Kalman filter: Theoretical formulation and practical implementation*, Ocean Dynam., 53 (2003), pp. 343–367.
[17] G. Evensen, *Data Assimilation: The Ensemble Kalman Filter*, Springer-Verlag, Berlin, Heidelberg, 2009.
[18] G. Evensen and P. V. Leeuwen, *Assimilation of Geosat altimeter data for the Agulhas current using the ensemble Kalman filter with a quasigeostrophic model*, Mon. Weather Rev., 124 (1996), pp. 85–96.
[19] G. Evensen, F. C. Vossepoel, and P. J. van Leeuwen, *Data Assimilation Fundamentals: A Unified Formulation of the State and Parameter Estimation Problem*, Springer, Cham, 2022.
[20] R. Furrer and T. Bengtsson, *Estimation of high-dimensional prior and posterior covariance matrices in Kalman filter variants*, J. Multivariate Anal., 98 (2007), pp. 227–255.
[21] Y. Gu and D. S. Oliver, *An iterative ensemble Kalman filter for multiphase fluid flow data assimilation*, SPE J., 12 (2007), pp. 438–446.
[22] J. Harlim and A. J. Majda, *Catastrophic filter divergence in filtering nonlinear dissipative systems*, Commun. Math. Sci., 8 (2010), pp. 27–43.
[23] P. L. Houtekamer and F. Zhang, *Review of the ensemble Kalman filter for atmospheric data assimilation*, Mon. Weather Rev., 144 (2016), pp. 4489–4532.
[24] M. A. Iglesias, K. J. H. Law, and A. M. Stuart, *Ensemble Kalman methods for inverse problems*, Inverse Problems, 29 (2013), 045001.
[25] R. E. Kalman, *A new approach to linear filtering and prediction problems*, J. Basic Eng., 82 (1960), pp. 35–45.
[26] D. Kelly, A. J. Majda, and X. T. Tong, *Concrete ensemble Kalman filters with rigorous catastrophic filter divergence*, Proc. Natl. Acad. Sci., 112 (2015), pp. 10589–10594.
[27] V. Koltchinskii and K. Lounici, *Concentration inequalities and moment bounds for sample covariance operators*, Bernoulli, 23 (2017), pp. 110–133.
[28] E. Kwiatkowski and J. Mandel, *Convergence of the square root ensemble Kalman filter in the large ensemble limit*, SIAM/ASA J. Uncertain. Quantif., 3 (2015), pp. 1–17, https://doi.org/10.1137/140965363.

[29] K. Law, A. Stuart, and K. Zygalakis, *Data Assimilation*, Texts Appl. Math. 62, Springer, Cham, 2015.
[30] K. J. Law, D. Sanz-Alonso, A. Shukla, and A. M. Stuart, *Filter accuracy for the Lorenz 96 model: Fixed versus adaptive observation operators*, Phys. D, 325 (2016), pp. 1–13.
[31] K. J. Law and A. M. Stuart, *Evaluating data assimilation algorithms*, Mon. Weather Rev., 140 (2012), pp. 3757–3782.
[32] W. G. Lawson and J. A. Hansen, *Implications of stochastic and deterministic filters as ensemble-based data assimilation methods in varying regimes of error growth*, Mon. Weather Rev., 132 (2004), pp. 1966–1981.
[33] G. Li and A. C. Reynolds, *An iterative ensemble Kalman filter for data assimilation*, in SPE Annual Technical Conference and Exhibition, Society of Petroleum Engineers, 2007.
[34] E. N. Lorenz, *Predictability: A problem partly solved*, in Proceedings Seminar on Predictability, Vol. 1, Reading, UK, 1996.
[35] A. Majda and X. Wang, *Nonlinear Dynamics and Statistical Theories for Basic Geophysical Flows*, Cambridge University Press, 2006.
[36] A. J. Majda and J. Harlim, *Filtering Complex Turbulent Systems*, Cambridge University Press, 2012.
[37] A. J. Majda and X. T. Tong, *Performance of ensemble Kalman filters in large dimensions*, Comm. Pure Appl. Math., 71 (2018), pp. 892–937.
[38] J. Mandel, L. Cobb, and J. D. Beezley, *On the convergence of the ensemble Kalman filter*, Appl. Math., 56 (2011), pp. 533–541.
[39] I. Myrseth, J. Sætrom, and H. Omre, *Resampling the ensemble Kalman filter*, Comput. Geosci., 55 (2013), pp. 44–53.
[40] C. Naesseth, S. Linderman, R. Ranganath, and D. Blei, *Variational sequential Monte Carlo*, in International Conference on Artificial Intelligence and Statistics, PMLR, 2018, pp. 968–977.
[41] O. Papaspiliopoulos and M. Ruggiero, *Optimal filtering and the dual process*, Bernoulli, 20 (2014), pp. 1999–2019.
[42] S. Reich and C. Cotter, *Probabilistic Forecasting and Bayesian Data Assimilation*, Cambridge University Press, 2015.
[43] D. Sanz-Alonso, A. Stuart, and A. Taeb, *Inverse Problems and Data Assimilation*, London Math. Soc. Stud. Texts 107, Cambridge University Press, 2023.
[44] D. Sanz-Alonso and A. M. Stuart, *Long-time asymptotics of the filtering distribution for partially observed chaotic dynamical systems*, SIAM/ASA J. Uncertain. Quantif., 3 (2015), pp. 1200–1220, https://doi.org/10.1137/140997336.
[45] S. Särkkä and L. Svensson, *Bayesian Filtering and Smoothing*, IMS Textb. 17, Cambridge University Press, Cambridge, 2023.
[46] M. K. Tippett, J. L. Anderson, C. H. Bishop, T. M. Hamill, and J. S. Whitaker, *Ensemble square root filters*, Mon. Weather Rev., 131 (2003), pp. 1485–1490.
[47] J. A. Tropp, *An Introduction to Matrix Concentration Inequalities*, Found. Trends Mach. Learn. 8, Now Publishers, Tampa, FL, 2015.
[48] P. J. Van Leeuwen, *A consistent interpretation of the stochastic version of the Ensemble Kalman Filter*, Q. J. Roy. Meteorol. Soc., 146 (2020), pp. 2815–2825.
[49] R. Vershynin, *High-Dimensional Probability: An Introduction with Applications in Data Science*, Camb. Ser. Stat. Probab. Math. 47, Cambridge University Press, 2018.
[50] J. Wu, J.-X. Wang, and S. C. Shadden, *Improving the Convergence of the Iterative Ensemble Kalman Filter by Resampling*, preprint, https://arxiv.org/abs/1910.04247, 2019.
[51] J. Wu, L. Wen, and J. Li, *Resampled ensemble Kalman inversion for Bayesian parameter estimation with sequential data*, Discrete Contin. Dyn. Syst. Ser. S, 15 (2022), pp. 837–850.
[52] Y. Zhang and D. S. Oliver, *Improving the ensemble estimate of the Kalman gain by bootstrap sampling*, Math. Geosci., 42 (2010), pp. 327–345.