Information and Inference: A Journal of the IMA (2023) **13**, 1–66
https://doi.org/10.1093/imaiai/iaad043

Non-asymptotic analysis of ensemble Kalman updates: effective dimension and localization

Omar Al-Ghattas* and Daniel Sanz-Alonso
Department of Statistics, The University of Chicago, 60637 IL, USA
*Corresponding author. Email: ghattas@uchicago.edu

[Received on 5 August 2022; revised on 6 May 2023; accepted on 2 September 2023]

Many modern algorithms for inverse problems and data assimilation rely on ensemble Kalman updates to blend prior predictions with observed data. Ensemble Kalman methods often perform well with a small ensemble size, which is essential in applications where generating each particle is costly. This paper develops a non-asymptotic analysis of ensemble Kalman updates, which rigorously explains why a small ensemble size suffices if the prior covariance has moderate effective dimension due to fast spectrum decay or approximate sparsity. We present our theory in a unified framework, comparing several implementations of ensemble Kalman updates that use perturbed observations, square root filtering and localization. As part of our analysis, we develop new dimension-free covariance estimation bounds for approximately sparse matrices that may be of independent interest.

Keywords: ensemble Kalman updates; effective dimension; localization; covariance estimation.

1. Introduction

Many algorithms for inverse problems and data assimilation rely on ensemble Kalman updates to blend prior predictions with observed data. The main motivation behind ensemble Kalman methods is that they often perform well with a small ensemble size N, which is essential in applications where generating each particle is costly. However, theoretical studies have primarily focused on large ensemble asymptotics, that is, on the limit $N \to \infty$. While these *mean-field* results are mathematically interesting and have led to significant practical improvements, they fail to explain the empirical success of ensemble Kalman methods when deployed with a small ensemble size. The aim of this paper is to develop a *non-asymptotic* analysis of ensemble Kalman updates that rigorously explains why, and under what circumstances, a small ensemble size may suffice. To that end, we establish non-asymptotic error bounds in terms of suitable notions of effective dimension of the prior covariance model that account for spectrum decay (which may represent smoothness of a prior random field) and approximate sparsity (which may represent spatial decay of correlations). Our work complements mean-field analyses of ensemble Kalman updates and identifies scenarios where mean-field behavior holds with moderate N.

In addition to demystifying the practical success of ensemble Kalman methods with a small ensemble size, our non-asymptotic perspective allows us to tell apart, on accuracy grounds, implementations of ensemble Kalman updates that use perturbed observations (POs) and square root (SR) filtering. These implementations become equivalent in the large N limit, and therefore their differences in accuracy cannot be captured by asymptotic results. Furthermore, our non-asymptotic perspective provides new understanding on the importance of localization, a procedure widely used by practitioners that involves tapering or 'localizing' empirical covariance estimates to avoid spurious correlations.

Rather than providing a complete, definite analysis of any particular ensemble Kalman method, our goal is to bring to bear a new set of tools from high-dimensional probability and statistics to the study of

these algorithms. In particular, our work builds on and contributes to the theory of high-dimensional covariance estimation, which we believe is fundamental to the understanding of ensemble Kalman methods. To make the presentation accessible to a wide audience, we assume no background knowledge on covariance estimation or on ensemble Kalman methods.

1.1 *Problem description*

Consider the inverse problem of recovering $u \in \mathbb{R}^d$ from data $y \in \mathbb{R}^k$, corrupted by noise η, where

$$y = \mathcal{G}(u) + \eta, \tag{1.1}$$

$\mathcal{G} : \mathbb{R}^d \to \mathbb{R}^k$ is the forward model, and $\eta \sim \mathbb{P}_\eta = \mathcal{N}(0, \Gamma)$ is the observation error with positive-definite covariance matrix Γ. An ensemble Kalman update takes as input a *prior ensemble* $\{u_n\}_{n=1}^N$ and observed data y, and returns as output an *updated ensemble* $\{\upsilon_n\}_{n=1}^N$ that blends together the information in the prior ensemble and in the data. Two main types of problems will be investigated: posterior approximation and sequential optimization. In the former, ensemble Kalman updates are used to approximate a posterior distribution in a Bayesian linear setting; in the latter, they are used within optimization algorithms for nonlinear inverse problems:

1.1.1 *Posterior approximation* If the forward model is linear, i.e. $\mathcal{G}(u) = Au$ for some matrix $A \in \mathbb{R}^{k\times d}$, and A is ill-conditioned or $d \gg k$, naive inversion of the data by means of the (generalized) inverse of A results in an amplification of small observation error η into large error in the reconstruction of u. In such situations, regularization is needed to stabilize the solution. To this end, one may adopt a Bayesian approach and place a Gaussian prior on the unknown $u \sim \mathbb{P}_u = \mathcal{N}(m, C)$ with positive-definite C; the prior distribution then acts as a probabilistic regularizer. The Bayesian solution to the inverse problem (1.1) is a full characterization of the posterior distribution $\mathbb{P}_{u|y}$, that is, the distribution of u given y. A standard calculation shows that $\mathbb{P}_{u|y} = \mathcal{N}(\mu, \Sigma)$, with

$$\begin{aligned} \mu &= m + CA^\top (ACA^\top + \Gamma)^{-1}(y - Am), \\ \Sigma &= C - CA^\top (ACA^\top + \Gamma)^{-1}AC, \end{aligned} \tag{1.2}$$

which require the storage of $d \times d$ matrices and consequently are difficult to compute explicitly when the state dimension d is large. A posterior-approximation ensemble Kalman update transforms a prior ensemble $\{u_n\}_{n=1}^N$ drawn from $\mathbb{P}_u$ into an updated ensemble $\{\upsilon_n\}_{n=1}^N$ whose sample mean and sample covariance approximate the mean and covariance of $\mathbb{P}_{u|y}$. Ensemble Kalman updates enjoy a low computational and memory cost when the ensemble size N is smaller than the state dimension d. In Section 2, we establish non-asymptotic error bounds that ensure that if N is larger than a suitably defined *effective dimension*, then the sample mean and sample covariance of the updated ensemble approximate well the true posterior mean and covariance in (1.2). We refer to methods that are capable of approximating well the posterior $\mathbb{P}_{u|y}$ in a linear-Gaussian setting as *posterior-approximation* algorithms.

1.1.2 *Sequential optimization* When faced with a general nonlinear model $\mathcal{G}$, exact characterization of the posterior can be challenging. One may then opt for an optimization framework and solve the inverse problem (1.1) by minimizing a user-chosen objective function. Starting from a prior ensemble

$\{u_n\}_{n=1}^N$ drawn from a measure $\mathbb{P}_u$ that encodes prior beliefs about u, an ensemble Kalman update returns an updated ensemble $\{v_n\}_{n=1}^N$ whose sample mean approximates the desired minimizer. The process can be iterated by taking the updated ensemble to be the prior ensemble of a new ensemble Kalman update. Under suitable conditions on $\mathcal{G}$, and after a sufficient number of such updates, all particles in the ensemble collapse into the minimizer of the objective. Ensemble Kalman optimization algorithms are derivative-free methods, and are therefore particularly useful when derivatives of the model $\mathcal{G}$ are unavailable or expensive to compute. As for posterior-approximation algorithms, implementing each update has low computational and memory cost when the ensemble size N is small. In Section 3, we will establish non-asymptotic error bounds that ensure that if N is larger than a suitably defined *effective dimension*, then each particle update $u_n \mapsto v_n$, $1 \leq n \leq N$, approximates well an idealized mean-field update computed with an infinite number of particles; this suggests that the evolution of particles along an ensemble-based sequential optimizer is close to an idealized mean-field evolution. We refer to methods that solve the inverse problem (1.1) by minimization of an objective function as *sequential-optimization* algorithms.

1.2 *Summary of contributions and outline*

- Section 2 is concerned with posterior-approximation algorithms. The main results, Theorems 2.3 and 2.5, give non-asymptotic bounds on the estimation of the posterior mean and covariance in terms of a standard notion of effective dimension that accounts for spectrum decay in the prior covariance model. Our analysis explains the statistical advantage of SR updates over PO ones. We also discuss the deterioration of our bounds in small noise limits where the prior and the posterior become mutually singular.
- Section 3 is concerned with sequential-optimization algorithms. The main results, Theorems 3.5 and 3.7, give non-asymptotic bounds on the approximation of mean-field particle updates using ensemble Kalman updates with and without localization. Our analysis explains the advantage of localized updates if the prior covariance satisfies a soft-sparsity condition. For the study of localized updates, we show in Theorems 3.1 and 3.3 new dimension-free covariance estimation bounds in terms of a new notion of effective dimension that simultaneously accounts for spectrum decay and approximate sparsity in the prior covariance model.
- Section 4 concludes with a summary of our work and several research directions that stem from our non-asymptotic analysis of ensemble Kalman updates. We also discuss the potential and limitations of localization in posterior-approximation algorithms.
- The proofs of all our results are deferred to three appendices.

1.3 *Related work*

Ensemble Kalman methods—overviewed in [21,33,53,50,81,83]—first appeared as filtering algorithms in the data assimilation literature [15,31,34,47,48]. The goal of data assimilation is to estimate a time-evolving state as new observations become available [3,62,66,71,79,83,85]. Ensemble Kalman filters (EnKFs) solve an inverse problem of the form (1.1) every time a new observation is acquired. In that filtering context, (1.1) encodes the relationship between the state u and observation y at a given time t, and the prior on u is specified by propagating a probabilistic estimate of the state at time $t-1$ through the dynamical system that governs the state evolution. To approximate this prior, EnKFs propagate an ensemble of N particles through the dynamics, and subsequently update this prior *forecast* ensemble into an updated *analysis* ensemble that assimilates the new observation. Thus, an ensemble Kalman

update is performed every time a new observation is acquired. The goal is that the sample mean and sample covariance of the updated ensemble approximate well the mean and covariance of the filtering distribution, that is, the conditional distribution of the state at time t given all observations up to time t. While only giving provably accurate posterior approximation in linear settings [30], EnKFs are among the most popular methods for high-dimensional nonlinear filtering, in particular in numerical weather forecasting. In such applications the state dimension can be very large, but the effective dimension of the filter update is often much lower due to smoothness of the state and decay of correlations in space. Moreover, in practice, the analysis step can be constrained to the subspace determined by the expanding directions of the dynamics [98].

The papers [41,69,80] introduced ensemble Kalman methods for inverse problems in petroleum engineering and the geophysical sciences. Application-agnostic ensemble Kalman methods for inverse problems were developed in [51,52], inspired by classical regularization schemes [43]. Since then, a wide range of sequential-optimization algorithms for inverse problems have been proposed that differ in the objective function they seek to minimize and in how ensemble Kalman updates are implemented. We refer to Subsection 2.1 for further background and to [21] for a review.

Ensemble Kalman methods for inverse problems and data assimilation have been studied extensively from a large N asymptotic point of view; (see e.g. [10,24,26,27,30,37,45,59,63,65,70,73]). A complementary line of work [40,44,54,95,96] has focused on challenges faced by ensemble Kalman methods, including loss of stability and catastrophic filter divergence. Two overarching themes that underlie large N asymptotic analyses are to ensure consistency and to derive equations for the mean-field evolution of the ensemble. Related to this second theme, several works (e.g. [12,13,21,42,86,97]) set the analysis in a *continuous time limit*; the idea is to view Kalman updates as occurring over an artificial discrete-time variable, and then take the time between updates to be infinitesimally small to formally derive differential equations for the evolution of the ensemble or its density. Large N asymptotics and continuous time limits have resulted in new theoretical insights and practical advancements. However, an important caveat of these results is that they cannot tell apart implementations of ensemble Kalman methods that become equivalent in large N or continuous time asymptotic regimes. Moreover, several papers (e.g. [5,6,55,72,86]) have noted that large N asymptotic analyses fail to explain empirical results that report good performance with a moderately sized ensemble in problems with high state dimension; for instance, $d \sim 10^9$ and $N \sim 10^2$ in operational numerical weather prediction. Finally, the note [76] shows subtle but important differences in the evolution of interacting particle systems with finite ensemble size when compared to their mean-field counterparts [37].

In this paper we adopt a non-asymptotic viewpoint to establish sufficient conditions on the ensemble size for posterior-approximation and sequential-optimization algorithms. Empirical evidence in [77] suggests that there is a sample size N^* above which ensemble Kalman methods are effective. The seminal work [36] conducts insightful explicit calculations that motivate our more general theory. Following the analysis of ensemble Kalman methods in [36] and the study of importance sampling and particle filters in [1,82,84,9,88,4,87,25,89], we focus on analyzing a single ensemble Kalman update rather than on investigating the propagation of error across multiple updates. While in practice ensemble Kalman methods for posterior approximation in data assimilation and for sequential optimization in inverse problems often perform many updates, focusing on a single update enables us to clearly demonstrate the tight connection between the sample complexity of ensemble updates and the effective dimension of the prior; additionally, for some posterior-approximation algorithms, our theory generalizes in a straightforward way to multi-step implementations, as we shall demonstrate in Section 2. More importantly, the focus on a single update allows us to tell apart, on accuracy grounds, POs and SR implementations of ensemble Kalman updates, as well as implementations with and without localization. Similar considerations

motivate the study of sufficient sample size for importance sampling in [1,22,75,82,84,88], where the focus on a single update facilitates establishing clear comparisons between standard and optimal proposals, and identifying meaningful notions of dimension to characterize necessary and sufficient conditions on the required sample size. Our work builds on and develops tools from high-dimensional probability and statistics [7,16,17,23,68,102,103]. In particular, we bring to bear thresholded [7,16] and masked covariance estimators [23,68] to the understanding of localization in ensemble Kalman methods. In so doing, we establish new dimension-free covariance and cross-covariance estimation bounds under approximate sparsity—see Theorems 3.1 and 3.3.

1.4 *Notation*

Given two positive sequences $\{a_n\}$ and $\{b_n\}$, the relation $a_n \lesssim b_n$ denotes that $a_n \leq cb_n$ for some constant $c > 0$. If the constant c depends on some quantity τ, then we write $a \lesssim_\tau b$. If both $a_n \lesssim b_n$ and $b_n \lesssim a_n$ hold simultaneously, then we write $a_n \asymp b_n$. 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. For a vector $v \in \mathbb{R}^N$, $\|v\|_p^p = \sum_{n=1}^N |v_n|^p$. For a matrix $A \in \mathbb{R}^{n\times m}$, the operator norm is given by $\|A\| = \sup_{\|v\|_2=1} \|Av\|_2$. $\mathcal{S}_+^d$ 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. $A^\dagger$ denotes the pseudo-inverse of A. 1_N denotes the N-dimensional vector vector of ones, 0_d denotes the d-dimensional vector of zeroes, and $O_{d\times k}$ is the $d \times k$ matrix of zeroes. $\mathbf{1}_B$ denotes the indicator of the set B. $\equiv$ denotes a definition. $\circ$ denotes the matrix Hadamard or Schur (elementwise) product. Given a non-decreasing, non-zero convex function $\psi : [0, \infty] \to [0, \infty]$ with $\psi(0) = 0$, the Orlicz norm of a real random variable X is $\|X\|_\psi = \inf\{t > 0 : \mathbb{E}[\psi(t^{-1}|X|)] \leq 1\}$. In particular, for the choice $\psi_p(x) \equiv e^{x^p} - 1$ for $p \geq 1$, real random variables that satisfy $\|X\|_{\psi_2} < \infty$ are referred to as sub-Gaussian. A random vector X is sub-Gaussian if $\|v^\top X\|_{\psi_2} < \infty$ for any v such that $\|v\|_2 = 1$. For a differentiable function $g : \mathbb{R}^d \to \mathbb{R}^k$, $Dg \in \mathbb{R}^{d\times k}$ denotes the Jacobian of g.

All the methods we study have the same starting point of a prior ensemble

$$u_1, \ldots, u_N \overset{\text{i.i.d.}}{\sim} \mathcal{N}(m, C),$$

and observed data y generated according to (1.1), which are to be used in generating an updated ensemble $\{v_n\}_{n=1}^N$. We denote the prior sample means by

$$\widehat{m} \equiv \frac{1}{N}\sum_{n=1}^N u_n, \qquad \bar{\mathcal{G}} \equiv \frac{1}{N}\sum_{n=1}^N \mathcal{G}(u_n),$$

and the prior sample covariances by

$$\begin{aligned} \widehat{C} &\equiv \frac{1}{N-1}\sum_{n=1}^N (u_n - \widehat{m})(u_n - \widehat{m})^\top, \qquad \widehat{C}^{pp} \equiv \frac{1}{N-1}\sum_{n=1}^N (\mathcal{G}(u_n) - \bar{\mathcal{G}})(\mathcal{G}(u_n) - \bar{\mathcal{G}})^\top, \\ \widehat{C}^{up} &\equiv \frac{1}{N-1}\sum_{n=1}^N (u_n - \widehat{m})(\mathcal{G}(u_n) - \bar{\mathcal{G}})^\top. \end{aligned} \tag{1.3}$$

The population versions will be denoted by

$$C^{pp} \equiv \mathbb{E}\Big[\big(\mathcal{G}(u_n) - \mathbb{E}[\mathcal{G}(u_n)]\big)\big(\mathcal{G}(u_n) - \mathbb{E}[\mathcal{G}(u_n)]\big)^\top\Big], \qquad C^{up} \equiv \mathbb{E}\Big[\big(u_n - m\big)\big(\mathcal{G}(u_n) - \mathbb{E}[\mathcal{G}(u_n)]\big)^\top\Big].$$

2. Ensemble Kalman updates: posterior approximation algorithms

In posterior-approximation algorithms we consider the inverse problem (1.1) with a linear forward model, i.e.

$$y = Au + \eta, \qquad \eta \sim \mathcal{N}(0, \Gamma). \tag{2.1}$$

In order to establish comparisons between different posterior-approximation algorithms, as well as to streamline our analysis, we follow the exposition in [59] 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

$$\mathscr{K} : \mathcal{S}_+^d \to \mathbb{R}^{d\times k}, \qquad \mathscr{K}(C; A, \Gamma) = \mathscr{K}(C) = CA^\top(ACA^\top + \Gamma)^{-1}, \tag{2.2}$$

$$\mathscr{M} : \mathbb{R}^d \times \mathcal{S}_+^d \to \mathbb{R}^d, \qquad \mathscr{M}(m, C; A, y, \Gamma) = \mathscr{M}(m, C) = m + \mathscr{K}(C; A, \Gamma)(y - Am), \tag{2.3}$$

$$\mathscr{C} : \mathcal{S}_+^d \to \mathcal{S}_+^d, \qquad \mathscr{C}(C; A, \Gamma) = \mathscr{C}(C) = \big(I - \mathscr{K}(C; A, \Gamma)A\big)C. \tag{2.4}$$

The pointwise continuity and boundedness of all three operators was established in [59], and we summarize these results in Lemmas A.4, A.5 and A.6. We note that the Kalman update (1.2) can be rewritten succinctly as

$$\begin{aligned} \mu &= \mathscr{M}(m, C), \\ \Sigma &= \mathscr{C}(C). \end{aligned} \tag{2.5}$$

2.1 *Ensemble algorithms for posterior approximation*

We study two main classes of posterior-approximation algorithms based on PO and SR ensemble Kalman updates. In both implementations, the updated ensemble has sample mean $\widehat{\mu}$ and sample covariance $\widehat{\Sigma}$ that are, by design, consistent estimators of the posterior mean μ and covariance Σ in (2.5). Although PO and SR updates are asymptotically equivalent, differences between the two algorithms do exist in finite ensembles, and this difference is captured in our non-asymptotic analysis in Subsection 2.3.

2.1.1 *Perturbed Observation update* The PO update, introduced in [31], transforms each particle of the prior ensemble according to

$$\begin{aligned} \upsilon_n &= u_n + \mathscr{K}(\widehat{C})\big(y - Au_n - \eta_n\big) \\ &= \mathscr{M}(u_n, \widehat{C}) - \mathscr{K}(\widehat{C})\eta_n, \qquad \eta_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \Gamma), \quad 1 \le n \le N. \end{aligned}$$

The form of the update is similar to the Kalman mean update (2.5) albeit with the n-th ensemble member being assigned a PO $y - \eta_n$. Consequently, denoting the sample mean of the perturbations by

$\bar{\eta} \equiv N^{-1} \sum_{n=1}^{N} \eta_n$, the updated ensemble has the sample mean

$$\widehat{\mu} \equiv \frac{1}{N} \sum_{n=1}^{N} \upsilon_n = \mathscr{M}(\widehat{m}, \widehat{C}) - \mathscr{K}(\widehat{C})\bar{\eta},$$

and sample covariance

$$\begin{aligned} \widehat{\Sigma} &\equiv \frac{1}{N-1} \sum_{n=1}^{N} (\upsilon_n - \widehat{\mu})(\upsilon_n - \widehat{\mu})^\top \\ &= \big(I - \mathscr{K}(\widehat{C})A\big)\widehat{C}\big(I - \mathscr{K}(\widehat{C})A\big)^\top + \mathscr{K}(\widehat{C})\widehat{\Gamma}\mathscr{K}^\top(\widehat{C}) \\ &\qquad - \big(I - \mathscr{K}(\widehat{C})A\big)\widehat{C}^{u\eta}\mathscr{K}^\top(\widehat{C}) - \mathscr{K}(\widehat{C})(\widehat{C}^{u\eta})^\top\big(I - A^\top\mathscr{K}^\top(\widehat{C})\big), \end{aligned} \tag{2.6}$$

where

$$\widehat{\Gamma} \equiv \frac{1}{N-1} \sum_{n=1}^{N} (\eta_n - \bar{\eta}_N)(\eta_n - \bar{\eta})^\top, \qquad \text{and} \qquad \widehat{C}^{u\eta} \equiv \frac{1}{N-1} \sum_{n=1}^{N} (u_n - \widehat{m})(\eta_n - \bar{\eta})^\top.$$

To facilitate comparison with the Kalman update in (2.5), we rewrite the PO update as follows:

$$\begin{aligned} \widehat{\mu} &= \mathscr{M}(\widehat{m}, \widehat{C}) - \mathscr{K}(\widehat{C})\bar{\eta}, \\ \widehat{\Sigma} &= \mathscr{C}(\widehat{C}) + \widehat{O}, \end{aligned} \tag{2.7}$$

where the *offset* term $\widehat{O}$, obtained as the difference between (2.6) and $\mathscr{C}(\widehat{C})$, is given by

$$\widehat{O} = \mathscr{K}(\widehat{C})(\widehat{\Gamma} - \Gamma)\mathscr{K}^\top(\widehat{C}) - \big(I - \mathscr{K}(\widehat{C})A\big)\widehat{C}^{u\eta}\mathscr{K}^\top(\widehat{C}) - \mathscr{K}(\widehat{C})(\widehat{C}^{u\eta})^\top\big(I - A^\top\mathscr{K}^\top(\widehat{C})\big). \tag{2.8}$$

The offset term $\widehat{O}$ was introduced in [36, Proposition 4]. The addition of perturbations serves the purpose of correcting the sample covariance, in the sense that without perturbations the sample covariance is an inconsistent estimator of Σ. To see the consistency of the PO covariance estimator $\widehat{\Sigma}$ in (2.7), note that from Lemma A.6 the map $\mathscr{C}$ is continuous, and so the continuous mapping theorem together with the fact that $\widehat{C}$ is consistent for C imply that $\mathscr{C}(\widehat{C}) \xrightarrow{p} \mathscr{C}(C) = \Sigma$. Further, the offset $\widehat{O}$ converges in probability to zero, which can be shown using that $\widehat{\Gamma} \xrightarrow{p} \Gamma$, $\widehat{C}^{u\eta} \xrightarrow{p} O_{d\times k}$ and the continuity of $\mathscr{K}$ established in Lemma A.4.

2.1.2 *Square Root Update* The PO update relies crucially on the added perturbations to maintain consistency and, as noted, for example, in [11,32,93], is asymptotically equivalent to the exact posterior update (1.2). However, for a finite ensemble of size N, the addition of random perturbations introduces an extra source of error into the ensemble Kalman update. The SR update, introduced in [32] and surveyed in [61,93], is a deterministic alternative to the PO update. It updates the prior ensemble in a manner that ensures that $\widehat{\Sigma} \equiv \mathscr{C}(\widehat{C})$. This is achieved by first identifying a map $g : \mathbb{R}^{d\times N} \to \mathbb{R}^{d\times N}$ such that

$\widehat{\Pi} = g(\widehat{P})$, where

$$\widehat{C} = \widehat{P}\widehat{P}^\top, \qquad \text{and} \qquad \mathscr{C}(\widehat{C}) = \widehat{\Pi}\widehat{\Pi}^\top,$$

with both factorizations guaranteed to exist since $\widehat{C}, \mathscr{C}(\widehat{C}) \in \mathcal{S}_+^d$. Consistency of $\widehat{\Sigma}$ can then be ensured by choosing g to satisfy $g(\widehat{P})g(\widehat{P})^\top \equiv \mathscr{C}(\widehat{C})$, with this being referred to as the *consistency condition* in [61]. There are infinitely many such g, each of which lead to a variant of the SR update. Here we describe two of the most popular variants in the literature as outlined in [93]: the Ensemble Transform Kalman update [11] and the Ensemble Adjustment Kalman update [2] with respective transformations $g_T(\widehat{P}) = \widehat{P}T$ and $g_A(\widehat{P}) = B\widehat{P}$, for matrices T and B. Both g_T and g_A are therefore linear maps, with g_T post-multiplying $\widehat{P}$, which implies a transformation on the N-dimensional space spanned by the ensemble, and g_A pre-multiplying $\widehat{P}$, so that the transformation is applied to the d-dimensional state-space instead. In both approaches we identify the relevant matrix by first writing

$$\widehat{\Pi}\widehat{\Pi}^\top = \mathscr{C}(\widehat{C}) = \widehat{P}(I - VD^{-1}V^\top)\widehat{P}^\top,$$

where $V = (A\widehat{P})^\top$ and $D = V^\top V + \Gamma$.

1. *Ensemble Transform Kalman update:* taking $\widehat{\Pi} = \widehat{P}FU$ for any F satisfying $FF^\top = I - VD^{-1}V^\top$ and arbitrary orthogonal U satisfies the consistency condition. One approach for finding such a matrix F is by rewriting

$$I - VD^{-1}V^\top = (I + \widehat{P}^\top A^\top \Gamma^{-1} A\widehat{P})^{-1} = E(I + \Lambda)^{-1/2}(I + \Lambda)^{-1/2}E^\top = FF^\top,$$

   where the first equality follows by the Sherman–Morrison formula, and $E\Lambda E^\top$ is the eigenvalue decomposition of $\widehat{P}^\top A^\top \Gamma^{-1} A\widehat{P}$. In summary, we have $g_T(\widehat{P}) = \widehat{P}E(I + \Lambda)^{-1/2}U$.

2. *Ensemble Adjustment Kalman update:* Introducing $M = V\Gamma^{-1/2}$, we can write

$$\widehat{P}(I - VD^{-1}V^\top)\widehat{P}^\top = \widehat{P}(I + MM^\top)^{-1}\widehat{P}^\top.$$

   Noting that $\widehat{P}$ has full column rank, we may then define $B = \widehat{P}(I + MM^\top)^{-1/2}(\widehat{P}^\top)^\dagger$, and so

$$g_A(\widehat{P}) = B\widehat{P} = \widehat{P}(I + MM^\top)^{-1/2}(\widehat{P}^\top)^\dagger\widehat{P} = \widehat{P}(I + MM^\top)^{-1/2}.$$

Once a choice of g has been made, and an estimate $\widehat{\Sigma}$ has been computed, the updated ensemble has first two moments given by

$$\begin{aligned} \widehat{\mu} &= \mathscr{M}(\widehat{m}, \widehat{C}), \\ \widehat{\Sigma} &= \mathscr{C}(\widehat{C}). \end{aligned} \tag{2.9}$$

Frequently, only $\widehat{\mu}, \widehat{\Sigma}$ are of concern to the practitioner, but it is still possible to *back-out* the individual members of the updated ensemble as they may be of interest. It is clear that one choice for $\widehat{P}$ is

$$\widehat{P} = \frac{1}{\sqrt{N-1}}\left[u_1 - \widehat{m}, \cdots, u_N - \widehat{m}\right],$$

in which case it holds that $\widehat{P}1_N = 0_d$, and so

$$\upsilon_n = \sqrt{N-1}[\widehat{\Pi}]_n + \mathscr{M}(\widehat{m}, \widehat{C}), \qquad 1 \le n \le N, \tag{2.10}$$

where $[\widehat{\Pi}]_n$ denotes the n-th column of $\widehat{\Pi}$.

In Subsection 2.3, we establish error bounds for the approximation of the posterior mean and covariance (μ, Σ) in (1.2) by $(\widehat{\mu}, \widehat{\Sigma})$ as estimated using the PO and SR updates in (2.7) and (2.9). It is clear from (2.9) that as long as the choice of g is valid, in the sense that the resulting $\widehat{\Sigma}$ is consistent, then the specific choice of g is irrelevant to the accuracy of a single SR update. We therefore make no assumptions in our subsequent analysis of the SR algorithm beyond that of g satisfying the consistency condition. Note that, when compared to the SR update in (2.9), the PO update in (2.7) contains additional stochastic terms that will, as our bounds indicate, hinder the estimation of (μ, Σ). As noted in the literature, for example in [93], the PO update increases the probability of underestimating the analysis error covariance. While our presentation and analysis of PO and SR updates is carried out in the linear-Gaussian setting, both updates are frequently utilized in nonlinear and non-Gaussian settings, with empirical evidence suggesting that the PO updates can outperform SR updates [64,67]. In fact, the consistency argument outlined above is only valid in the linear case $\mathcal{G}(u) = Au$, and the statistical advantage of SR implementations in linear settings may not translate into the nonlinear case.

2.2 *Dimension-free covariance estimation*

We define the *effective dimension* [103] of a matrix $Q \in \mathcal{S}_+^d$ by

$$r_2(Q) \equiv \frac{\mathrm{Tr}(Q)}{\|Q\|}. \tag{2.11}$$

The effective dimension quantifies the number of directions where Q has significant spectral content [99]. The monographs [99,102] refer to $r_2(Q)$ as the intrinsic dimension, while [57] 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 [99]. In situations where the eigenvalues of Q decay rapidly, $r_2(Q)$ is a better measure of dimension than the state dimension d. The following result [57, Theorem 9] gives a non-asymptotic sufficient sample size requirement for accurate covariance estimation in terms of the effective dimension of the covariance matrix. We recall that the sample covariance estimator $\widehat{C}$ is defined in (1.3).

PROPOSITION 2.1. (Covariance Estimation with Sample Covariance—Unstructured Case) Let $u_1, \ldots, u_N$ be d-dimensional i.i.d. sub-Gaussian random vectors with $\mathbb{E}[u_1] = m$ and $\mathrm{var}[u_1] = C$. Then, for all

$t \geq 1$, it holds with probability at least $1 - ce^{-t}$ that

$$\|\widehat{C} - C\| \lesssim \|C\| \left(\sqrt{\frac{r_2(C)}{N}} \vee \frac{r_2(C)}{N} \vee \sqrt{\frac{t}{N}} \vee \frac{t}{N} \right).$$

REMARK 2.2. (Effective Dimension and Smoothness) Proposition 2.1 motivates defining $r_2(C) \equiv \mathrm{Tr}(C)/\|C\|$ to be the effective dimension of a d-dimensional sub-Gaussian random vector u with $\mathrm{var}[u] = C$. As in the definition for matrices, $r_2(C)$ quantifies the number of directions where the distribution of u has significant spread. Proposition 2.1 and our results in Subsection 2.3 may be extended to sub-Gaussian random variables defined in an infinite-dimensional separable Hilbert space, say $\mathcal{H} = L^2(0, 1)$. It is then illustrative to note that any Gaussian measure $\mathcal{N}(m, C)$ in $\mathcal{H}$ satisfies that $\mathrm{Tr}(C) < \infty$; in other words, all Gaussian measures have finite effective dimension. In this context, $r_2(C)$ is related to the rate of decay of the eigenvalues of C, and hence to the almost sure Sobolev regularity of functions u drawn from the Gaussian measure $\mathcal{N}(m, C)$ on $\mathcal{H} = L^2(0, 1)$, see e.g. [14, 91]. In computational inverse problems and data assimilation, u is often a d-dimensional vector that represents a fine discretization of a Gaussian random field; then, $r_2(C)$ quantifies the smoothness of the undiscretized field.

2.3 *Main results: posterior approximation with finite ensemble*

In this subsection, we state finite ensemble approximation results for the posterior mean and covariance with PO and SR ensemble updates. To highlight some key insights, including the dependence of the bounds on the effective dimension of C and the differences between PO and SR updates, we opt to present expectation bounds in Theorems 2.3 and 2.5 that are less notationally cumbersome than the stronger exponential tail bounds in Theorems A.8 and A.9 in Appendix A.3. Throughout this section, the data y are treated as a fixed quantity.

THEOREM 2.3. (Posterior Mean Approximation with Finite Ensemble—Expectation Bound) Consider the PO and SR ensemble Kalman updates given by (2.7) and (2.9), respectively, leading to an estimate $\widehat{\mu}$ of the posterior mean μ defined in (1.2). Set $\phi = 1$ for the PO update and $\phi = 0$ for the SR update. Then, for any $p \geq 1$,

$$\left[\mathbb{E}\|\widehat{\mu} - \mu\|_2^p\right]^{1/p} \lesssim_p c_1 \left(\sqrt{\frac{r_2(C)}{N}} \vee \left(\frac{r_2(C)}{N}\right)^{3/2} \right) + \phi c_2 \left(\sqrt{\frac{r_2(\Gamma)}{N}} \vee \frac{r_2(C)}{N} \sqrt{\frac{r_2(\Gamma)}{N}} \right), \tag{2.12}$$

where $c_1 = c_1(\|C\|, \|A\|, \|\Gamma^{-1}\|, \|y - Am\|_2)$ and $c_2 = c_2(\|C\|, \|A\|, \|\Gamma^{-1}\|)$.

Importantly, the bound (2.12) does not depend on the dimension d of the state-space, and the only dependence on C is through its operator norm and the effective dimension $r_2(C)$. The term multiplied by ϕ in the PO update accounts for the additional error incurred by the presence of the offset term (2.8) in the PO update (2.7). The following remark discusses another important consequence of Theorem 2.3: the stable performance of ensemble Kalman updates in small noise regimes when compared with other sampling algorithms.

REMARK 2.4. (Dependence of Constants on Model Parameters) The proof of Theorem 2.3 in Appendix A provides an explicit definition of c_1 and c_2 up to constants, i.e. it describes how these quantities rely on their arguments, and Theorem A.8 establishes a high probability bound on $\|\widehat{\mu} - \mu\|_2$. In particular, it

is important to note that the constants c_1 and c_2 in Theorem 2.3 deteriorate in the small noise limit where the observation noise goes to zero, and c_2 deteriorates with $r_2(\Gamma)$. In the small noise limit, the posterior and prior distribution become mutually singular, and it is hence expected for ensemble updates to be unstable. To illustrate this intuition in a concrete setting, assume that $\Gamma = \gamma I$ for a positive constant γ, and, for simplicity, that $N \geq r_2(C)$ as well as $\|C\| = \|A\| = \|y - Am\|_2 = 1$. Then, the expression for c_1 established in Theorem 2.3 implies that for the SR update, for any error $\varepsilon > 0$ and $p \geq 1$,

$$N \gtrsim \frac{r_2(C)}{\varepsilon^2 \gamma^4} \implies \mathbb{E}\big[\|\widehat{\mu} - \mu\|_2^p\big]^{1/p} \lesssim_p \varepsilon.$$

Similarly, the expressions for c_1 and c_2 imply that for the PO update,

$$N \gtrsim \frac{r_2(C)}{\varepsilon^2 \gamma^4} \vee \frac{k}{\varepsilon^2 \gamma} \implies \mathbb{E}\big[\|\widehat{\mu} - \mu\|_2^p\big]^{1/p} \lesssim_p \varepsilon,$$

where we recall that k denotes the dimension of the data y. The papers [1, 84] show the need to increase the sample size along small noise limits in importance sampling when target and proposal are given, respectively, by posterior and prior. While our bounds here only give sufficient rather than necessary conditions on the ensemble size, it is noteworthy that, for fixed k, the scaling of N as $\gamma \to 0$ shown here is independent of k. In contrast, necessary sample size conditions for importance sampling show a polynomial dependence on k, see [84].

THEOREM 2.5. (Posterior Covariance Approximation with Finite Ensemble—Expectation Bound) Consider the PO and SR ensemble Kalman updates given by (2.7) and (2.9), respectively, leading to an estimate $\widehat{\Sigma}$ of the posterior covariance Σ defined in (1.2). Set $\phi = 1$ for the PO update and $\phi = 0$ for the SR update. Then, for any $p \geq 1$,

$$\big[\mathbb{E}\|\widehat{\Sigma} - \Sigma\|^p\big]^{1/p} \lesssim_p c_1 \left(\sqrt{\frac{r_2(C)}{N}} \vee \left(\frac{r_2(C)}{N} \right)^2 \right) + \phi \mathscr{E},$$

where

$$\mathscr{E} = c_2 \left(\sqrt{\frac{r_2(C)}{N}} \vee \left(\frac{r_2(C)}{N} \right)^3 \vee \left(\sqrt{\frac{r_2(\Gamma)}{N}} \vee \frac{r_2(\Gamma)}{N} \right) \left(1 \vee \left(\frac{r_2(C)}{N} \right)^2 \right) \right),$$

where $c_1 = c_1(\|C\|, \|A\|, \|\Gamma^{-1}\|)$ and $c_2 = c_2(\|C\|, \|A\|, \|\Gamma^{-1}\|, \|\Gamma\|)$.

As in Theorem 2.3, the bound in Theorem 2.5 does not depend on the dimension d of the state-space, and the dependence on C is through the operator norm and the effective dimension $r_2(C)$.

REMARK 2.6. (Dependence of Constants on Model Parameters) The proof of Theorem 2.5 in Appendix A provides an explicit definition of c_1 and c_2 up to constants and Theorem A.9 establishes a high probability bound on $\|\widehat{\Sigma} - \Sigma\|$. As discussed in Remark 2.4, these bounds may be used to establish sufficient ensemble size requirements in small noise limits and other singular limits of practical importance.

REMARK 2.7. (Comparison to the Literature) The results in this section complement many of the existing analyses of ensemble Kalman updates in the literature. In one direction, our Theorems 2.3 and 2.5 can be viewed in the context of [36, Section 3.4], which claims that for finite ensembles the SR filter is always more efficient than the PO filter, since the latter introduces additional variability through noisy perturbations of the data. Our results quantify this additional variability both in probability and in expectation. In [72], the authors put forward a non-asymptotic analysis of a multi-step EnKF augmented by a spectral projection step in which the Kalman gain matrix is projected onto the linear span of its leading eigenvalues exceeding a threshold level. They refer to the dimension d_{subs} of this subspace as the effective dimension and provide guarantees on the performance of the algorithm so long as the ensemble size scales with d_{subs}. In contrast, our one-step analysis does not require any augmentation of standard implementations (see e.g. [36]) of the ensemble update. They also employ (forecast) covariance inflation, which is a de-biasing technique standard in the literature; see, for example, [36, Section 5], which our results do not require. In another direction, our results can be directly compared with [59, Theorem 6.1], which states that for iteration t of the SR EnKF and for any $p \geq 1$,

$$\left[\mathbb{E}\|\hat{\mu}^{(t)} - \mu^{(t)}\|_2^p\right]^{1/p} \leq \frac{c(p,t)}{\sqrt{N}} \qquad \text{and} \qquad \left[\mathbb{E}\|\widehat{\Sigma}^{(t)} - \Sigma^{(t)}\|^p\right]^{1/p} \leq \frac{c(p,t)}{\sqrt{N}}, \tag{2.13}$$

where $\hat{\mu}^{(t)}$ and $\widehat{\Sigma}^{(t)}$ are the sample mean and covariance of the updated (analysis) ensemble at iteration t, and $\mu^{(t)}$ and $\Sigma^{(t)}$ are the corresponding Kalman Filter posterior mean and covariance, respectively. The term $c(p,t)$ that arises in both of their bounds denotes a constant that depends only on p, the iteration index t, and the norms of the non-random inputs of the algorithm, but do not depend on dimension or ensemble size. Importantly, they do not distinguish between settings with different effective dimensions as our bounds do. In Appendix A.4, we provide an explicit outline of the multi-step algorithm considered in their paper along with definitions of all quantities described here. As previously noted, our bounds cover the PO setting, whereas (2.13) is specific to the SR setting. In Appendix A.4, we also establish (see Corollary A.12) a simple extension of Theorems 2.3 and 2.5 to the multi-step SR setting, which shows that for any $p \geq 1$, iteration t, and assuming for simplicity that $N \gtrsim r_2(\Sigma^{(0)})$, then

$$\begin{aligned}
\left[\mathbb{E}\|\hat{\mu}^{(t)} - \mu^{(t)}\|_2^p\right]^{1/p} &\lesssim_p \sqrt{\frac{r_2(\Sigma^{(0)})}{N}} \times c(\{\|M^{(l)}\|, \|A^{(l)}\|, \|\Sigma^{(l-1)}\|, \|y^{(l)} - A^{(l)}m^{(l)}\|_2\}_{l=1}^t, \|\Gamma^{-1}\|), \\
\left[\mathbb{E}\|\widehat{\Sigma}^{(t)} - \Sigma^{(t)}\|^p\right]^{1/p} &\lesssim_p \sqrt{\frac{r_2(\Sigma^{(0)})}{N}} \times c(\{\|M^{(l)}\|, \|A^{(l)}\|, \|\Sigma^{(l-1)}\|\}_{l=1}^t, \|\Gamma^{-1}\|).
\end{aligned} \tag{2.14}$$

Our bounds therefore refine those in [59] as they explicitly capture the dependence on the state dimension through the effective dimension of the initial distribution, $r_2(\Sigma^{(0)})$. It follows then that in the case of the SR EnKF, it suffices to use an ensemble on the order of the effective dimension of $\Sigma^{(0)}$ multiplied by constants depending on the operator norms of the forward model matrices $\{\|A^{(l)}\|\}_{l=1}^t$, analysis covariance matrices $\{\|\Sigma^{(l)}\|\}_{l=1}^t$, inverse of the noise covariance, $\|\Gamma^{-1}\|$ and ℓ_2-norm of the model errors $\{\|y^{(l)} - A^{(l)}m^{(l)}\|_2\}_{l=1}^t$. We note that extensions to the multi-step setting for other variants of the EnKF that do not use SR updates may not follow as easily. In this direction, the recent work [39] studies a multi-step EnKF with PO updates, which incorporates an additional resampling step.

3. Ensemble Kalman updates: sequential optimization algorithms

In the optimization approach, the solution to the inverse problem (1.1) is found by minimizing an objective function. As discussed in [21], an entire suite of ensemble algorithms have been derived that differ in the choice of objective function and optimization scheme. In this subsection we introduce the Ensemble Kalman Inversion (EKI) algorithm [52] and a new localized implementation of EKI, which we call localized EKI (LEKI) following [97]. Both EKI and LEKI use an ensemble approximation of a Levenberg–Marquardt (LM) optimization scheme to minimize a data-misfit objective

$$\mathsf{J}(u) = \frac{1}{2}\|\Gamma^{-1/2}\big(y - \mathcal{G}(u)\big)\|_2^2, \tag{3.1}$$

which promotes fitting the data y. Before deriving EKI in Subsection 3.1.1 and LEKI in Subsection 3.1.2, we give some background that will help us interpret both methods as ensemble-based implementations of classical gradient-based LM schemes. The finite ensemble approximation of an idealized mean-field EKI update using EKI and LEKI updates will be studied in Subsection 3.3.

Recall that classical iterative optimization algorithms choose an initialization $u^{(0)}$ and set

$$u^{(t+1)} = u^{(t)} + w^{(t)}, \qquad t = 0, 1, \ldots, \tag{3.2}$$

until a pre-specified convergence criterion is met. Here, $w^{(t)}$ is some favorable direction determined by the optimization algorithm at iteration t, given the current estimate $u^{(t)}$. In the case that the inverse problem is ill-posed, directly minimizing (3.1) leads to a solution that over-fits the data. Then, implicit regularization can be achieved through the optimization scheme used to obtain the update $w^{(t)}$. Under the assumption that $r(u) \equiv y - \mathcal{G}(u)$ is differentiable, LM algorithm chooses $w^{(t)}$ by solving the constrained minimization problem

$$\min_w \mathsf{J}_t^{\text{lin}}(w) \quad \text{subject to} \quad \|C^{-1/2}w\|_2^2 \le \delta_l,$$

where

$$\mathsf{J}_t^{\text{lin}}(w) \equiv \frac{1}{2}\|Dr(u^{(t)})w + r(u^{(t)})\|_2^2,$$

and Dr denotes the Jacobian of r. The LM algorithm belongs to the class of trust region optimization methods, and it chooses each increment to minimize a linearized objective, $\mathsf{J}_t^{\text{lin}}$, but with the added constraint that the minimizer belongs to the ball $\{\|C^{-1/2}w\|^2 \le \delta_l\}$, in which we *trust* that the objective may be replaced by its linearization. Equivalently, $w^{(t)}$ can be viewed as the unconstrained minimizer of a regularized objective,

$$\min_w \mathsf{J}_t^{\text{U}}(w), \qquad \mathsf{J}_t^{\text{U}}(w) \equiv \mathsf{J}_t^{\text{lin}}(w) + \frac{1}{2\alpha_t}\|C^{-1/2}w\|_2^2, \tag{3.3}$$

where $\alpha_t > 0$ acts as a Lagrange multiplier.

We are interested in ensemble sequential-optimization algorithms, which instead of updating a single estimate $u^{(t)}$—as in (3.2)—propagate an *ensemble* of estimates. Ensemble-based optimization schemes

often rely on *statistical linearization* to avoid the computation of derivatives. Underpinning this idea [21,56,100] is the argument that if $\mathcal{G}(u) = Au$ were linear, then $\widehat{C}^{up} = \widehat{C}A^\top$, leading to the approximation in the general nonlinear case

$$D\mathcal{G}(u_n) \approx (\widehat{C}^{up})^\top \widehat{C}^\dagger \equiv G. \tag{3.4}$$

This approximation motivates the derivative-free label often attached to ensemble-based algorithms [58], and we note that they may be employed whenever computing $D\mathcal{G}(u)$ is expensive or when $\mathcal{G}$ is not differentiable. For the remainder, our analysis focuses on a single step of EKI and LEKI, and so we drop the iteration index t from our notation; we will use instead our previous terminology of prior ensemble and updated ensemble. Finally, similar to our presentation of posterior-approximation algorithms, our exposition is simplified by introducing the *nonlinear gain-update* operator $\mathscr{P}$,

$$\mathscr{P} : \mathbb{R}^{d\times k} \times \mathcal{S}_+^k \to \mathbb{R}^{d\times k}, \qquad \mathscr{P}(C^{up}, C^{pp}; \Gamma) = \mathscr{P}(C^{up}, C^{pp}) = C^{up}(C^{pp} + \Gamma)^{-1}, \tag{3.5}$$

which is shown to be both pointwise continuous and bounded in Lemma A.7.

3.1 *Ensemble algorithms for sequential optimization*

3.1.1 *EKI update*

In the EKI, each particle in the prior ensemble is updated according to the LM algorithm, so that

$$\upsilon_n = u_n + w_n, \qquad 1 \le n \le N,$$

where w_n is the minimizer of a linearized and regularized data-misfit objective

$$\mathsf{J}_n^{\text{lin}}(w) = \frac{1}{2}\|\Gamma^{-1/2}\big(y - \eta_n - \mathcal{G}(u_n) - Gw\big)\|_2^2 + \frac{1}{2\alpha}\|\widehat{C}^{-1/2}w\|_2^2, \qquad \eta_n \sim \mathcal{N}(0,\Gamma). \tag{3.6}$$

Following [52], we henceforth set $\alpha = 1$, but note that our main results can be readily extended to any $\alpha > 0$. Note that each ensemble member solves the optimization (3.6) with a POperturbed observation $y - \eta_n$, similar in spirit to the PO update of Subsection 2.1.1. The minimizer of (3.6) (with $\alpha = 1$) is given by

$$w_n = \widehat{C}G^\top(G\widehat{C}G^\top + \Gamma)^{-1}\big(y - \eta_n - \mathcal{G}(u_n)\big).$$

Substituting $\widehat{C}G^\top = \widehat{C}^{up}$, and approximating

$$G\widehat{C}G^\top = G\widehat{C}^{up} = (\widehat{C}^{up})^\top\widehat{C}^\dagger\widehat{C}^{up} \approx \widehat{C}^{pp}$$

leads to the EKI update

$$\upsilon_n = u_n + \mathscr{P}(\widehat{C}^{up}, \widehat{C}^{pp})\big(y - \mathcal{G}(u_n) - \eta_n\big), \qquad 1 \le n \le N. \tag{3.7}$$

In the linear forward-model setting, $\mathscr{P}(\widehat{C}^{up}, \widehat{C}^{pp}) = \mathscr{K}(\widehat{C})$, and (3.7) takes on a form identical to the PO update in (2.7). We further define the *mean-field* EKI update

$$\upsilon_n^* = u_n + \mathscr{P}(C^{up}, C^{pp})\big(y - \mathcal{G}(u_n) - \eta_n\big), \qquad 1 \le n \le N, \tag{3.8}$$

which is the update that would be performed if one had access to the population quantities C^{up} and C^{pp} or, equivalently, to an infinite ensemble. We will analyze the approximation of the update (3.7) to the mean-field update (3.8) in Subsection 3.3. The study of mean-field ensemble Kalman methods of the form (3.8) was proposed in [45] and is overviewed in [19]. While mean-field algorithms are not useful for practical implementation, they facilitate a transparent mathematical analysis that can provide understanding on the performance of practical ensemble approximations. Desirable properties of mean-field algorithms include convergence to the desired target in a continuous-time limit [20], a gradient flow structure [37] or the ability to approximate derivative-based optimization algorithms [21]. The transfer of theoretical insights from mean-field algorithms to particle-based algorithms tacitly presupposes, however, that the ensemble is large enough for ensemble-based updates to approximate well idealized mean-field updates. In this direction, [27] establishes a $\mathcal{O}(N^{-1/2})$ rate for an approximation of a mean-field evolution equation in terms of the ensemble size N. Our first main result of this section, Theorem 3.5, will show that the mean-field update (3.8) can be well approximated with the EKI update (3.7) with a number of particles of the order of the *effective dimension* of the problem, which is defined as for posterior-approximation algorithms.

3.1.2 *LEKI update* In practice, ensemble-based algorithms are often implemented with $N \ll d$, that is, with an ensemble that is much smaller than the state dimension. In this setting, the update is augmented with an additional *localization* procedure applied to $\widehat{C}$ in the case of linear forward model, and to both $\widehat{C}^{pp}$ and $\widehat{C}^{up}$ in the case of a nonlinear forward model. In either case, localization is seen as an approach to deal both with the extreme rank deficiency and the sampling error that arise from using an ensemble that is significantly smaller than the dimension of the state and/or the dimension of the observation; see, for example, [35,49,50]. Localization is also useful when the state u, or the transformed state $\mathcal{G}(u)$, has elements $\mathsf{E}(i)$ and $\mathsf{E}(j)$ that represent the values of a variable of interest at physical locations that are a known distance $\mathsf{d}(i,j)$ apart: correlations may decay quickly with the physical distance of the variables and localization may help to remove spurious correlations in the sample covariance estimator. In ensemble Kalman methods, localization has most commonly been carried out via the Schur (elementwise) product of the estimator and a positive-semidefinite matrix M of equal dimension. In the vast majority of cases, the elements of M are taken to be $\mathsf{M}_{ij} = \kappa(\mathsf{d}(i,j)/b)$, where κ is a locally supported correlation function—usually the Gaspari Cohn 5^{th}-order compact piecewise polynomial [38]—and $b > 0$ is a length-scale parameter chosen by the practitioner. Since κ tapers off to zero as its argument becomes larger, i.e. when the underlying variables are further apart, the Schur-product operation zeroes out the corresponding elements of the estimator, and the rate at which this tapering occurs is controlled by the size of the length-scale. The LEKI, recently studied in [97], replaces both $\widehat{C}^{pp}$ and $\widehat{C}^{up}$ with their localized counterparts, $\mathsf{M}_1 \circ \widehat{C}^{pp}$ and $\mathsf{M}_2 \circ \widehat{C}^{up}$, where M_1 and M_2 are localization matrices of appropriate dimension. Two important issues have, in our opinion, hindered the rigorous study of localized ensemble algorithms, and we highlight these next before moving on to introduce our localization framework:

1. Optimality: The justification outlined above for localization in the ensemble Kalman literature has been largely heuristic, and relying on these arguments alone one cannot hope to define

a localization procedure that is demonstrably optimal. Notably, the widespread usage of the Gaspari-Cohn correlation function is not rooted in any sense of optimality. Generally, focusing solely on a band of entries near the diagonal is a sub-optimal approach to covariance estimation, as noted in the high-dimensional covariance estimation literature; see, for example, [8,23,68]. Moreover, even in cases where focusing on elements near the diagonal is justified, for example by assuming that the underlying target is a banded matrix, the bandwidth $b > 0$ must be chosen carefully as a function of the ensemble size, problem dimension, and dependence structure [7]. This type of analysis has, to the best of our knowledge, not been carried out for the Gaspari–Cohn localization scheme. An important message in the covariance estimation literature is that localization—regardless of how it is employed—can only be optimal if the target of estimation itself is sparse, and such sparsity assumptions must be made explicit in order to facilitate a rigorous mathematical analysis of the procedure. The difficulty of optimal localization in ensemble updates has also been highlighted in [36], where the authors derive an optimal localization matrix M under the unrealistic assumption that C is a diagonal matrix.

2. Schur-Product Approximations: In the literature on ensemble Kalman methods, a consensus has not been reached on how best to apply localization in practice. The issue here can be sufficiently described by deferring to the linear forward-model setting, i.e. $\mathcal{G}(u) = Au$, in which the Kalman gain is a central quantity. As mentioned for example in [49], in a localized update, the Kalman gain operator should in theory be applied to $\mathsf{M} \circ \widehat{C}$, i.e. one should study the quantity

$$\mathscr{K}(\mathsf{M} \circ \widehat{C}) = (\mathsf{M} \circ \widehat{C})A^\top \big(A(\mathsf{M} \circ \widehat{C})A^\top + \Gamma\big)^{-1},$$

although their experimental results are based on the more computationally convenient approximation

$$\mathscr{K}(\mathsf{M} \circ \widehat{C}) \approx \big(\mathsf{M} \circ (\widehat{C}A^\top)\big)\big(\mathsf{M} \circ (A\widehat{C}A^\top) + \Gamma\big)^{-1}, \tag{3.9}$$

which, as they mention, is a reasonable approximation in the case that A is diagonal. Subsequently, much of the literature on localization in ensemble Kalman updates has adopted this or similar approximations, as discussed in greater depth in [78, Section 3.3]. In general, however, approximations made on the Schur product are difficult to justify without strong assumptions on the forward model $\mathcal{G}$.

With these issues in mind, we opt to study an alternative, data-driven approach to localization often employed in the high-dimensional covariance estimation literature [7,17,18], where it is referred to as *thresholding*. We ground our analysis in the assumption that the target of estimation belongs to the following soft sparsity matrix class:

$$\mathscr{U}_{d_1,d_2}(q, R_q) \equiv \left\{ B \in \mathbb{R}^{d_1 \times d_2} : \max_{i \le d_1} \sum_{j=1}^{d_2} |B_{ij}|^q \le R_q \right\}, \tag{3.10}$$

where $q \in [0, 1)$ and $R_q > 0$, and write $\mathscr{U}_d(q, R_q)$ in the case $d_1 = d_2 = d$. In the special case $q = 0$, matrices in $\mathscr{U}_{d_1,d_2}(0, R_0)$ possess rows that have no more than R_0 non-zero entries—a special case of which are banded matrices—which is the classical hard-sparsity constraint. In contrast, for $q \in (0, 1)$, the class $\mathscr{U}_{d_1,d_2}(q, R_q)$ contains matrices with rows belonging to the ℓ_q ball of radius R_q^q. This

includes matrices with rows that contain possibly many non-zero entries so long as their magnitudes decay sufficiently rapidly, and so is often referred to as a *soft*-sparsity constraint. Importantly, the class $\mathscr{U}_d(q, R_q)$ is sufficiently rich to capture the motivating intuition that correlations decay with physical distance in a rigorous manner that avoids the optimality issues mentioned above. Structured covariance matrices, such as those belonging to $\mathscr{U}_{d_1,d_2}(q, R_q)$ are optimally estimated using localized versions of their sample covariances. To this end, we study the localized matrix estimator $B_{\rho_N} \equiv \mathcal{L}_{\rho_N}(B)$, where $\mathcal{L}_{\rho_N}(u) = u\mathbf{1}_{\{|u|\geq\rho_N\}}$ is a localization operator with localization radius ρ_N, and which is applied elementwise to its argument B. In Section 3, we detail how the localization radius ρ_N can be chosen optimally in terms of the parameters of the inverse problem (1.1) and the ensemble size N.

Throughout our analysis, we refrain from using approximations such as the one outlined in (3.9); that is, our analysis of localization replaces all non-localized quantities in the original update (3.7) with their localized counterparts. We introduce the LEKI update:

$$\upsilon_n^\rho = u_n + \mathscr{P}(\widehat{C}^{up}_{\rho_{N,1}}, \widehat{C}^{pp}_{\rho_{N,2}})\big(y - \mathcal{G}(u_n) - \eta_n\big), \qquad 1 \leq n \leq N, \tag{3.11}$$

where $\rho_{N,1}$ and $\rho_{N,2}$ are two, potentially different localization radii. As in the non-localized case, in Subsection 3.3 we provide finite sample bounds on the deviation of the LEKI update from the mean-field update of (3.8), and describe in detail how the additional structure imposed on C^{up} and C^{pp} leads to improved bounds relative to the non-localized setting. Our second main result of this section, Theorem 3.7, will be based on new covariance estimation bounds that may be of independent interest, and on a suitable notion of effective dimension that we introduce in Subsection 3.2. Our theory explains the improved sample complexity that can be achieved by simultaneously exploiting spectral decay and sparsity of the covariance model.

An important issue that warrants discussion is that of positive-semidefiniteness of the estimator $\widehat{B}_{\rho_N}$ when the target B is a square covariance matrix. In the case of the Schur-product estimator, any localization matrix M derived from a valid correlation function κ is guaranteed to be positive-semidefinite by definition [38], and so by the Schur-product Theorem [46, Theorem 7.5.3] the estimator $\mathsf{M} \circ \widehat{B}$ is positive-semidefinite as well. In contrast, the localization operator $\mathcal{L}_{\rho_N}$ thresholds the sample covariance $\widehat{B}$ elementwise and does not in general preserve positive-semidefiniteness. As discussed in [18,29], $\widehat{B}_{\rho_N}$ is positive-semidefinite with high probability, but in practice one may opt to use an augmented estimator that guarantees positive-semidefiniteness. We describe this estimator here for completeness: let $\widehat{B}_{\rho_N} = \sum_{j=1}^d \hat{\lambda}_j v_j v_j^\top$ be the eigen-decomposition of $\widehat{B}_{\rho_N}$, so that λ_j, v_j are the jth eigenvalue and eigenvector of $\widehat{B}_{\rho_N}$. Consider then the positive-part estimator $\widehat{B}^+_{\rho_N} \equiv \sum_{j=1}^d (0 \vee \hat{\lambda}_j) v_j v_j^\top$. Clearly then, $\widehat{B}^+_{\rho_N}$ is positive-semidefinite, and furthermore it achieves the same rate as $\widehat{B}_{\rho_N}$ since

$$\|\widehat{B}^+_{\rho_N} - B\| \leq \|\widehat{B}^+_{\rho_N} - \widehat{B}_{\rho_N}\| + \|\widehat{B}_{\rho_N} - B\| \leq \max_{j:\hat{\lambda}_j<0} |\hat{\lambda}_j - \lambda_j| + \|\widehat{B}_{\rho_N} - B\| \leq 2\|\widehat{B}_{\rho_N} - B\|,$$

where λ_j is the jth eigenvalue of B. In light of this fact, we abuse notation slightly and assume that B_{ρ_N} is positive-semidefinite throughout this work.

3.2 *Dimension-free covariance estimation under soft sparsity*

For the covariance estimation problem under (approximate) sparsity, there are estimators that significantly improve upon the sample covariance. In particular, [103, Chapter 6.5] notes that for sub-Gaussian data the operator-norm covariance estimation error depends logarithmically on the state dimension d

for localized estimators, while the error depends linearly on d for the sample covariance. If no sparse structure is assumed, the effective dimension r_2 defined in (2.11) characterizes the error of the sample covariance estimator, as described in Proposition 2.1. We introduce an analogous notion of effective dimension that is more suitable than r_2 in the sparse covariance estimation problem, termed the *max-log effective dimension* and which, for $Q \in \mathcal{S}_+^d$, is given by

$$r_\infty(Q) \equiv \frac{\max_{j \le d} Q_{(j)} \log(j+1)}{Q_{(1)}},$$

where $Q_{(1)} \ge Q_{(2)} \ge \ldots \ge Q_{(d)}$ is the decreasing rearrangement of the diagonal entries of Q. To the best of our knowledge, this notion of dimension has not been previously considered in the literature, and, as will be shown, refines the rate of covariance estimation under sparsity by incorporating intrinsic properties of the underlying matrix, albeit differently to (2.11). In particular, $r_\infty(Q)$ is small whenever Q exhibits a decay of the ordered elements $Q_{(1)}, Q_{(2)}, \ldots$ that is faster than $\log(j+1)$. We use the subscript ∞ to highlight that the quantity r_∞ is related to the dimension-free sub-Gaussian maxima result of Lemma B.6. Similarly, we use the subscript 2 to draw the connection between r_2 and the sub-Gaussian 2-norm concentration of Theorem A.1. Importantly, bounds based on r_∞ will be dimension-free, in the sense that they exhibit no dependence on the state dimension d. The next result is our analog of Proposition 2.1 for estimation under sparsity using the localized sample covariance estimator. Recall that $C_{(1)}$ denotes the largest element on the diagonal of C, $\widehat{C}_{\rho_N} \equiv \mathcal{L}_{\rho_N}(\widehat{C})$ denotes the localized sample covariance matrix and $\mathscr{U}_d(q, R_q)$ is the sparse matrix class defined in (3.10). All proofs in this subsection have been deferred to Appendix B.1.

THEOREM 3.1. (Covariance Estimation with Localization—Soft Sparsity Assumption) Let $u_1, \ldots, u_N$ be d-dimensional i.i.d. sub-Gaussian random vectors with $\mathbb{E}[u_1] = m$ and $\mathrm{var}[u_1] = C$. Further, assume that $C \in \mathscr{U}_d(q, R_q)$ for some $q \in [0, 1)$ and $R_q > 0$. For any $t \ge 1$, set

$$\rho_N \asymp C_{(1)} \left(\sqrt{\frac{r_\infty(C)}{N}} \vee \frac{t r_\infty(C)}{N} \vee \sqrt{\frac{t}{N}} \vee \frac{t}{N} \right)$$

and let $\widehat{C}_{\rho_N} \equiv \mathcal{L}_{\rho_N}(\widehat{C}_{\rho_N})$ be the localized sample covariance estimator. There exists a constant $c > 0$ such that, with probability at least $1 - ce^{-t}$,

$$\|\widehat{C}_{\rho_N} - C\| \lesssim R_q \rho_N^{1-q}.$$

REMARK 3.2. (Max-Log Effective Dimension) The proof of Theorem 3.1 can be found in Section B.1 and, up to the choice of ρ_N, follows an identical approach to the standard proof for localized covariance estimators in the literature, for example [103, Theorem 6.27]. The result depends crucially on the order of the maximum elementwise distance between the sample and true covariance matrices, $\|\widehat{C} - C\|_{\max}$, which is where our analysis differs from the exiting literature. Our proof utilizes techniques in [57] combined with the dimension-free sub-Gaussian maxima bound of Lemma B.6 to obtain a bound in terms of r_∞. In the worst case, for example when $C = cI_d$ for some constant $c > 0$ so that the ordered diagonal elements of C exhibit no decay, we recover exactly the standard logarithmic dependence on the state dimension. In particular, when $N \ge r_\infty(C) (= \log d)$, Theorem 3.1 matches the result for recovering

C in operator norm in the sub-Gaussian setting over the class $\mathcal{U}_d(q, R_q)$, as shown in [7, Theorem 1]. If the ordered variances exhibit sufficiently fast decay, our upper bound is significantly better. (Recall that in many applications $d \sim 10^9$ and $N \sim 10^2$, and so the logarithmic dependence on d may play a significant role in determining a sufficient ensemble size.) Importantly, many of the results in the structured covariance estimation literature rely similarly on the maximum elementwise norm, and so our results can be utilized to achieve refined bounds on the estimation error of the localized estimator under structural assumptions on C that differ from the soft-sparsity assumption considered in this work.

A result analogous to Theorem 3.1 holds for cross-covariance estimation under sparsity. For a formal statement we refer to Theorem B.11 whose proof is based on a deep generic chaining bound for product empirical processes [74, Theorem 1.12]. Here we present a cross-covariance estimation result that is specific to the LEKI setting in that it relies on a smoothness assumption on the forward model.

THEOREM 3.3. (Cross-Covariance Estimation with Localization—Soft Sparsity Assumption) Let $u_1, \dots, u_N$ be d-dimensional i.i.d. sub-Gaussian random vectors with $\mathbb{E}[u_1] = m$ and $\mathrm{var}[u_1] = C$. Let $\mathcal{G} : \mathbb{R}^d \to \mathbb{R}^k$ be a Lipschitz continuous forward model and assume that $C^{up} \in \mathcal{U}_{d,k}(q_1, R_{q_1})$ and $C^{pu} \in \mathcal{U}_{k,d}(q_2, R_{q_2})$ where $q_1, q_2 \in [0, 1)$ and R_{q_1}, R_{q_2} are positive constants. For any $t \geq 1$, set

$$\rho_N \asymp (C_{(1)} \vee C^{pp}_{(1)}) \left(\left(\frac{t}{N} \vee \sqrt{\frac{t}{N}} \right) \left(\sqrt{r_\infty(C)} \vee \sqrt{r_\infty(C^{pp})} \right) \vee \sqrt{\frac{r_\infty(C)}{N}} \sqrt{\frac{r_\infty(C^{pp})}{N}} \right),$$

and let $\widehat{C}^{up}_{\rho_N} \equiv \mathcal{L}_{\rho_N}(\widehat{C}^{up})$ be the localized sample cross-covariance estimator. There exist positive universal constants c_1, c_2 such that, with probability at least $1 - c_1 e^{-c_2 t}$,

$$\|\widehat{C}^{up}_{\rho_N} - C^{up}\| \lesssim R_{q_1} \rho_N^{1-q_1} \vee R_{q_2} \rho_N^{1-q_2}.$$

REMARK 3.4. (Sparsity of the Cross-Covariance) To the best of our knowledge, estimation of the cross-covariance matrix under structural assumptions has not been a point of focus in the literature. Indeed, one may implicitly estimate the cross-covariance by applying Theorem 3.1 to the full covariance matrix

$$\begin{bmatrix} C & C^{up} \\ C^{pu} & C^{pp} \end{bmatrix}$$

of the sub-Gaussian vector $[u^\top, \mathcal{G}(u)^\top]^\top$, and extracting a bound on $\|C^{up}_{\rho_N} - C^{up}\|$. This approach however requires one to place sparsity assumptions on the full covariance matrix, making the result potentially less useful in practice. That is, one may wish to make structural assumptions on C^{up} and C^{pp} without imposing any restrictions on C, which our result allows for.

3.3 *Main results: approximation of mean-field particle updates with finite ensemble size*

In this subsection, we state finite ensemble approximation results for EKI and LEKI updates. The main results, Theorems 3.5 and 3.7, showcase the dependence on the effective dimension of C and C^{pp} for EKI and on the max-log dimension of these matrices for LEKI. For both algorithms, we study the update of a generic particle u_n and the analysis is carried out conditional on both u_n and the noise perturbation η_n.

THEOREM 3.5. (Approximation of Mean-Field EKI with EKI—Operator-Norm Bound) Let y be generated according to (1.1) with Lipschitz forward model $\mathcal{G} : \mathbb{R}^d \to \mathbb{R}^k$. Let υ_n and υ_n^* be the EKI and mean-field EKI updates defined in (3.7) and (3.8), respectively. Then, for any $t \geq 1$, there exists universal positive constants c_1, c_2 such that, with probability at least $1 - c_1 e^{-c_2 t}$,

$$\|\upsilon_n - \upsilon_n^*\|_2 \leq c_1 \left(\frac{c_2}{N} \vee \sqrt{\frac{r_2(C)}{N}} \vee \frac{r_2(C)}{N} \vee \sqrt{\frac{r_2(C^{pp})}{N}} \vee \frac{r_2(C^{pp})}{N} \vee \sqrt{\frac{t}{N}} \vee \frac{t}{N} \right),$$

where $c_1 = c_1(\|y - \mathcal{G}(u_n) - \eta_n\|_2, \|\Gamma^{-1}\|, \|C\|, \|C^{up}\|, \|C^{pp}\|)$ and for $u \sim \mathcal{N}(m, C)$, $c_2 = c_2(\|u_n\|_2, \|m\|_2, \|\mathcal{G}(u_n)\|_2, \|\mathbb{E}[\mathcal{G}(u)]\|_2)$.

REMARK 3.6. (Dependence of Constants on Model Parameters) The proof of Theorem 3.5 in Appendix B.2 gives an explicit expression for the dependence of c on its arguments. These bounds may be used to establish the sufficient ensemble size to ensure that the EKI update approximates well the mean-field EKI update in the unstructured covariance setting.

THEOREM 3.7. (Approximation of Mean-Field EKI with LEKI—Operator-Norm Bound) Let y be generated according to (1.1) with Lipschitz forward model $\mathcal{G} : \mathbb{R}^d \to \mathbb{R}^k$. Assume that $C^{up} \in \mathscr{U}_{d,k}(q_1, R_{q_1})$, $C^{pu} \in \mathscr{U}_{k,d}(q_2, R_{q_2})$ and $C^{pp} \in \mathscr{U}_k(q_3, R_{q_3})$ for $q_1, q_2, q_3 \in [0, 1)$, and positive constants $R_{q_1}, R_{q_2}, R_{q_3}$. Let υ_n^ρ and υ_n^* be the LEKI and mean-field EKI updates outlined in (3.11) and (3.8) respectively. For any $t \geq 1$, set

$$\rho_{N,1} = \rho_{N,2} \asymp \frac{c_1}{N} + (C_{(1)} \vee C_{(1)}^{pp}) \left(\left(\frac{t}{N} \vee \sqrt{\frac{t}{N}} \right) \left(\sqrt{r_\infty(C)} \vee \sqrt{r_\infty(C^{pp})} \right) \vee \sqrt{\frac{r_\infty(C)}{N}} \sqrt{\frac{r_\infty(C^{pp})}{N}} \right),$$

and

$$\rho_{N,3} \asymp \frac{c_2}{N} + C_{(1)}^{pp} \left(\sqrt{\frac{r_\infty(C^{pp})}{N}} \vee \sqrt{\frac{t}{N}} \vee \frac{t}{N} \vee \frac{t r_\infty(C^{pp})}{N} \right),$$

where $c_1 = c_1(\|u_n\|_\infty, \|m\|_\infty, \|\mathcal{G}(u_n)\|_\infty, \|\mathbb{E}[\mathcal{G}(u)]\|_\infty)$ and $c_2 = c_2(\|\mathcal{G}(u_n)\|_\infty, \|\mathbb{E}[\mathcal{G}(u)]\|_\infty)$, with $u \sim \mathcal{N}(m, C)$. There exist positive universal constants c_3, c_4 such that, with probability at least $1 - c_3 e^{-c_4 t}$,

$$\|\upsilon_n^\rho - \upsilon_n^*\|_2 \leq c_5 (R_{q_1} \rho_{N,1}^{1-q_1} \vee R_{q_2} \rho_{N,2}^{1-q_2} \vee R_{q_3} \rho_{N,3}^{1-q_3}),$$

where $c_5 = c_5(\|y - \mathcal{G}(u_n) - \eta_n\|_2, \|\Gamma^{-1}\|, \|C^{up}\|)$.

REMARK 3.8. (Dependence of Constants on Model Parameters) The proof of Theorem 3.7 in Appendix B.2 gives an explicit expression for the dependence of c on its arguments. As discussed in Remark 3.6, these bounds may be used to establish the sufficient ensemble size to ensure that the LEKI update approximates well the mean-field EKI update in the structured covariance setting.

REMARK 3.9. (On the Soft-Sparsity Assumptions) Importantly, Theorem 3.7 makes no assumptions on the covariance matrix C, and so can be used even in cases where C is dense, but the covariances C^{up}, C^{pu}, and C^{pp} can be reasonably assumed to be sparse. In the case that sparsity assumptions on C are appropriate, then an interesting question is: what (explicit) assumptions on $\mathcal{G}$ ensure sparsity of C^{up}, C^{pu} and C^{pp}? We provide here two simple arguments that may provide some insight. Throughout, c_1, c_2, c_2, c_4, c_5 are arbitrary positive constants independent of both state and observation dimensions d and k, and $q \in [0, 1)$:

1. Suppose $C \in \mathcal{U}_d(q, c_1)$ and $\mathbb{E}[D\mathcal{G}]^\top \in \mathcal{U}_{d,k}(q, c_2)$. Then there exists c_3 such that $C^{up} \in \mathcal{U}_{d,k}(q, c_3)$. We provide a formal statement of this result in Lemma B.14. Similarly, if $\mathbb{E}[D\mathcal{G}] \in \mathcal{U}_{k,d}(q, c_4)$, then there exists c_5 such that $C^{pu} \in \mathcal{U}_{k,d}(q, c_5)$. The assumptions on the expected Jacobian $\mathbb{E}[D\mathcal{G}]$ can be understood as the requirements that, in expectation:
   a. Any coordinate function $\mathcal{G}_j$ of $\mathcal{G}$ depends on its input u only through a subset of u whose size does not grow with k nor d.
   b. Any state coordinate u_j of u is acted on only by a subset of the coordinate-functions of $\mathcal{G}$ whose size does not grow with k nor d.
   For example, a Jacobian that is banded in expectation would satisfy these two properties.
2. Suppose $C \in \mathcal{U}_d(q, c_1)$. Then there exists c_2 such that $C^{pp} \in \mathcal{U}_k(q, c_2)$ whenever $\mathcal{G}(u) = Au$ is a linear map with $A \in \mathcal{U}_{k,d}(q, c_3)$ and $A^\top \in \mathcal{U}_{d,k}(q, c_4)$, i.e. whenever A has both rows and columns that are sparse. This condition holds, for example, for banded A. We provide a formal statement of this result in Lemma B.16.

The two arguments above indicate that if $\mathcal{G}$ acts on local subsets of u, which holds for instance for convolution or moving average operators, then one can expect the sparsity of C to carry on to C^{up}, C^{pu}, and C^{pp}.

REMARK 3.10. (Comparison to the Literature) Although the focus of this subsection is the LEKI, it is useful to compare our Theorems 3.5 and 3.7 to existing results for the performance of ensemble based algorithms with localization. In this regard, our results are closest to those of [94], which shows that an ensemble that scales with the logarithm of the state dimension times a localization radius suffices for good performance of the localized EnKF (LEnKF). They study performance over multiple time steps and linear dynamics under a stability assumption which enforces control over the model matrices as well as a sparse ($q = 0$) structure of the underlying true covariance matrices. They consider *domain localization* whereas we study *covariance localization*. In contrast to our results, [94] employs covariance localization and utilizes a Schur-product localization scheme in which elements whose indices are beyond a certain bandwidth are set to zero, whereas we study localization via thresholding (recall our discussion comparing these two approaches in Subsection 3.1.2). Consequently, our required localization radius is in terms of the max-log effective dimension whereas theirs is in terms of the bandwidth of the underlying covariance matrix. Our results are dimension-free in that they do not rely on the state dimension d, and so as noted in Remark 3.2, our bounds can have significantly better than logarithmic dependence on dimension. Our setting also differs from [94] in that our dynamics are allowed to be nonlinear, and our prior ensemble can be sub-Gaussian as opposed to Gaussian. Related to this point is that the analysis in [94] does not account for noise introduced from adding perturbations to the ensemble update, which is justified by a law of large numbers argument; however in the non-asymptotic and nonlinear settings, it is likely that one must account for this noise especially when considering the covariance between the current ensemble and the perturbation noise at a given iteration of the algorithm.

We view it as an important avenue to extend the results of this subsection to a multi-step analysis, and a particularly important question is whether dimension-free control of the LEnKF can be rigorously shown utilizing a combination of our results and those of [94]. The LEKI has also been recently studied in [97] under a nonlinear, multi-step setting. The authors study convergence of the iterates to a global minimizer and the rate of collapse of the ensemble. They argue that localization is a remedy for the 'subspace property' of the EKI, which refers to the fact that ensembles at any given iteration live in the linear subspace spanned by the initial ensemble, which cannot capture the true state if $N < d$. Their analysis differs from ours in that they study the continuous-time setting whereas we analyze discrete time updates as implemented in practice. Further, while they discuss that the size of the ensemble may be much smaller than the state dimension, as well as illustrate this with simulations, they do not provide an explicit characterization of the sufficient ensemble size. Our results also show that the LEKI is close to the mean field version of the problem, which is not considered in their set-up. An interesting open question is whether the results of this section can be used in conjunction with results in [97] to provide a sufficient ensemble size for LEKI over multiple iterations.

4. Conclusions, discussion and future directions

This paper has introduced a non-asymptotic approach to the study of ensemble Kalman methods. Our theory explains why these algorithms may be accurate provided that the ensemble size is larger than a suitable notion of effective dimension, which may be dramatically smaller than the state dimension due to spectrum decay and/or approximate sparsity. Our non-asymptotic results in Section 2 tell apart PO and SR updates for posterior approximation, and our results in Section 3 demonstrate the potential advantage of using localization in sequential-optimization algorithms.

As discussed in Subsection 3.1.2, localization is also often used in posterior-approximation algorithms. For instance, one may define a localized PO update by

$$\begin{aligned} \widehat{\mu} &= \mathscr{M}(\widehat{m}, \widehat{C}_{\rho_N}) - \mathscr{K}(\widehat{C}_{\rho_N})\bar{\eta}, \\ \widehat{\Sigma} &= \mathscr{C}(\widehat{C}_{\rho_N}) + \widehat{O}_{\rho_N}, \end{aligned} \tag{4.1}$$

where $\widehat{O}_{\rho_N}$ is defined replacing $\widehat{C}$ with $\widehat{C}_{\rho_N}$ in (2.8). Similarly, one may define a localized SR update by

$$\begin{aligned} \widehat{\mu} &= \mathscr{M}(\widehat{m}, \widehat{C}_{\rho_N}), \\ \widehat{\Sigma} &= \mathscr{C}(\widehat{C}_{\rho_N}). \end{aligned} \tag{4.2}$$

It is then natural to ask if localized PO and SR updates can yield better approximation of the posterior mean and covariance than those without localization in Theorems 2.3 and 2.5. The answer for the posterior mean seems to be negative.

To see why, consider for intuition that we are given a random sample $X_1, \ldots, X_N$ from a normal distribution with mean μ^X and covariance Σ^X with the objective to estimate μ^X. Standard results, see e.g. [60, Example 1.14], show that the sample mean $\overline{X}$ is minimax optimal for ℓ_2-loss regardless of whether or not Σ^X is known. In other words, the minimax rate of estimating μ^X can be achieved without making use of information regarding Σ^X. It follows then that placing assumptions on Σ^X can lead to impressive improvements in the covariance estimation problem (as shown in Section 3) but cannot be expected to affect the mean estimation problem. Similarly, in our inverse problem setting, sparsity assumptions on the prior covariance C cannot be expected to translate into a better bound on $\|\widehat{\mu} - \mu\|_2$: this quantity

is a function of both the covariance deviation $\|\widehat{C}_{\rho_N} - C\|$ and the prior mean deviation $\|\widehat{m} - m\|_2$ and since the latter is unaffected it dominates the overall bound, yielding an error bound of the same order as that in Theorem 2.3. As discussed in Remark 2.7, a potential avenue for future investigation is to utilize techniques introduced in this manuscript to study alternative localization schemes in the posterior approximation setting, such as *domain localization* considered in [94]. In short, covariance localization as defined in (4.2) does not lead to improved bounds for the posterior-approximation problem.

Similar issues to those arising in the estimation of the posterior mean affect the analysis of the localized offset $\widehat{O}_{\rho_N}$, and we therefore do not expect improvement on the bound in Theorem 2.5 for covariance estimation with the localized PO update. We note, however, that for localized SR it is possible to derive an analog to the high probability version of Theorem 2.5 (see Theorem A.9) with an improved error bound, which we present in Theorem C.2.

Our discussion here should not be taken to imply that localization in posterior-approximation algorithms is not useful; it is plausible that localization in one step of the algorithm can lead to improved bounds in later steps, and we leave this multi-step analysis of localized posterior approximation ensemble updates as an important line for future work. A related phenomenon is known to occur in sequential Monte Carlo, where a proposal density that may be optimal for one step of the filter may not be optimal over multiple steps [1]. Another interesting direction for future study is the non-asymptotic analysis of ensemble Kalman methods for likelihood approximations in state-space models [24]. Finally, we envision that the non-asymptotic approach set forth here may be adopted to design and analyze new multi-step methods for posterior-approximation and sequential-optimization in inverse problems and data assimilation.

Acknowledgments

The authors are grateful to Jiaheng Chen, Subhodh Kotekal, Yandi Shen, and Nathan Waniorek for many helpful discussions. The authors are also thankful to the anonymous reviewers for their insightful suggestions that improved the manuscript.

Funding

National Science Foundation (grant nos NSF DMS-2027056 and NSF DMS-2237628 to D.S.A.); the BBVA Foundation (José Luis Rubio de Francia start-up grant); and Department of Energy for (grant no. DOE DE-SC0022232).

Data Availability Statement

No new data were generated or analysed in support of this research.

References

1. Agapiou, S., Papaspiliopoulos, O., Sanz-Alonso, D. & Stuart, A. M. (2017) Importance sampling: intrinsic dimension and computational cost. *Statist. Sci.*, **32**, 405–431.
2. Anderson, J. L. (2001) An ensemble adjustment Kalman filter for data assimilation. *Month. Weather Rev.*, **129**, 2884–2903.
3. Asch, M., Bocquet, M. & Nodet, M. (2016) *Data Assimilation: Methods, Algorithms, and Applications*, vol. **11**. Philadelphia, USA: SIAM.
4. Bengtsson, T., Bickel, P. J., Li, B., et al. (2008) Curse-of-dimensionality revisited: collapse of the particle filter in very large scale systems. In: *Probability and statistics: Essays in honor of David A. Freedman*. Institute of Mathematical Statistics, pp. 316–334.

5. Bergemann, K. & Reich, S. (2010) A localization technique for Ensemble Kalman Filters. *Quart. J. Royal Meteorol. Soc.*, **136**, 701–707.
6. Bergemann, K. & Reich, S. (2010) A mollified Ensemble Kalman Filter. *Quart. J. Roy. Meteorol. Soc.*, **136**, 1636–1643.
7. Bickel, P. J. & Levina, E. (2008) Covariance regularization by thresholding. *Ann. Stat.*, **36**, 2577–2604.
8. Bickel, P. J. & Levina, E. (2008) Regularized estimation of large covariance matrices. *Ann. Stat.*, **36**, 199–227.
9. Bickel, P. J., Li, B. & Bengtsson, T. (1994) Pushing the limits of contemporary statistics: contributions in honor of Jayanta K. Ghosh: sharp failure rates for the bootstrap particle filter in high dimensions. Institute of Mathematical Statistics, 318–329.
10. Bishop, A. N. & Del Moral (2023) On the mathematical theory of ensemble (linear-Gaussian) Kalman–Bucy filtering. *Math. Control Signals Syst.*, 1–69.
11. Bishop, C. H., Etherton, B. J. & Majumdar, S. J. (2001) Adaptive sampling with the ensemble transform Kalman filter. Part I: theoretical aspects. *Mon. Weather Rev.*, **129**, 420–436.
12. Blömker, D., Schillings, C. & Wacker, P. (2018) A strongly convergent numerical scheme from Ensemble Kalman Inversion. *SIAM J. Numer. Anal.*, **56**, 2537–2562.
13. Blömker, D., Schillings, C., Wacker, P. & Weissmann, S. (2019) Well posedness and convergence analysis of the Ensemble Kalman Inversion. *Inverse Probl.*, **35**, 085007.
14. Bogachev, V. I. (1998) *Gaussian Measures*. Providence, Rhode Island, USA: American Mathematical Soc.
15. Burgers, G., Jan van Leeuwen, P. & Evensen, G. (1998) Analysis scheme in the Ensemble Kalman Filter. *Mon. Weather Rev.*, **126**, 1719–1724.
16. Cai, T. T. & Yuan, M. (2012) Adaptive covariance matrix estimation through block thresholding. *Ann. Stat.*, **40**, 2014–2042.
17. Cai, T. T. & Zhou, H. H. (2012) Minimax estimation of large covariance matrices under ℓ_1-norm. *Stat. Sin.*, 1319–1349.
18. Cai, T. T. & Zhou, H. H. (2012) Optimal rates of convergence for sparse covariance matrix estimation. *Ann. Stat.*, **40**, 2389–2420.
19. Calvello, E., Reich, S. & Stuart, A. M. (2022) Ensemble Kalman methods: a mean field perspective. arXiv preprint arXiv:2209.11371.
20. Carrillo, J. A. & Vaes, U. (2021) Wasserstein stability estimates for covariance-preconditioned Fokker–Planck equations. *Nonlinearity*, **34**, 2275.
21. Chada, N. K., Chen, Y. & Sanz-Alonso, D. (2021) Iterative ensemble Kalman methods: a unified perspective with some new variants. *Found. Data Sci.*, **3**, 331–369.
22. Chatterjee, S. & Diaconis, P. (2018) The sample size required in importance sampling. *Ann. Appl. Probab.*, **28**, 1099–1135.
23. Chen, R. Y., Gittens, A. & Tropp, J. A. (2012) The masked sample covariance estimator: an analysis using matrix concentration inequalities. *Inf. Inference*, **1**, 2–20.
24. Chen, Y., Sanz-Alonso, D. & Willett, R. (2022) Autodifferentiable Ensemble Kalman Filters. *SIAM J. Math. Data Sci.*, **4**, 801–833.
25. Chorin, A. J. & Morzfeld, M. (2013) Conditions for successful data assimilation. *J. Geophys. Res.: Atmos.*, **118**, 11–522.
26. Del Moral, P. & Tugaut, J. (2018) On the stability and the uniform propagation of chaos properties of ensemble Kalman–Bucy filters. *Ann. Appl. Probab.*, **28**, 790–850.
27. Ding, Z. & Li, Q. (2021) Ensemble Kalman Inversion: mean-field limit and convergence analysis. *Stat. Comput.*, **31**, 1–21.
28. Dirksen, S. (2015) Tail bounds via generic chaining. *Electron. J. Probab.*, **20**, 1–29.
29. El Karoui, N. (2008) Operator norm consistent estimation of large-dimensional sparse covariance matrices. *Ann. Statist.*, **36**, 2717–2756.
30. Ernst, O. G., Sprungk, B. & Starkloff, H.-J. (2015) Analysis of the ensemble and polynomial chaos Kalman filters in Bayesian inverse problems. *SIAM/ASA J. Uncertain. Quantif.*, **3**, 823–851.

31. Evensen, G. (1995) Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.: Oceans*, **99**, 10143–10162.
32. Evensen, G. (2004) Sampling strategies and square root analysis schemes for the EnKF. *Ocean Dyn.*, **54**, 539–560.
33. Evensen, G. (2009) *Data Assimilation: the Ensemble Kalman Filter*. New York City, USA: Springer Science and Business Media.
34. Evensen, G. & Van Leeuwen, P. (1996) Assimilation of Geosat altimeter data for the Agulhas current using the Ensemble Kalman Filter with a quasigeostrophic model. *Mon. Weather Rev.*, **124**, 85–96.
35. Farchi, A. & Bocquet, M. (2019) On the efficiency of covariance localisation of the Ensemble Kalman Filter using augmented ensembles. *Front. Appl. Math. Stat.*, 3.
36. Furrer, R. & Bengtsson, T. (2007) Estimation of high-dimensional prior and posterior covariance matrices in Kalman filter variants. *J. Multivariate Anal.*, **98**, 227–255.
37. Garbuno-Inigo, A., Hoffmann, F., Li, W. & Stuart, A. M. (2020) Interacting Langevin diffusions: gradient structure and ensemble Kalman sampler. *SIAM J. Appl. Dyn. Syst.*, **19**, 412–441.
38. Gaspari, G. & Cohn, S. E. (1999) Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteorol. Soc.*, **125**, 723–757.
39. Al Ghattas, Bao, J. & Sanz-Alonso, D. (2023) Ensemble Kalman Filters with resampling. arXiv preprint arXiv:2308.08751.
40. Gottwald, G. A. & Majda, A. J. (2013) A mechanism for catastrophic filter divergence in data assimilation for sparse observation networks. *Nonlinear Process. Geophys.*, **20**, 705–712.
41. Gu, Y. & Oliver, D. S. (2007) An iterative Ensemble Kalman Filter for multiphase fluid flow data assimilation. *SPE J.*, **12**, 438–446.
42. Guth, P. A., Schillings, C. & Weissmann, S. (2022) *Ensemble Kalman Filter for Neural Network-based One-Shot Inversion*. Boston: de Gruyter & Co Berlin.
43. Hanke, M. (1997) A regularizing Levenberg-Marquardt scheme, with applications to inverse groundwater filtration problems. *Inverse Probl.*, **13**, 79–95.
44. Harlim, J. & Majda, A. J. (2010) Catastrophic filter divergence in filtering nonlinear dissipative systems. *Commun. Math. Sci.*, **8**, 27–43.
45. Herty, M. & Visconti, G. (2019) Kinetic methods for inverse problems. *Kinet. Related Models*, **12**, 1109.
46. Horn, R. A. & Johnson, C. R. (2012) *Matrix Analysis*. Cambridge, United Kingdom: Cambridge University Press.
47. Houtekamer, P. L. & Derome, J. (1995) Methods for ensemble prediction. *Mon. Weather Rev.*, **123**, 2181–2196.
48. Houtekamer, P. L. & Mitchell, H. L. (1998) Data assimilation using an Ensemble Kalman Filter technique. *Mon. Weather Rev.*, **126**, 796–811.
49. Houtekamer, P. L. & Mitchell, H. L. (2001) A sequential Ensemble Kalman Filter for atmospheric data assimilation. *Mon. Weather Rev.*, **129**, 123–137.
50. Houtekamer, P. L. & Zhang, F. (2016) Review of the Ensemble Kalman Filter for atmospheric data assimilation. *Mon. Weather Rev.*, **144**, 4489–4532.
51. Iglesias, M. A. (2016) A regularizing iterative ensemble Kalman method for PDE-constrained inverse problems. *Inverse Probl.*, **32**, 025002.
52. Iglesias, M. A., Law, K. J. H. & Stuart, A. M. (2013) Ensemble Kalman methods for inverse problems. *Inverse Probl.*, **29**, 045001.
53. Katzfuss, M., Stroud, J. R. & Wikle, C. K. (2016) Understanding the Ensemble Kalman Filter. *Am. Statist.*, **70**, 350–357.
54. Kelly, D., Majda, A. J. & Tong, X. T. (2015) Concrete Ensemble Kalman Filters with rigorous catastrophic filter divergence. *Proc. Natl. Acad. Sci.*, **112**, 10589–10594.
55. Kelly, D. & Stuart, A. M. (2014) Well-posedness and accuracy of the Ensemble Kalman Filter in discrete and continuous time. *Nonlinearity*, **27**.
56. Kim, H., Sanz-Alonso, D. & Strang, A. (2023) Hierarchical ensemble Kalman methods with sparsity-promoting generalized gamma hyperpriors. *Found. Data Sci.*, **5**, 366–388.

57. Koltchinskii, V. & Lounici, K. (2017) Concentration inequalities and moment bounds for sample covariance operators. *Bernoulli*, **23**, 110–133.
58. Kovachki, N. B. & Stuart, A. M. (2019) Ensemble Kalman Inversion: a derivative-free technique for machine learning tasks. *Inverse Probl.*, **35**, 095005.
59. Kwiatkowski, E. & Mandel, J. (2015) Convergence of the square root Ensemble Kalman Filter in the large ensemble limit. *SIAM/ASA J. Uncertain. Quantif.*, **3**, 1–17.
60. Lehmann, L. E. & Casella, G. (2006) *Theory of Point Estimation*. New York City, New York, USA: Springer Science & Business Media.
61. Lange, T. & Stannat, W. (2021) Mean field limit of ensemble square root filters-discrete and continuous time. *Found. Data Sci.*, **3**, 563–588.
62. Law, K. J. H., Stuart, A. M. & Zygalakis, K. (2015) *Data Assimilation*. New York City, New York, USA: Springer.
63. Law, K. J. H., Tembine, H. & Tempone, R. (2016) Deterministic mean-field ensemble Kalman filtering. *SIAM J. Sci. Comput.*, **38**, A1251–A1279.
64. Lawson, W. G. & Hansen, J. A. (2004) Implications of stochastic and deterministic filters as ensemble-based data assimilation methods in varying regimes of error growth. *Mon. Weather Rev.*, **132**, 1966–1981.
65. Le Gland, Monbet, V. & Tran, V.-D. (2009) *Large Sample Asymptotics for the Ensemble Kalman Filter* PhD thesis. Rocquencourt, France: INRIA.
66. Van Leeuwen, Cheng, Y. & Reich, S. (2015) *Nonlinear Data Assimilation*. New York City, New York, USA: Springer.
67. Leeuwenburgh, O., Evensen, G. & Bertino, L. (2005) The impact of ensemble filter definition on the assimilation of temperature profiles in the tropical Pacific. *Quart. J. Roy. Meteorol. Soc.*, **131**, 3291–3300.
68. Levina, E. & Vershynin, R. (2012) Partial estimation of covariance matrices. *Probab. Theory Related Fields*, **153**, 405–419.
69. Li G. & Reynolds A. C.. An iterative Ensemble Kalman Filter for data assimilation. In: *SPE Annual Technical Conference and Exhibition*. Society of Petroleum Engineers, 2007.
70. Li, J. & Xiu, D. (2008) On numerical properties of the Ensemble Kalman Filter for data assimilation. *Comput. Methods Appl. Mech. Engrg.*, **197**, 3574–3583.
71. Majda, A. J. & Harlim, J. (2012) *Filtering Complex Turbulent Systems*. Cambridge, United Kingdom: Cambridge University Press.
72. Majda, A. J. & Tong, X. T. (2018) Performance of Ensemble Kalman Filters in large dimensions. *Comm. Pure Appl. Math.*, **71**, 892–937.
73. Mandel, J., Cobb, L. & Beezley, J. D. (2011) On the convergence of the Ensemble Kalman Filter. *Appl. Math.*, **56**, 533–541.
74. Mendelson, S. (2016) Upper bounds on product and multiplier empirical processes. *Stochastic Process. Appl.*, **126**, 3652–3680.
75. Morzfeld, M., Hodyss, D. & Snyder, C. (2017) What the collapse of the Ensemble Kalman Filter tells us about particle filters. *Tellus A: Dyn. Meteorol. Oceanogr.*, **69**, 1283809.
76. Nüsken, N. & Reich, S. (2019) Note on interacting Langevin diffusions: gradient structure and ensemble Kalman sampler by Garbuno-Inigo, Hoffmann, Li and Stuart. arXiv preprint arXiv:1908.10890.
77. Ott, E., Hunt, B. R., Szunyogh, I., Zimin, A. V., Kostelich, E. J., Corazza, M., Kalnay, E., Patil, D. J. & Yorke, J. A. (2004) A local Ensemble Kalman Filter for atmospheric data assimilation. *Tellus A: Dyn. Meteorol. Oceanogr.*, **56**, 415–428.
78. Petrie, R. (2008) *Localization in the Ensemble Kalman Filter*. Reading, United Kingdom: MSc Atmosphere, Ocean and Climate University of Reading.
79. Reich, S. & Cotter, C. (2015) *Probabilistic Forecasting and Bayesian Data Assimilation*. Cambridge, United Kingdom: Cambridge University Press.
80. Reynolds A. C., Zafari M. & Li G.. Iterative forms of the Ensemble Kalman Filter. In: *ECMOR X-10th European Conference on the Mathematics of Oil Recovery*. pp. cp–23. European Association of Geoscientists & Engineers, 2006.

81. Roth, M., Hendeby, G., Fritsche, C. & Gustafsson, F. (2017) The Ensemble Kalman Filter: a signal processing perspective. *EURASIP J. Adv. Signal Process.*, **2017**, 1–16.
82. Sanz-Alonso, D. (2018) Importance sampling and necessary sample size: an information theory approach. *SIAM/ASA J. Uncertain. Quantif.*, **6**, 867–879.
83. Sanz-Alonso, D., Stuart, A. M. & Taeb, A. (2023) *Inverse Problems and Data Assimilation*, vol. **107**. Cambridge University Press.
84. Sanz-Alonso, D. & Wang, Z. (2021) Bayesian update with importance sampling: required sample size. *Entropy*, **23**, 22.
85. Särkkä S.. (2013) *Bayesian Filtering and Smoothing*, Vol. **3**. Cambridge, United Kingdom: Cambridge University Press.
86. Schillings, C. & Stuart, A. M. (2017) Analysis of the Ensemble Kalman Filter for inverse problems. *SIAM J. Numer. Anal.*, **55**, 1264–1290.
87. Snyder C.. Particle filters, the "optimal" proposal and high-dimensional systems. In: *Proceedings of the ECMWF Seminar on Data Assimilation for Atmosphere and Ocean*, 2011.
88. Snyder, C., Bengtsson, T., Bickel, P. J. & Anderson, J. L. (2016) Obstacles to high-dimensional particle filtering. *Mon. Weather Rev.*, **136**, 4629–4640.
89. Snyder, C., Bengtsson, T. & Morzfeld, M. (2015) Performance bounds for particle filters using the optimal proposal. *Mon. Weather Rev.*, **143**, 4750–4761.
90. Stein C.. (1972) A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In: *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Volume 2: Probability Theory*, pp. 583–603. California, USA: University of California Press.
91. Stuart, A. M. (2010) Inverse problems: a Bayesian perspective. *Acta Numer.*, **19**, 451–559.
92. Talagrand, M. (2014) *Upper and Lower Bounds for Stochastic Processes*, vol. **60**. Springer.
93. Tippett, M. K., Anderson, J. L., Bishop, C. H., Hamill, T. M. & Whitaker, J. S. (2003) Ensemble square root filters. *Mon. Weather Rev.*, **131**, 1485–1490.
94. Tong, X. T. (2018) Performance analysis of local Ensemble Kalman Filter. *J. Nonlinear Sci.*, **28**, 1397–1442.
95. Tong, X. T., Majda, A. J. & Kelly, D. (2015) Nonlinear stability of the Ensemble Kalman Filter with adaptive covariance inflation. *Nonlinearity*, **29**, 54–60.
96. Tong, X. T., Majda, A. J. & Kelly, D. (2016) Nonlinear stability and ergodicity of ensemble based Kalman filters. *Nonlinearity*, **29**, 657.
97. Tong, X. T. & Morzfeld, M. (2023) Localized Ensemble Kalman Inversion. *Inverse Probl.*, **39**, 064002.
98. Trevisan, A. & Uboldi, F. (2004) Assimilation of standard and targeted observations within the unstable subspace of the observation–analysis–forecast cycle system. *J. Atmos. Sci.*, **61**, 103–113.
99. Tropp, J. A. (2015) *An Introduction to Matrix Concentration Inequalities*, vol. **8**. Norwell, MA, USA: Now Publishers, Inc.
100. Ungarala, S. (2012) On the iterated forms of Kalman filters using statistical linearization. *J. Process Control*, **22**, 935–943.
101. Van Handel, R. (2017) On the spectral norm of Gaussian random matrices. *Trans. Amer. Math. Soc.*, **369**, 8161–8178.
102. Vershynin, R. (2018) *High-Dimensional Probability: An Introduction with Applications in Data Science*, vol. **47**. Cambridge, United Kingdom: Cambridge University Press.
103. Wainwright, M. J. (2019) *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*, vol. **48**. Cambridge, United Kingdom: Cambridge University Press.

Appendix

We provide proofs of all theorems in the main body. We will use the following result extensively and summarise it here for brevity. Given events $E_1, \ldots, E_J$ that each occur with probability at least $1 - ce^{-t}$,

where $t \geq 1$ and $c > 0$ is a universal constant that may be different for each event, then

$$\mathbb{P}\left(\bigcap_{j=1}^{J} E_j\right) = 1 - \mathbb{P}\left(\bigcup_{j=1}^{J} \bar{E}_j\right) \geq 1 - \sum_{j=1}^{J} \mathbb{P}(\bar{E}_j) \geq 1 - ce^{-t}.$$

A. Proofs: Section 2

This appendix contains the proofs of all the theorems in Section 2. Background results on covariance estimation are reviewed in Subsection A.1 and the continuity and boundedness of the Kalman gain, mean-update, covariance-update and nonlinear gain-update operators are summarized in Subsection A.2. These preliminary results are used in Subsection A.3 to establish our main theorems.

A.1 *Preliminaries: concentration and covariance estimation*

THEOREM A.1. (Sub-Gaussian Norm Concentration, [102, Exercise 6.3.5]) Let X be a d-dimensional sub-Gaussian random vector with $\mathbb{E}[X] = \mu^X$, $\text{var}[X] = \Sigma^X$. Then, for any $t \geq 1$, with probability at least $1 - ce^{-t}$ it holds that

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

Proof of Proposition 2.1 For $n = 1, \ldots, N$, let $u_n = Z_n + m$, where Z_n is a centered sub-Gaussian random vector with $\text{var}[Z_n] = C$. Then we may write

$$\widehat{C} = \frac{1}{N-1}\sum_{n=1}^{N}(Z_n - \bar{Z})(Z_n - \bar{Z})^\top \asymp \frac{1}{N}\sum_{n=1}^{N} Z_n Z_n^\top - \bar{Z}\bar{Z}^\top \equiv \widehat{C}^0 - \bar{Z}\bar{Z}^\top.$$

Therefore,

$$\|\widehat{C} - C\| \leq \|\widehat{C}^0 - C\| + \|\bar{Z}\bar{Z}^\top\| = \|\widehat{C}^0 - C\| + \|\bar{Z}\|_2^2.$$

Let E_1 denote the event on which

$$\|\widehat{C}^0 - C\| \lesssim \|C\|\left(\sqrt{\frac{r_2(C)}{N}} \vee \frac{r_2(C)}{N} \vee \sqrt{\frac{t}{N}} \vee \frac{t}{N}\right),$$

and E_2 the event on which

$$\|\bar{Z}\|_2^2 \lesssim \|C\|\left(\frac{r_2(C)}{N} \vee \frac{t}{N}\right).$$

Then by Theorem 9 of [57], $\mathbb{P}(E_1) \geq 1 - e^{-t}$, and from Theorem A.1, $\mathbb{P}(E_2) \geq 1 - e^{-t}$. Therefore, the result holds on $E_1 \cap E_2$, which has probability at least $1 - ce^{-t}$. □

LEMMA A.2. (Sample Covariance Operator Norm Bound) Let $u_1, \ldots, u_N$ and $\widehat{C}$ be as in Proposition 2.1. Then, for any $t \geq 1$, it holds with probability at least $1 - ce^{-t}$ such that

$$\|\widehat{C}\| \lesssim \|C\| \left(1 \vee \frac{r_2(C)}{N} \vee \frac{t}{N}\right).$$

Proof. By the triangle inequality $\|\widehat{C}\| \leq \|\widehat{C} - C\| + \|C\|$. The result follows by Proposition 2.1 noting that, for any $x \geq 0$, $1 \vee \sqrt{x} \vee x = 1 \vee x$. □

LEMMA A.3. (Cross-Covariance Estimation—Unstructured Case) Let $u_1, \ldots, u_N$ be d-dimensional i.i.d. sub-Gaussian random vectors with $\mathbb{E}[u_1] = m$ and $\mathrm{var}[u_1] = C$. Let $\eta_1, \ldots, \eta_N$ be k-dimensional i.i.d. sub-Gaussian random vectors with $\mathbb{E}[\eta_1] = 0$ and $\mathrm{var}[\eta_1] = \Gamma$, and assume that the two sequences are independent. Consider the estimator

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

of the cross-covariance $C^{u\eta} \equiv \mathbb{E}\big[(u_1 - m)\eta_1^\top\big]$. Then there exists a constant c such that, for all $t \geq 1$, it holds with probability at least $1 - ce^{-t}$ that

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

Proof. First, we note that

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

and so it suffices to prove the claim for the biased sample covariance estimator, which we denote by $\widetilde{C}^{u\eta}$. Letting $Z_n = u_n - m$, it follows that

$$\|\widetilde{C}^{u\eta}\| = \left\| \frac{1}{N} \sum_{n=1}^{N} Z_n \eta_n^\top - \bar{Z}\bar{\eta}^\top \right\| \leq \left\| \frac{1}{N} \sum_{n=1}^{N} Z_n \eta_n^\top \right\| + \|\bar{Z}\bar{\eta}^\top\|. \tag{A1}$$

For the second term in the right-hand side of (A1), let E_1 denote the event on which

$$\|\bar{Z}\|_2 \lesssim \sqrt{\|C\|\left(\frac{r_2(C)}{N} \vee \frac{t}{N}\right)},$$

and E_2 the event on which

$$\|\bar{\eta}\|_2 \lesssim \sqrt{\|\Gamma\|\left(\frac{r_2(\Gamma)}{N} \vee \frac{t}{N}\right)},$$

each of which have probability at least $1 - e^{-t}$ from Theorem A.1. Therefore, the event $E_1 \cap E_2$ occurs with probability at least $1 - ce^{-t}$, and on which it follows that

$$\|\bar{Z}\bar{\eta}^\top\| = \|\bar{Z}\|_2\|\bar{\eta}\|_2 \lesssim (\|C\| \vee \|\Gamma\|)\left(\frac{r_2(C)}{N} \vee \frac{r_2(\Gamma)}{N} \vee \frac{t}{N}\right),$$

where the inequality follows since $\sqrt{ab} \lesssim a \vee b$ for $a, b \geq 0$. To control the first term in the right-hand side of (A1), we define the vector

$$W_n = \begin{bmatrix} Z_n \\ \eta_n \end{bmatrix} \in \mathbb{R}^{d+k}, \qquad 1 \leq n \leq N,$$

and note that $W_1, \ldots, W_N$ is an i.i.d. sub-Gaussian sequence with $\mathbb{E}[W_1] = [m^\top, 0_k^\top]^\top$ and variance $C^W = \mathrm{diag}(C, \Gamma)$. Let E_3 denote the event on which

$$\begin{aligned}
\left\|\frac{1}{N}\sum_{n=1}^N W_n W_n^\top - C^W\right\| &\lesssim \|C^W\|\left(\sqrt{\frac{r_2(C^W)}{N}} \vee \frac{r_2(C^W)}{N} \vee \sqrt{\frac{t}{N}} \vee \frac{t}{N}\right) \\
&\lesssim (\|C\| \vee \|\Gamma\|)\left(\left(\sqrt{\frac{\mathrm{Tr}(C)}{N\|C\|}} + \sqrt{\frac{\mathrm{Tr}(\Gamma)}{N\|\Gamma\|}}\right) \vee \frac{\mathrm{Tr}(C) + \mathrm{Tr}(\Gamma)}{N(\|C\| \vee \|\Gamma\|)} \vee \sqrt{\frac{t}{N}} \vee \frac{t}{N}\right) \\
&\lesssim (\|C\| \vee \|\Gamma\|)\left(\left(\sqrt{\frac{r_2(C)}{N}} + \sqrt{\frac{r_2(\Gamma)}{N}}\right) \vee \left(\frac{r_2(C)}{N} + \frac{r_2(\Gamma)}{N}\right) \vee \left(\sqrt{\frac{t}{N}} \vee \frac{t}{N}\right)\right) \\
&\lesssim (\|C\| \vee \|\Gamma\|)\left(\sqrt{\frac{r_2(C)}{N}} \vee \frac{r_2(C)}{N} \vee \sqrt{\frac{r_2(\Gamma)}{N}} \vee \frac{r_2(\Gamma)}{N} \vee \sqrt{\frac{t}{N}} \vee \frac{t}{N}\right).
\end{aligned}$$

By Proposition 2.1, it holds for any $t \geq 1$ that $\mathbb{P}(E_3) \geq 1 - e^{-t}$. Note that we can express

$$\mathsf{P} \equiv \frac{1}{N}\sum_{n=1}^N W_n W_n^\top - \begin{bmatrix} C & O \\ O & \Gamma \end{bmatrix} = \begin{bmatrix} N^{-1}\sum_{n=1}^N Z_n Z_n^\top - C & N^{-1}\sum_{n=1}^N Z_n \eta_n^\top \\ N^{-1}\sum_{n=1}^N \eta_n Z_n^\top & N^{-1}\sum_{n=1}^N \eta_n \eta_n^\top - \Gamma \end{bmatrix},$$

and that

$$\left\| \frac{1}{N} \sum_{n=1}^{N} Z_n \eta_n^\top \right\| = \|E_{11} \mathsf{P} E_{12}\| \leq \|E_{11}\| \|\mathsf{P}\| \|E_{12}\| = \|\mathsf{P}\|,$$

where E_{11}, E_{12} are block *selection* matrices that pick the relevant sub-block matrix of P. Therefore, it holds on E_3 that

$$\left\| \frac{1}{N} \sum_{n=1}^{N} Z_n \eta_n^\top \right\| \lesssim (\|C\| \vee \|\Gamma\|) \left(\sqrt{\frac{r_2(C)}{N}} \vee \frac{r_2(C)}{N} \vee \sqrt{\frac{r_2(\Gamma)}{N}} \vee \frac{r_2(\Gamma)}{N} \vee \sqrt{\frac{t}{N}} \vee \frac{t}{N} \right).$$

The final result follows by noting that the intersection $E_1 \cap E_2 \cap E_3$ has probability at least $1 - ce^{-t}$. □

A.2 *Continuity and boundedness of update operators*

The next three lemmas, shown in [59], ensure the continuity and boundedness of the Kalman gain, mean-update, and covariance-update operators introduced in Section 2. We include them here for completeness. Lemma A.7 below establishes similar properties for the nonlinear gain-update operator introduced in Section 3.

LEMMA A.4. (Continuity and Boundedness of Kalman Gain Operator [59, Lemma 4.1 and Corollary 4.2]) Let $\mathscr{K}$ be the Kalman gain operator defined in (2.2). Let $P, Q \in \mathcal{S}_+^d$, $\Gamma \in \mathcal{S}_{++}^k$ and $A \in \mathbb{R}^{k \times d}$. The following holds:

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

LEMMA A.5. (Continuity and Boundedness of Mean-Update Operator [59, Corollary 4.3 and Lemma 4.7]) Let $\mathscr{M}$ be the mean-update operator defined in (2.3). Let $P, Q \in \mathcal{S}_+^d$, $\Gamma \in \mathcal{S}_{++}^k$, $A \in \mathbb{R}^{k \times d}$, $y \in \mathbb{R}^k$ and $m, m' \in \mathbb{R}^d$. The following holds:

$$\begin{aligned} \|\mathscr{M}(m, Q)\| &\leq \|m\| + \|Q\| \|A\| \|\Gamma^{-1}\| \|y - Am\|_2, \\ \left\|\mathscr{M}(m, Q) - \mathscr{M}(m', P)\right\| &\leq \left\|m - m'\right\| \left(1 + \|A\|^2 \|\Gamma^{-1}\| \|Q\|\right) \\ &\quad + \|Q - P\| \|A\| \|\Gamma^{-1}\| \left(1 + \|A\|^2 \|\Gamma^{-1}\| \|P\|\right) \left\|y - Am'\right\|_2. \end{aligned}$$

LEMMA A.6. (Continuity and Boundedness of Covariance-Update Operator [59, Lemma 4.4 and Lemma 4.6]) Let $\mathscr{C}$ be the covariance-update operator defined in (2.4). Let $P, Q \in \mathcal{S}_+^d$, $\Gamma \in \mathcal{S}_{++}^k$, $A \in \mathbb{R}^{k \times d}$,

$y \in \mathbb{R}^k$ and $m, m' \in \mathbb{R}^d$. The following holds:

$$\begin{aligned}\|\mathscr{C}(Q) - \mathscr{C}(P)\| &\leq \|Q - P\|\Big(1 + \|A\|^2\|\Gamma^{-1}\|(\|Q\| + \|P\|) + \|A\|^4\|\Gamma^{-1}\|^2\|Q\|\|P\|\Big),\\ 0 &\preccurlyeq \mathscr{C}(Q) \preccurlyeq Q,\\ \|\mathscr{C}(Q)\| &\leq \|Q\|.\end{aligned}$$

LEMMA A.7. (Continuity and Boundedness of Nonlinear Gain-Update Operator) Let $\mathscr{P}$ be the nonlinear gain-update operator defined in (3.5). Let $P, \widetilde{P} \in \mathbb{R}^{d\times k}$, $Q, \widetilde{Q} \in \mathcal{S}_+^k$ and $\Gamma \in \mathcal{S}_{++}^k$. The following holds:

$$\begin{aligned}\|\mathscr{P}(P, Q) - \mathscr{P}(\widetilde{P}, \widetilde{Q})\| &\leq \|\Gamma^{-1}\|\|P - \widetilde{P}\| + \|\Gamma^{-1}\|^2\|P\|\|Q - \widetilde{Q}\|,\\ \|\mathscr{P}(P, Q)\| &\leq \|\Gamma^{-1}\|\|P\| + \|\Gamma^{-1}\|^2\|Q\|.\end{aligned}$$

Proof. The proof follows in similar style to Lemma 4.1 in [59]. We note that

$$\begin{aligned}\|P(Q + \Gamma)^{-1} - \widetilde{P}(\widetilde{Q} + \Gamma)^{-1}\| &\leq \|P(Q + \Gamma)^{-1} - P(\widetilde{Q} + \Gamma)^{-1}\| + \|P(\widetilde{Q} + \Gamma)^{-1} - \widetilde{P}(\widetilde{Q} + \Gamma)^{-1}\|\\ &\leq \|P\|\|(Q + \Gamma)^{-1} - (\widetilde{Q} + \Gamma)^{-1}\| + \|\widetilde{P} - P\|\|(Q + \Gamma)^{-1}\|.\end{aligned}$$

Since $\Gamma \succ 0$ and $Q \succeq 0$, it holds that $Q + \Gamma \succeq \Gamma$ and so $(Q + \Gamma)^{-1} \preccurlyeq \Gamma^{-1}$, which, in turn, implies $\|(Q + \Gamma)^{-1}\| \leq \|\Gamma^{-1}\|$. Further,

$$\begin{aligned}\|(Q + \Gamma)^{-1} - (\widetilde{Q} + \Gamma)^{-1}\| &= \|\Gamma^{-1/2}[(\Gamma^{-1/2}Q\Gamma^{-1/2} + I)^{-1} - (\Gamma^{-1/2}\widetilde{Q}\Gamma^{-1/2} + I)^{-1}]\Gamma^{-1/2}\|\\ &\leq \|\Gamma^{-1}\|\|(\Gamma^{-1/2}Q\Gamma^{-1/2} + I)^{-1} - (\Gamma^{-1/2}\widetilde{Q}\Gamma^{-1/2} + I)^{-1}\|\\ &\leq \|\Gamma^{-1}\|\|\Gamma^{-1/2}Q\Gamma^{-1/2} - \Gamma^{-1/2}\widetilde{Q}\Gamma^{-1/2}\|\\ &\leq \|\Gamma^{-1}\|^2\|Q - \widetilde{Q}\|,\end{aligned}$$

where the second to last equality follows by the fact that $\|(I + A)^{-1} - (I + B)^{-1}\| \leq \|B - A\|$ for $A, B \in \mathcal{S}_+^k$. To prove the pointwise boundedness of $\mathscr{P}$, take $\widetilde{P}$ to be the $d \times k$ matrix of zeroes, and $\widetilde{Q}$ to be the $k \times k$ matrix of zeroes, and plug these values into the continuity bound. □

A.3 *Proof of main results in Section 2*

THEOREM A.8. (Posterior Mean Approximation with Finite Ensemble—High Probability Bound) Consider the PO and SR ensemble Kalman updates given by (2.7) and (2.9), respectively, leading to an estimate $\widehat{\mu}$ of the posterior mean μ defined in (1.2). Set $\phi = 1$ for the PO update and $\phi = 0$ for the SR update. Then there exists a constant c such that, for all $t \geq 1$, it holds with probability at least $1 - ce^{-t}$

that

$$\|\widehat{\mu}-\mu\|_2 \lesssim (\|C\|^{1/2}\vee\|C\|^2)(\|A\|\vee\|A\|^4)(\|\Gamma^{-1}\|\vee\|\Gamma^{-1}\|^2)(1\vee\|y-Am\|_2)$$
$$\times\left(\sqrt{\frac{r_2(C)}{N}}\vee\sqrt{\frac{t}{N}}\vee\left(\frac{r_2(C)}{N}\right)^{3/2}\vee\left(\frac{t}{N}\right)^{3/2}\vee\frac{r_2(C)}{N}\sqrt{\frac{t}{N}}\vee\sqrt{\frac{r_2(C)}{N}}\frac{t}{N}\right)+\phi\mathscr{E},$$

where

$$\mathscr{E}=\|A\|\|\Gamma^{-1}\|\|\Gamma\|^{1/2}\|C\|\left(\sqrt{\frac{r_2(\Gamma)}{N}}\vee\sqrt{\frac{t}{N}}\vee\frac{r_2(C)}{N}\sqrt{\frac{r_2(\Gamma)}{N}}\vee\frac{r_2(C)}{N}\sqrt{\frac{t}{N}}\vee\frac{t}{N}\sqrt{\frac{r_2(\Gamma)}{N}}\vee\left(\frac{t}{N}\right)^{3/2}\right).$$

Proof. It follows from Lemma A.5 that

$$\begin{aligned}\|\widehat{\mu}-\mu\|_2 &= \|\mathscr{M}(\widehat{m},\widehat{C})-\phi\mathscr{K}(\widehat{C})\bar{\eta}-\mathscr{M}(m,C)\|_2\\ &\le \|\mathscr{M}(\widehat{m},\widehat{C})-\mathscr{M}(m,C)\|_2+\phi\|\mathscr{K}(\widehat{C})\bar{\eta}\|_2\\ &\le \|\widehat{m}-m\|_2\left(1+\|A\|^2\|\Gamma^{-1}\|\|\widehat{C}\|\right) \qquad \text{(A2)}\\ &+\|\widehat{C}-C\|\|A\|\|\Gamma^{-1}\|\left(1+\|A\|^2\|\Gamma^{-1}\|\|C\|\right)\|y-Am\|_2 \qquad \text{(A3)}\\ &+\phi\|\mathscr{K}(\widehat{C})\|\|\bar{\eta}\|_2. \qquad \text{(A4)}\end{aligned}$$

We now control each of the terms in equations (A2), (A3) and (A4) separately. For (A2), we note that $\widehat{m}-m\sim\mathcal{N}(0,C/N)$. Let E_1 be the set on which

$$\|\widehat{m}-m\|_2\lesssim\sqrt{\|C\|\left(\frac{r_2(C)}{N}\vee\frac{t}{N}\right)},$$

let E_2 be the set on which

$$\|\widehat{C}-C\|\lesssim\|C\|\left(\sqrt{\frac{r_2(C)}{N}}\vee\frac{r_2(C)}{N}\vee\sqrt{\frac{t}{N}}\vee\frac{t}{N}\right)$$

and

$$\|\widehat{C}\|\lesssim\|C\|\left(1\vee\frac{r_2(C)}{N}\vee\frac{t}{N}\right),$$

and let E_3 be the set on which

$$\|\bar{\eta}\|_2 \lesssim \sqrt{\|\Gamma\|\left(\frac{r_2(\Gamma)}{N} \vee \frac{t}{N}\right)}.$$

From Theorem A.1, Proposition 2.1 and Lemma A.2, the set $E = E_1 \cap E_2 \cap E_3$ has probability at least $1 - ce^{-t}$, and it holds on this set that (A2) is bounded above by

$$(\|C\|^{1/2} \vee \|C\|^{3/2})(1 \vee \|A\|^2\|\Gamma^{-1}\|)\left(\sqrt{\frac{r_2(C)}{N}} \vee \sqrt{\frac{t}{N}} \vee \left(\frac{r_2(C)}{N}\right)^{3/2} \vee \left(\frac{t}{N}\right)^{3/2} \vee \frac{r_2(C)}{N}\sqrt{\frac{t}{N}} \vee \sqrt{\frac{r_2(C)}{N}}\frac{t}{N}\right). \quad \text{(A5)}$$

Further, on the set E, we can bound (A3) above by

$$(\|C\| \vee \|C\|^2)(\|A\| \vee \|A\|^3)\left(\|\Gamma^{-1}\| \vee \|\Gamma^{-1}\|^2\right)\|y - Am\|\left(\sqrt{\frac{r_2(C)}{N}} \vee \frac{r_2(C)}{N} \vee \sqrt{\frac{t}{N}} \vee \frac{t}{N}\right). \quad \text{(A6)}$$

Finally, for (A4), it follows from Lemma A.4:

$$\|\mathscr{K}(\widehat{C})\|\|\bar{\eta}\| \leq \|A\|\|\Gamma^{-1}\|\|\widehat{C}\|\|\bar{\eta}\|$$

and so on the set E, we can show that (A4) is bounded above by

$$\|A\|\|\Gamma^{-1}\|\|C\|\|\Gamma\|^{1/2}\left(\sqrt{\frac{r_2(\Gamma)}{N}} \vee \sqrt{\frac{t}{N}} \vee \frac{r_2(C)}{N}\sqrt{\frac{r_2(\Gamma)}{N}} \vee \frac{r_2(C)}{N}\sqrt{\frac{t}{N}} \vee \frac{t}{N}\sqrt{\frac{r_2(\Gamma)}{N}} \vee \left(\frac{t}{N}\right)^{3/2}\right) \equiv \mathsf{E}. \quad \text{(A7)}$$

Putting the three bounds (A5), (A6) and (A7) together, we see that on E, it holds that

$$\begin{aligned}\|\widehat{\mu} - \mu\|_2 \lesssim\ & (\|C\|^{1/2} \vee \|C\|^2)(\|A\| \vee \|A\|^4)(\|\Gamma^{-1}\| \vee \|\Gamma^{-1}\|^2)(1 \vee \|y - Am\|_2) \\ & \times \left(\sqrt{\frac{r_2(C)}{N}} \vee \sqrt{\frac{t}{N}} \vee \left(\frac{r_2(C)}{N}\right)^{3/2} \vee \left(\frac{t}{N}\right)^{3/2} \vee \frac{r_2(C)}{N}\sqrt{\frac{t}{N}} \vee \sqrt{\frac{r_2(C)}{N}}\frac{t}{N}\right) \\ & + \phi\mathsf{E}.\end{aligned}$$

□

Proof of Theorem 2.3 Recall that from Theorem A.8, for all $t \geq 1$ with probability at least $1 - ce^{-t}$,

$$\|\widehat{\mu} - \mu\|_2 \lesssim (\|C\|^{1/2} \vee \|C\|^2)(\|A\| \vee \|A\|^4)(\|\Gamma^{-1}\| \vee \|\Gamma^{-1}\|^2)(1 \vee \|y - Am\|_2)$$
$$\times \left(\sqrt{\frac{r_2(C)}{N}} \vee \left(\frac{r_2(C)}{N}\right)^{3/2} \vee \sqrt{\frac{t}{N}} \vee \left(\frac{t}{N}\right)^{3/2} \vee \frac{r_2(C)}{N}\sqrt{\frac{t}{N}} \vee \sqrt{\frac{r_2(C)}{N}}\frac{t}{N} \right) + \phi\mathscr{E}.$$

For notational brevity, let

$$\mathcal{W} \equiv (\|C\|^{1/2} \vee \|C\|^2)(\|A\| \vee \|A\|^4)(\|\Gamma^{-1}\| \vee \|\Gamma^{-1}\|^2)(1 \vee \|y - Am\|_2),$$

and let $B \equiv \mathcal{W}\left(\sqrt{\frac{r_2(C)}{N}} \vee \left(\frac{r_2(C)}{N}\right)^{3/2}\right)$. Then, for $\phi = 0$ and $p \geq 1$,

$$\begin{aligned}
\mathbb{E}\big[\|\widehat{\mu} - \mu\|_2^p\big] &= p\int_0^\infty x^{p-1}\mathbb{P}(\|\widehat{\mu} - \mu\|_2 > x)dx \\
&\leq p\int_0^B x^{p-1}dx + p\int_B^\infty x^{p-1}\mathbb{P}(\|\widehat{\mu} - \mu\|_2 > x)dx \\
&\lesssim B^p + p\int_0^\infty x^{p-1}\exp\left(-\min\left(\frac{Nx^2}{\mathcal{W}^2}, \frac{Nx^{2/3}}{\mathcal{W}^{2/3}}, \frac{N^3x^2}{\mathcal{W}^2 r_2^2(C)}, \frac{N^{3/2}x}{\mathcal{W}\sqrt{r_2(C)}}\right)\right)dx \\
&= B^p + p\max\left\{\frac{1}{2}\Gamma\left(\frac{p}{2}\right)\left(\frac{\mathcal{W}}{\sqrt{N}}\right)^p, \frac{1}{2}\Gamma\left(\frac{3p}{2}\right)\left(\frac{\mathcal{W}}{N^{3/2}}\right)^p,\right. \\
&\qquad \left.\frac{1}{2}\Gamma\left(\frac{p}{2}\right)\left(\frac{\mathcal{W}r_2(C)}{N^{3/2}}\right)^p, \Gamma(p)\left(\frac{\mathcal{W}\sqrt{r_2(C)}}{N^{3/2}}\right)^p\right\},
\end{aligned}$$

where the final equality follows by direct integration. It follows then that

$$\big[\mathbb{E}\|\widehat{\mu} - \mu\|_2^p\big]^{1/p} \lesssim B + c(p)\mathcal{W}\max\left(\frac{1}{\sqrt{N}}, \frac{1}{N^{3/2}}, \frac{r_2(C)}{N^{3/2}}, \frac{\sqrt{r_2(C)}}{N^{3/2}}\right) \lesssim c(p)B,$$

where the final inequality holds since $r_2(C) \geq 1$. The result for the $\phi = 1$ case is identical and thus omitted. The constants in the statement of the result are then

$$c_1 = (\|C\|^{1/2} \vee \|C\|^2)(\|A\| \vee \|A\|^4)(\|\Gamma^{-1}\| \vee \|\Gamma^{-1}\|^2)(1 \vee \|y - Am\|_2),$$
$$c_2 = \|A\|\|\Gamma^{-1}\|\|\Gamma\|^{1/2}\|C\|.$$

□

THEOREM A.9. (Posterior Covariance Approximation with Finite Ensemble—High Probability Bound) Consider the PO and SR ensemble Kalman updates given by (2.7) and (2.9), respectively, leading to an

estimate $\widehat{\Sigma}$ of the posterior covariance Σ defined in (1.2). Set $\phi = 1$ for the PO update and $\phi = 0$ for the SR update. For any $t \geq 1$, it holds with probability at least $1 - ce^{-t}$ that

$$\|\widehat{\Sigma} - \Sigma\| \lesssim (\|C\| \vee \|C\|^3)(\|A\|^2 \vee \|A\|^4)(\|\Gamma^{-1}\| \vee \|\Gamma^{-1}\|^2) \left(\sqrt{\frac{r_2(C)}{N}} \vee \left(\frac{r_2(C)}{N}\right)^2 \vee \sqrt{\frac{t}{N}} \vee \left(\frac{t}{N}\right)^2\right) + \phi\mathscr{E},$$

where

$$\mathscr{E} = (\|A\| \vee \|A\|^3)(\|\Gamma^{-1}\| \vee \|\Gamma^{-1}\|^2)(\|C\| \vee \|\Gamma\|)(\|C\| \vee \|C\|^2) \times \left(\sqrt{\frac{r_2(C)}{N}} \vee \left(\frac{r_2(C)}{N}\right)^3 \vee \sqrt{\frac{t}{N}} \vee \left(\frac{t}{N}\right)^3 \vee \left(\sqrt{\frac{r_2(\Gamma)}{N}} \vee \frac{r_2(\Gamma)}{N}\right)\left(1 \vee \left(\frac{r_2(C)}{N}\right)^2 \vee \left(\frac{t}{N}\right)^2\right)\right).$$

Proof. From Proposition 4 of [36], for the PO-ensemble Kalman update, we may write

$$\widehat{\Sigma} = \mathscr{C}(\widehat{C}) + \widehat{O},$$

while for the SR-ensemble Kalman update, we have $\widehat{\Sigma} = \mathscr{C}(\widehat{C})$. We deal initially with the $\mathscr{C}(\widehat{C})$ term that is common to both expressions, and then proceed to show how the operator norm of the additional $\widehat{O}$ term can be controlled. From Lemma A.6, the continuity of $\mathscr{C}$ immediately implies that

$$\begin{aligned}\|\mathscr{C}(\widehat{C}) - \mathscr{C}(C)\| &\leq \|\widehat{C} - C\| \left(1 + \|A\|^2\|\Gamma^{-1}\|\left(\|\widehat{C}\| + \|C\|\right) + \|A\|^4\|\Gamma^{-1}\|^2\|\widehat{C}\|\|C\|\right) \\ &= \left[\|A\|^2\|\Gamma^{-1}\| + \|A\|^4\|\Gamma^{-1}\|^2\|C\|\right]\|\widehat{C} - C\|\|\widehat{C}\| \\ &+ \left[1 + \|A\|^2\|\Gamma^{-1}\|\|C\|\right]\|\widehat{C} - C\|.\end{aligned}$$

For any $N \in \mathbb{N}$ and $a > 0$, let $\mathscr{R}_N(a) \equiv \sqrt{\frac{a}{N}} \vee \frac{a}{N}$. Let E_1 be the set on which both

$$\|\widehat{C} - C\| \lesssim \|C\|\left(\mathscr{R}_N(r_2(C)) \vee \mathscr{R}_N(t)\right), \qquad \text{and} \qquad \|\widehat{C}\| \lesssim \|C\|\left(1 \vee \mathscr{R}_N(r_2(C)) \vee \mathscr{R}_N(t)\right).$$

Let E_2 be the set on which

$$\|\hat{\Gamma} - \Gamma\| \lesssim \|\Gamma\|\left(\mathscr{R}_N(r_2(\Gamma)) \vee \mathscr{R}_N(t)\right),$$

and E_3 the set on which

$$\|\widehat{C}^{u\eta} - C^{u\eta}\| \lesssim (\|C\| \vee \|\Gamma\|) \left(\mathscr{R}_N(r_2(C)) \vee \mathscr{R}_N(r_2(\Gamma)) \vee \mathscr{R}_N(t)\right).$$

Then, by Proposition 2.1 applied separately to E_1 and E_2, and Lemma A.3 applied to E_3, the intersection $E = E_1 \cap E_2 \cap E_3$ has probability at least $1 - ce^{-t}$. It follows that on E:

$$\begin{aligned}\|\widehat{C} - C\|\|\widehat{C}\| &\lesssim \|C\|^2 \left(\mathscr{R}_N(r_2(C)) \vee \mathscr{R}_N(t)\right)\left(1 \vee \mathscr{R}_N(r_2(C)) \vee \mathscr{R}_N(t)\right) \\ &\lesssim \|C\|^2 \left(\mathscr{R}_N(r_2(C)) \vee \mathscr{R}_N(t) \vee \mathscr{R}^2_{N,2}(C) \vee \mathscr{R}^2_{N,2}(t)\right), \qquad \text{(A8)}\end{aligned}$$

$$\|\hat{\Gamma} - \Gamma\|\|\widehat{C}\|^2 \lesssim \|C\|^2\|\Gamma\| \left(\mathscr{R}_N(r_2(\Gamma)) \vee \mathscr{R}_N(t)\right)\left(1 \vee \mathscr{R}^2_{N,2}(C) \vee \mathscr{R}^2_{N,2}(t)\right), \qquad \text{(A9)}$$

$$\|\widehat{C}^{u\eta} - C^{u\eta}\|\|\widehat{C}\| \lesssim \|C\|(\|C\| \vee \|\Gamma\|)\left(1 \vee \mathscr{R}_N(r_2(C)) \vee \mathscr{R}_N(t)\right)\left(\mathscr{R}_N(r_2(C)) \vee \mathscr{R}_N(r_2(\Gamma)) \vee \mathscr{R}_N(t)\right), \qquad \text{(A10)}$$

$$\|\widehat{C}^{u\eta} - C^{u\eta}\|\|\widehat{C}\|^2 \lesssim \|C\|^2(\|C\| \vee \|\Gamma\|)\left(1 \vee \mathscr{R}^2_{N,2}(C) \vee \mathscr{R}^2_{N,2}(t)\right)\left(\mathscr{R}_N(r_2(C)) \vee \mathscr{R}_N(r_2(\Gamma)) \vee \mathscr{R}_N(t)\right). \qquad \text{(A11)}$$

Using (A8), it follows that on E,

$$\begin{aligned}\|\widehat{\Sigma} - \Sigma\| &\lesssim (\|C\| \vee \|C\|^3)(\|A\|^2 \vee \|A\|^4)(\|\Gamma^{-1}\| \vee \|\Gamma^{-1}\|^2)\left(\mathscr{R}_N(r_2(C)) \vee \mathscr{R}_N(t) \vee \mathscr{R}^2_{N,2}(C) \vee \mathscr{R}^2_{N,2}(t)\right) \\ &= (\|C\| \vee \|C\|^3)(\|A\|^2 \vee \|A\|^4)(\|\Gamma^{-1}\| \vee \|\Gamma^{-1}\|^2)\left(\sqrt{\frac{r_2(C)}{N}} \vee \left(\frac{r_2(C)}{N}\right)^2 \vee \sqrt{\frac{t}{N}} \vee \left(\frac{t}{N}\right)^2\right)\end{aligned}$$

Next, for the PO-ensemble Kalman update, it follows by the triangle inequality that

$$\begin{aligned}\|\widehat{O}\| &\le \|\mathscr{K}(\widehat{C})(\widehat{\Gamma} - \Gamma)\mathscr{K}^\top(\widehat{C})\| &\text{(A12)}\\ &\quad + \|(I - \mathscr{K}(\widehat{C})A)\widehat{C}^{u\eta}\mathscr{K}^\top(\widehat{C})\| &\text{(A13)}\\ &\quad + \|\mathscr{K}(\widehat{C})(\widehat{C}^{u\eta})^\top(I - A^\top\mathscr{K}^\top(\widehat{C}))\|, &\text{(A14)}\end{aligned}$$

and so we may proceed by bounding each of the three terms (A12), (A13) and (A14) separately. For (A12), invoking first the bound on $\mathscr{K}$ from Lemma A.4 as well as the inequality in (A9), it holds on E that

$$\begin{aligned}\|\mathscr{K}(\widehat{C})(\widehat{\Gamma} - \Gamma)\mathscr{K}^\top(\widehat{C})\| &\le \|\mathscr{K}(\widehat{C})\|^2\|\widehat{\Gamma} - \Gamma\| \\ &\le \|A\|^2\|\Gamma^{-1}\|^2\|\widehat{C}\|^2\|\widehat{\Gamma} - \Gamma\| \\ &\lesssim \|A\|^2\|\Gamma^{-1}\|^2\|C\|^2\|\Gamma\|\left(\mathscr{R}_N(r_2(\Gamma)) \vee \mathscr{R}_N(t)\right)\left(1 \vee \mathscr{R}^2_{N,2}(C) \vee \mathscr{R}^2_{N,2}(t)\right)\end{aligned}$$

Both (A13) and (A14) are equal in operator norm, and so we consider only (A13). We use Lemmas A.4 and A.2, along with the inequalities (A10) and (A11) to show that on E,

$$\begin{aligned}
\|(I-\mathscr{K}(\widehat{C})A)\widehat{C}^{u\eta}\mathscr{K}^{\top}(\widehat{C})\| &\leq \|\mathscr{K}(\widehat{C})\|\|I-\mathscr{K}(\widehat{C})A\|\|\widehat{C}^{u\eta}\| \\
&\leq \|\mathscr{K}(\widehat{C})\|\left(1+\|\mathscr{K}(\widehat{C})\|\|A\|\right)\|\widehat{C}^{u\eta}\| \\
&\leq \|A\|\|\Gamma^{-1}\|\|\widehat{C}\|\left(1+\|A\|^2\|\Gamma^{-1}\|\|\widehat{C}\|\right)\|\widehat{C}^{u\eta}\| \\
&\lesssim (\|A\|\vee\|A\|^3)(\|\Gamma^{-1}\|\vee\|\Gamma^{-1}\|^2)[\|\widehat{C}\|+\|\widehat{C}\|^2]\|\widehat{C}^{u\eta}\| \\
&\lesssim (\|A\|\vee\|A\|^3)(\|\Gamma^{-1}\|\vee\|\Gamma^{-1}\|^2)(\|C\|\vee\|\Gamma\|)(\|C\|\vee\|C\|^2) \\
&\times\left(1\vee\mathscr{R}_N(r_2(C))\vee\mathscr{R}^2_{N,2}(C)\vee\mathscr{R}_N(t)\vee\mathscr{R}^2_{N,2}(t)\right) \\
&\times\left(\mathscr{R}_N(r_2(C))\vee\mathscr{R}_N(r_2(\Gamma))\vee\mathscr{R}_N(t)\right).
\end{aligned}$$

Some algebra shows that

$$\begin{aligned}
&\left(1\vee\mathscr{R}_N(r_2(C))\vee\mathscr{R}^2_{N,2}(C)\vee\mathscr{R}_N(t)\vee\mathscr{R}^2_{N,2}(t)\right)\left(\mathscr{R}_N(r_2(C))\vee\mathscr{R}_N(r_2(\Gamma))\vee\mathscr{R}_N(t)\right) \\
&=\left(\sqrt{\frac{r_2(C)}{N}}\vee\left(\frac{r_2(C)}{N}\right)^3\vee\sqrt{\frac{t}{N}}\vee\left(\frac{t}{N}\right)^3\vee\mathscr{R}_N(r_2(\Gamma))\left(1\vee\left(\frac{r_2(C)}{N}\right)^2\vee\left(\frac{t}{N}\right)^2\right)\right),
\end{aligned}$$

and so

$$\begin{aligned}
\|\widehat{O}\| &\lesssim (\|A\|\vee\|A\|^3)(\|\Gamma^{-1}\|\vee\|\Gamma^{-1}\|^2)(\|C\|\vee\|\Gamma\|)(\|C\|\vee\|C\|^2) \\
&\times\left(\sqrt{\frac{r_2(C)}{N}}\vee\left(\frac{r_2(C)}{N}\right)^3\vee\sqrt{\frac{t}{N}}\vee\left(\frac{t}{N}\right)^3\vee\mathscr{R}_N(r_2(\Gamma))\left(1\vee\left(\frac{r_2(C)}{N}\right)^2\vee\left(\frac{t}{N}\right)^2\right)\right).
\end{aligned}$$

□

Proof of Theorem 2.5 The proof follows similarly to that of Theorem 2.3 and is therefore omitted. The constants in the statement of the result are as follows:

$$\begin{aligned}
c_1 &= (\|C\|\vee\|C\|^3)(\|A\|^2\vee\|A\|^4)(\|\Gamma^{-1}\|\vee\|\Gamma^{-1}\|^2), \\
c_2 &= (\|A\|\vee\|A\|^3)(\|\Gamma^{-1}\|\vee\|\Gamma^{-1}\|^2)(\|C\|\vee\|\Gamma\|)(\|C\|\vee\|C\|^2).
\end{aligned}$$

□

A.4 *Multi-step analysis of the square root ensemble Kalman filter*

Here we provide a description of the multi-step EnKF algorithm discussed in Remark 2.7. As described there, we focus on the square root EnKF studied in [59]. Given an initial ensemble $\{\upsilon_n^{(0)}\}_{n=1}^N$, the algorithm iterates the steps of the square root ensemble update (2.9) with new observations $y^{(t)}$ and with

TABLE A1 *Comparison of the Kalman filter and square root EnKF considered in [59]. The forecast and analysis steps are to be repeated for $t = 1, \dots, T$ iterations*

	Kalman filter	**Square root EnKF**
Input	$\{y^{(t)}, A^{(t)}, M^{(t)}\}_{t=1}^{T}$, Γ, $\mu^{(0)}$, $\Sigma^{(0)}$	$\{y^{(t)}, A^{(t)}, M^{(t)}\}_{t=1}^{T}$, Γ, $\{\upsilon_n^{(0)}\}_{n=1}^{N} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu^{(0)}, \Sigma^{(0)})$
Forecast	$m^{(t)} = M^{(t)}\mu^{(t-1)}$ $C^{(t)} = M^{(t)}\Sigma^{(t-1)}(M^{(t)})^\top$	$u_n^{(t)} = M^{(t)}\upsilon_n^{(t-1)}, \ \ n = 1, \dots, N$ $\widehat{m}^{(t)} = \frac{1}{N}\sum_{n=1}^{N} u_n^{(t)}$ $\widehat{C}^{(t)} = \frac{1}{N-1}\sum_{n=1}^{N}(u_n^{(t)} - \widehat{m}^{(t)})(u_n^{(t)} - \widehat{m}^{(t)})^\top$
Analysis	$\mu^{(t)} = \mathscr{M}(m^{(t)}, C^{(t)}; A^{(t)}, y^{(t)}, \Gamma)$ $\Sigma^{(t)} = \mathscr{C}(C^{(t)}; A^{(t)}, \Gamma)$	$\upsilon_n^{(t)} = \mathscr{M}(u_n^{(t)}, \widehat{C}^{(t)}; A^{(t)}, y^{(t)}, \Gamma), \ \ n = 1, \dots, N$ $\widehat{\mu}^{(t)} = \frac{1}{N}\sum_{n=1}^{N} \upsilon_n^{(t)}$ $\widehat{\Sigma}^{(t)} = \frac{1}{N-1}\sum_{n=1}^{N}(\upsilon_n^{(t)} - \widehat{\mu}^{(t)})(\upsilon_n^{(t)} - \widehat{\mu}^{(t)})^\top$
Output	$\{\mu^{(t)}, \Sigma^{(t)}\}_{t=1}^{T}$	$\{\widehat{\mu}^{(t)}, \widehat{\Sigma}^{(t)}\}_{t=1}^{T}$

possibly varying model matrices $A^{(t)}$. We assume that the noise distribution does not change over time, though this assumption can easily be relaxed at the expense of more cumbersome notation. We summarize both the Kalman filter and the square root EnKF in TABLE A1. In this filtering set-up, $M^{(t)} \in \mathbb{R}^{d\times d}$ is the dynamics map and $A^{(t)} \in \mathbb{R}^{k\times d}$ is the observation map at time $t \geq 1$. As detailed in [83], such a filtering set-up leads to a sequence of inverse problems of the form (2.1), where the forward model is given by the observation map, and the prior *forecast distribution* blends the dynamics map with previous probabilistic estimates. Throughout this subsection, we write $\|\widehat{\mu} - \mu^{(t)}\|_p \equiv \left[\mathbb{E}\|\hat{\mu}^{(t)} - \mu^{(t)}\|_2^p\right]^{1/p}$ and $\|\widehat{\Sigma}^{(t)} - \Sigma^{(t)}\|_p \equiv \left[\mathbb{E}\|\widehat{\Sigma}^{(t)} - \Sigma^{(t)}\|^p\right]^{1/p}$.

We will use two auxiliary lemmas to prove the main result of this subsection, Corollary A.12 below.

LEMMA A.10. (Continuity and Boundedness of Covariance-Update Operator in L^p [59, Corollary 4.8]) Let $\mathscr{C}$ be the covariance-update operator defined in (2.4). Let $Q \in \mathcal{S}_+^d$ be a random matrix and $P \in \mathcal{S}_+^d$ be a deterministic matrix, $\Gamma \in \mathcal{S}_{++}^k$, $A \in \mathbb{R}^{k\times d}$, $y \in \mathbb{R}^k$ and $m, m' \in \mathbb{R}^d$. Then, for any $1 \leq p < \infty$, the following holds:

$$\begin{aligned}\|\mathscr{C}(Q) - \mathscr{C}(P)\|_p \leq \|Q - P\|_p(1 + \|A\|^2\|\Gamma^{-1}\|\|P\|) \\ + (\|A\|^2\|\Gamma^{-1}\| + \|A\|^4\|\Gamma^{-1}\|^2\|P\|)\|Q\|_{2p}\|Q - P\|_{2p}.\end{aligned}$$

LEMMA A.11. (Continuity and Boundedness of Mean-Update Operator in L^p [59, Corollary 4.10]) Let $\mathscr{M}$ be the mean-update operator defined in (2.3). Let $P, Q \in \mathcal{S}_+^d$, $\Gamma \in \mathcal{S}_{++}^k$, $A \in \mathbb{R}^{k\times d}$, $y \in \mathbb{R}^k$ and $m, m' \in \mathbb{R}^d$. Assume that Q and m are random, and that P and m' are deterministic. The following holds:

$$\begin{aligned}\left\|\mathscr{M}(m, Q) - \mathscr{M}(m', P)\right\|_p \leq \left\|m - m'\right\|_p + \|A\|^2\|\Gamma^{-1}\|\|Q\|_{2p}\left\|m - m'\right\|_{2p} \\ + \|Q - P\|_p\|A\|\|\Gamma^{-1}\|\left(1 + \|A\|^2\|\Gamma^{-1}\|\|P\|\right)\left\|y - Am'\right\|_2.\end{aligned}$$

The next result shows how our one-step bounds in Theorems 2.3 and 2.5 can be extended to provide non-asymptotic bounds on the performance of the multi-step square root EnKF. The proof follows a similar argument to the proof of [59, Theorem 6.1].

COROLLARY A.12. Consider the square root EnKF defined in TABLE A1. Suppose that $N \gtrsim r_2(\Sigma^{(0)})$. Then, for any $t \geq 1$ and $p \geq 1$,

$$\|\widehat{\mu}^{(t)} - \mu^{(t)}\|_p \lesssim_p \sqrt{\frac{r_2(\Sigma^{(0)})}{N}} \times c(\{\|M^{(l)}\|, \|A^{(l)}\|, \|\Sigma^{(l-1)}\|, \|y^{(l)} - A^{(l)}m^{(l)}\|\}_{l=1}^t, \|\Gamma^{-1}\|),$$

$$\|\widehat{\Sigma}^{(t)} - \Sigma^{(t)}\|_p \lesssim_p \sqrt{\frac{r_2(\Sigma^{(0)})}{N}} \times c(\{\|M^{(l)}\|, \|A^{(l)}\|, \|\Sigma^{(l-1)}\|\}_{l=1}^t, \|\Gamma^{-1}\|).$$

Proof. The proof follows by strong induction on the predicate in the statement of the theorem. To that end, the base case ($t = 1$) holds by Theorems 2.3 and 2.5, which state that, for any $p \geq 1$,

$$\|\widehat{\mu}^{(1)} - \mu^{(1)}\|_p \lesssim_p \sqrt{\frac{r_2(\Sigma^{(0)})}{N}} \times c(\|M^{(1)}\|, \|A^{(1)}\|, \|\Sigma^{(0)}\|, \|y^{(1)} - A^{(1)}m^{(1)}\|, \|\Gamma^{-1}\|),$$

$$\|\widehat{\Sigma}^{(1)} - \Sigma^{(1)}\|_p \lesssim_p \sqrt{\frac{r_2(\Sigma^{(0)})}{N}} \times c(\|M^{(1)}\|, \|A^{(1)}\|, \|\Sigma^{(0)}\|, \|\Gamma^{-1}\|).$$

Suppose now that the claim holds for $l = 2, \ldots, t-1$. Then, for $l = t$, we have from Lemma A.10

$$\begin{aligned}\|\widehat{\Sigma}^{(t)} - \Sigma^{(t)}\|_p &= \|\mathscr{C}(\widehat{C}^{(t)}) - \mathscr{C}(C^{(t)})\|_p \\ &\leq \|\widehat{C}^{(t)} - C^{(t)}\|_p(1 + \|A^{(t)}\|^2\|\Gamma^{-1}\|\|C^{(t)}\|) \\ &+ (\|A^{(t)}\|^2\|\Gamma^{-1}\| + \|A^{(t)}\|^4\|\Gamma^{-1}\|^2\|C^{(t)}\|)\|\widehat{C}^{(t)}\|_{2p}\|\widehat{C}^{(t)} - C^{(t)}\|_{2p}. \qquad \text{(A15)}\end{aligned}$$

By the definition of $\widehat{C}^{(t)}, C^{(t)}$ together with the inductive hypothesis, it follows that, for $\mathfrak{p} \in \{p, 2p\}$,

$$\begin{aligned}\|\widehat{C}^{(t)} - C^{(t)}\|_{\mathfrak{p}} &= \|M^{(t)}(\widehat{\Sigma}^{(t-1)} - \Sigma^{(t-1)})(M^{(t)})^\top\|_{\mathfrak{p}} \\ &\leq \|M^{(t)}\|^2\|\widehat{\Sigma}^{(t-1)} - \Sigma^{(t-1)}\|_{\mathfrak{p}} \\ &\lesssim_p \|M^{(t)}\|^2\sqrt{\frac{r_2(\Sigma^{(0)})}{N}} \times c(\{\|M^{(l)}\|, \|A^{(l)}\|, \|\Sigma^{(l-1)}\|\}_{l=1}^{t-1}, \|\Gamma^{-1}\|) \\ &= \sqrt{\frac{r_2(\Sigma^{(0)})}{N}} \times c(\{\|M^{(l)}\|, \|A^{(l)}\|, \|\Sigma^{(l-1)}\|\}_{l=1}^t, \|\Gamma^{-1}\|).\end{aligned}$$

Further, we have

$$\begin{aligned}\|\widehat{C}^{(t)}\|_{2p} &\le \|\widehat{C}^{(t)} - C^{(t)}\|_{2p} + \|C^{(t)}\| \\ &\lesssim_p \sqrt{\frac{r_2(\Sigma^{(0)})}{N}} c(\{\|M^{(l)}\|, \|A^{(l)}\|, \|\Sigma^{(l-1)}\|\}_{l=1}^t, \|\Gamma^{-1}\|) + \|M^{(t)}\|^2 \|\Sigma^{(t-1)}\|.\end{aligned}$$

Plugging these two results into (A15) gives

$$\|\widehat{\Sigma}^{(t)} - \Sigma^{(t)}\|_p \lesssim_p \sqrt{\frac{r_2(\Sigma^{(0)})}{N}} \times c(\{\|M^{(l)}\|, \|A^{(l)}\|, \|\Sigma^{(l-1)}\|\}_{l=1}^t, \|\Gamma^{-1}\|).$$

Similarly, from Lemma A.11, we have

$$\begin{aligned}\|\widehat{\mu}^{(t)} - \mu^{(t)}\|_p &= \|\mathscr{M}(\widehat{m}^{(t)}, \widehat{C}^{(t)}) - \mathscr{M}(m^{(t)}, C^{(t)})\|_p \\ &\le \|\widehat{m}^{(t)} - m^{(t)}\|_p + \|A^{(t)}\|^2 \|\Gamma^{-1}\| \|\widehat{C}^{(t)}\|_{2p} \|\widehat{m}^{(t)} - m^{(t)}\|_{2p} \\ &+ \|\widehat{C}^{(t)} - C^{(t)}\|_p \|A^{(t)}\| \|\Gamma^{-1}\| \big(1 + \|A^{(t)}\|^2 \|\Gamma^{-1}\| \|C^{(t)}\|\big) \|y^{(t)} - A^{(t)} m^{(t)}\|_2. \qquad \text{(A16)}\end{aligned}$$

By the definition of $\widehat{m}^{(t)}, m^{(t)}$ together with the inductive hypothesis, we have, for $\mathfrak{p} \in \{p, 2p\}$,

$$\begin{aligned}\|\widehat{m}^{(t)} - m^{(t)}\|_{\mathfrak{p}} &= \|M^{(t)}(\widehat{\mu}^{(t-1)} - \mu^{(t-1)})\|_{\mathfrak{p}} \\ &\le \|M^{(t)}\| \|\widehat{\mu}^{(t-1)} - \mu^{(t-1)}\|_{\mathfrak{p}} \\ &\lesssim_p \|M^{(t)}\| \sqrt{\frac{r_2(\Sigma^{(0)})}{N}} \times c(\{\|M^{(l)}\|, \|A^{(l)}\|, \|\Sigma^{(l-1)}\|, \|y^{(l)} - A^{(l)} m^{(l)}\|\}_{l=1}^{t-1}, \|\Gamma^{-1}\|) \\ &= \sqrt{\frac{r_2(\Sigma^{(0)})}{N}} \times c(\{\|M^{(l)}\|, \|A^{(l)}\|, \|\Sigma^{(l-1)}\|, \|y^{(l)} - A^{(l)} m^{(l)}\|\}_{l=1}^{t}, \|\Gamma^{-1}\|).\end{aligned}$$

Plugging this bound and the one for $\|\widehat{C}^{(t)}\|_{2p}$ derived previously in the proof into (A16) yields

$$\|\widehat{\mu}^{(t)} - \mu^{(t)}\|_p \lesssim_p \sqrt{\frac{r_2(\Sigma^{(0)})}{N}} \times c(\{\|M^{(l)}\|, \|A^{(l)}\|, \|\Sigma^{(l-1)}\|, \|y^{(l)} - A^{(l)} m^{(l)}\|\}_{l=1}^{t}, \|\Gamma^{-1}\|).$$

□

B. Proofs: Section 3

This appendix contains the proofs of all the theorems in Section 3. Results on covariance estimation are in Subsection B.1 and our main results on ensemble Kalman updates are in Subsection B.2.

B.1 *Covariance estimation*

Here we establish Theorems 3.1 and 3.3. We first collect some required technical results in Subsection B.1.1. Next we study covariance and cross-covariance estimation under soft sparsity in Subsections B.1.2 and B.1.3, respectively.

B.1.1 *Background and preliminaries* DEFINITION B.1. ([92, Definition 2.2.17]) Given a set T, an admissible sequence of partitions of T is an increasing sequence (Δ_n) of partitions of T such that $\mathrm{card}(\Delta_0) = 1$ and $\mathrm{card}(\Delta_n) \leq 2^{2^n}$ for $n \geq 1$.
The notion of an admissible sequence of partitions allows us to define the following notion of complexity of a set T, often referred to as *generic complexity*.
DEFINITION B.2. ([92, Definition 2.2.19]) Let (T, d) be a possibly infinite metric space, and define

$$\gamma_2(T, \mathsf{d}) = \inf \sup_{t \in T} \sum_{n \geq 0} 2^{n/2} \mathrm{Diam}\big(\Delta_n(t)\big),$$

where $\Delta_n(t)$ denotes the unique element of the partition to which t belongs, and the infimum is taken over all admissible sequences of partitions.

The following theorem is known as the Majorizing Measure Theorem and provides upper and lower bounds for centered Gaussian processes in terms of the generic complexity.

THEOREM B.3. ([92, Theorem 2.4.1]) Let X_t, $t \in T$ be a centered Gaussian process that induces a metric $\mathsf{d}_X : T \times T \to [0, \infty]$ defined by

$$\mathsf{d}_X^2(s, t) = \mathbb{E}\left[(X_s - X_t)^2\right].$$

Then there exists an absolute constant $L > 0$ such that

$$\frac{1}{L}\gamma_2(T, \mathsf{d}_X) \leq \mathbb{E}\left[\sup_{t \in T} X_t\right] \leq L\gamma_2(T, \mathsf{d}_X).$$

We will be primarily interested in the case that $T = \mathcal{F}$ is some function class on the probability space $(\mathcal{X}, \mathsf{A}, \mathbb{P})$, and with d being the metric induced either by $\|\cdot\|_{L_2}$ or $\|\cdot\|_{\psi_2}$. We denote these spaces by $(\mathcal{F}, L_2)$ and $(\mathcal{F}, \psi_2)$, respectively, throughout this section. The next result is an exponential generic chaining bound, which was introduced in [28, Corollary 5.7] and described in [57, Theorem 8]. We present it as it was described in the latter reference.

THEOREM B.4. ([57, Theorem 8]) Let $(\mathcal{X}, \mathsf{A}, \mathbb{P})$ be a probability space and consider the random sample $X, X_1, \ldots, X_N \overset{\text{i.i.d.}}{\sim} \mathbb{P}$. Let $\mathcal{F}$ be a class of measurable functions on $(\mathcal{X}, \mathsf{A})$. There exists a universal

constant $c > 0$ such that, for all $t \geq 1$, it holds with probability at least $1 - e^{-t}$ that

$$\sup_{f\in\mathcal{F}}\left|\frac{1}{N}\sum_{n=1}^{N} f^2(X_n) - \mathbb{E}[f^2(X)]\right| \leq c\left(\sup_{f\in\mathcal{F}}\|f\|_{\psi_2}\frac{\gamma_2(\mathcal{F},\psi_2)}{\sqrt{N}} \vee \frac{\gamma_2^2(\mathcal{F},\psi_2)}{N} \vee \sup_{f\in\mathcal{F}}\|f\|_{\psi_2}^2\sqrt{\frac{t}{N}} \vee \sup_{f\in\mathcal{F}}\|f\|_{\psi_2}^2\frac{t}{N}\right).$$

LEMMA B.5. (Expectation Bound from Probability Bound, [92, Lemma 2.2.3]) Let $Y \geq 0$ be a random variable satisfying

$$\mathbb{P}(Y \geq r) \leq a\exp\left(-\frac{r^2}{b^2}\right), \qquad r \geq 0,$$

for certain numbers $a \geq 2$ and $b > 0$. Then there is a universal constant c such that

$$\mathbb{E}[Y] \leq cb\sqrt{\log a}.$$

Finally, we recall the following dimension-free bound for the maxima of sub-Gaussian random variables.

LEMMA B.6. (Dimension-Free Sub-Gaussian Maxima, [101, Lemma 2.4]) Let $X_1, \ldots X_N$ be not necessarily independent sub-Gaussian random variables with

$$\mathbb{P}(X_n > x) \leq ce^{-x^2/c\sigma_n^2}, \qquad \text{for all} x \geq 0,\ 1 \leq n \leq N,$$

where $\sigma_n \geq 0$ is given, or alternatively $\|X_n\|_{\psi_2} \lesssim \sigma_n$. Then, for any $t \geq 1$, it holds with probability at least $1 - ce^{-ct}$ that

$$\max_{n\leq N} X_n \lesssim \sqrt{t}\max_{n\leq N}\sigma_{(n)}\sqrt{\log(n+1)},$$

where $\sigma_{(1)} \geq \sigma_{(2)} \geq \ldots \sigma_{(N)}$ is the decreasing rearrangement of $\sigma_1, \ldots, \sigma_N$. Further,

$$\mathbb{E}\left[\max_{n\leq N} X_n\right] \lesssim \max_{n\leq N}\sigma_{(n)}\sqrt{\log(n+1)}.$$

Proof. The proof of the upper bound is based on the proof of Proposition 2.4.16 in [92]. By permutation invariance, we can assume without loss of generality that $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_N$. Then

$$\mathbb{P}\left(\max_{n\leq N} \frac{X_n}{\sigma_n\sqrt{\log(n+1)}} \geq \sqrt{t}\right) \leq \sum_{n=1}^{N} \mathbb{P}\left(X_n \geq \sigma_n\sqrt{t\log(n+1)}\right) \lesssim \sum_{n=1}^{N} \exp\left(-\frac{t}{c}\log(n+1)\right).$$

For $t \geq 2c$, the final expression in the above display is finite, and we may write

$$\sum_{n=1}^{N} \exp\left(-\frac{t}{c}\log(n+1)\right) = \sum_{n=2}^{N+1} \exp\left(-\frac{t}{c}\log(n)\right) \leq \exp\left(-\frac{t}{c}\log(2)\right) + \int_2^{\infty} x^{-t/c}dx \leq ce^{-t/c}.$$

Therefore, for any $t \geq 2c$, it holds with probability at least $1 - ce^{-t/c}$ that

$$\max_{n\leq N} X_n \lesssim \sqrt{t}\max_{n\leq N} \sigma_{(n)}\sqrt{\log(n+1)}.$$

This implies that, for any $t \geq 1$, it holds with probability at least $1 - ce^{-(t\vee 2c)/c}$ that

$$\max_{n\leq N} X_n \lesssim (\sqrt{t} \vee \sqrt{2c})\max_{n\leq N} \sigma_{(n)}\sqrt{\log(n+1)} \lesssim \sqrt{t}\max_{n\leq N} \sigma_{(n)}\sqrt{\log(n+1)}.$$

Since $1 - ce^{-(t\vee 2c)/c} \geq 1 - ce^{-t/c}$, it holds that, for any $t \geq 1$, with probability at least $1 - ce^{-t/c}$

$$\max_{n\leq N} X_n \lesssim \sqrt{t}\max_{n\leq N} \sigma_{(n)}\sqrt{\log(n+1)}.$$

It follows from Lemma B.5 that

$$\mathbb{E}\left[\max_{n\leq N} \frac{X_n}{\sigma_n\sqrt{\log(n+1)}}\right] \leq c,$$

which, in turn, implies

$$\mathbb{E}\left[\max_{n\leq N} X_n\right] \lesssim \max_{n\leq N} \sigma_n\sqrt{\log(n+1)} = \max_{n\leq N} \sigma_{(n)}\sqrt{\log(n+1)}.$$

□

B.1.2 *Covariance estimation under soft sparsity* This subsection contains the proof of Theorem 3.1. We follow the approach in [57, Theorem 4], but we restrict our attention to finite dimensional spaces. Our proof will rely on the following max-norm covariance estimation bound, which may be of independent interest.

THEOREM B.7. (Covariance Estimation with Sample Covariance—Max-Norm Bound) Let $X_1, \dots, X_N$ be d-dimensional i.i.d. sub-Gaussian random vectors with $\mathbb{E}[X_1] = \mu^X$ and $\mathrm{var}(X_1) = \Sigma^X$. Let $\widehat{\Sigma}^X = (N-1)^{-1} \sum_{n=1}^N (X_n - \mu^X)(X_n - \mu^X)^\top$. Then there exists a constant c such that, for all $t \geq 1$, it holds with probability at least $1 - ce^{-t}$ that

$$\|\widehat{\Sigma}^X - \Sigma^X\|_{\max} \leq c\Sigma^X_{(1)} \left(\sqrt{\frac{r_\infty(\Sigma^X)}{N}} \vee \sqrt{\frac{t}{N}} \vee \frac{t}{N} \vee \frac{t r_\infty(\Sigma^X)}{N} \right),$$

where

$$r_\infty(\Sigma^X) \equiv \frac{\max_j \Sigma^X_{(j)} \log(j+1)}{\Sigma^X_{(1)}}.$$

Proof. The proof of this result is based on the proof of the upper bound of Theorem 4 of [57], in conjunction with Theorem B.4. We deal with the case $\mu^X = 0$ first. To this end, let $Z_1, \dots, Z_N$ be d-dimensional i.i.d. sub-Gaussian random vectors with zero mean and $\mathrm{var}[Z_1] = \Sigma^X$. We denote the distribution of Z_1 by $\mathbb{P}$, and note that $\|\cdot\|_{\psi_1}$, $\|\cdot\|_{\psi_2}$, and $\|\cdot\|_{L_2}$ are defined implicitly with respect to $\mathbb{P}$. Let $\widehat{\Sigma}^0 = N^{-1} \sum_{n=1}^N Z_n Z_n^\top$. We rewrite the expectation of interest as a squared empirical process term over an appropriate class of functions. For $j \geq 1$ we denote the jth canonical vector (the vector with 1 in the jth index and zero otherwise) by e_j. Then, we note that

$$\begin{aligned}
\|\widehat{\Sigma}^0 - \Sigma^X\|_{\max} &= \sup_{i,j} \left\langle e_i, (\widehat{\Sigma}^0 - \Sigma^X) e_j \right\rangle \\
&= \sup_{i,j} \left[\left\langle \frac{e_i + e_j}{2}, (\widehat{\Sigma}^0 - \Sigma^X) \frac{e_i + e_j}{2} \right\rangle - \left\langle \frac{e_i - e_j}{2}, (\widehat{\Sigma}^0 - \Sigma^X) \frac{e_i - e_j}{2} \right\rangle \right] \\
&\leq 2 \sup_{u \in \mathcal{U}} \left| \left\langle (\widehat{\Sigma}^0 - \Sigma^X) u, u \right\rangle \right|,
\end{aligned}$$

where $\mathcal{U} = \left\{ u \in \mathbb{R}^d : u = \pm\frac{1}{2}(e_i \pm e_j),\ 1 \leq i, j \leq d \right\}$. Define the set of functions $\mathcal{F}_U = \left\{ \langle \cdot, u \rangle : u \in \mathcal{U} \right\}$, and note that, for any $f \in \mathcal{F}_\mathcal{U}$, $-f \in \mathcal{F}_\mathcal{U}$ and $\mathbb{E}[f(Z_1)] = 0$. It then follows from Theorem B.4 that for

the same universal constant c in the statement of the theorem,

$$
\begin{aligned}
&2\sup_{u\in\mathcal{U}}\left|\left\langle(\widehat{\Sigma}^0-\Sigma^X)u,u\right\rangle\right| \\
&=2\sup_{u\in\mathcal{U}}\left|\frac{1}{N}\sum_{n=1}^N\langle Z_n,u\rangle^2-\left\langle u,\Sigma^X u\right\rangle\right| \\
&=2\sup_{f\in\mathcal{F}_\mathcal{U}}\left|\frac{1}{N}\sum_{n=1}^N f^2(Z_n)-\mathbb{E}[f^2(Z_1)]\right| \\
&\le 2c\left(\sup_{f\in\mathcal{F}_\mathcal{U}}\|f\|_{\psi_2}\frac{\gamma_2(\mathcal{F}_\mathcal{U};\psi_2)}{\sqrt{N}}\vee\frac{\gamma_2^2(\mathcal{F}_\mathcal{U};\psi_2)}{N}\vee\sup_{f\in\mathcal{F}_\mathcal{U}}\|f\|_{\psi_2}^2\sqrt{\frac{t}{N}}\vee\sup_{f\in\mathcal{F}_\mathcal{U}}\|f\|_{\psi_2}^2\frac{t}{N}\right).
\end{aligned}
$$

Using the equivalence of the ψ_2 and L_2 norms for linear functionals, we have

$$
\begin{aligned}
\sup_{f\in\mathcal{F}_\mathcal{U}}\|f\|_{\psi_2}&\lesssim\sup_{f\in\mathcal{F}_\mathcal{U}}\|f\|_{L_2}=\max_{u\in\mathcal{U}}\sqrt{\mathbb{E}\big[\langle Z_1,u\rangle^2\big]}=\max_{u\in\mathcal{U}}\sqrt{\langle u,\Sigma^X u\rangle}=\frac{1}{2}\max_{i,j}\sqrt{\langle e_i\pm e_j,\Sigma^X(e_i\pm e_j)\rangle} \\
&=\frac{1}{2}\max_{i,j}\sqrt{\langle e_i,\Sigma^X e_i\rangle+\langle e_j,\Sigma^X e_j\rangle\pm 2\langle e_i,\Sigma^X e_j\rangle}=\frac{1}{2}\max_{i,j}\sqrt{\Sigma^X_{ii}+\Sigma^X_{jj}\pm 2\Sigma^X_{ij}}\le\sqrt{\Sigma^X_{(1)}}.
\end{aligned}
$$

To control the generic complexity $\gamma_2(\mathcal{F}_U,\psi_2)$, let $Y\sim\mathcal{N}(0,\Sigma^X)$ be a d-dimensional Gaussian vector, with induced metric

$$
\mathsf{d}_Y(u,v)=\sqrt{\mathbb{E}\big[(\langle Y,u\rangle-\langle Y,v\rangle)^2\big]}=\|\langle\cdot,u\rangle-\langle\cdot,v\rangle\|_{L_2},\qquad u,v\in\mathcal{U}.
$$

Using again the equivalence of the ψ_2 and L_2 norms for linear functionals, we have that

$$
\gamma_2(\mathcal{F}_\mathcal{U};\psi_2)\lesssim\gamma_2(\mathcal{F}_\mathcal{U};L_2)=\gamma_2(\mathcal{U};\mathsf{d}_Y).
$$

It follows then from Theorem B.3 that

$$
\begin{aligned}
\gamma_2(\mathcal{U};\mathsf{d}_Y)&\lesssim\mathbb{E}\left[\sup_{u\in\mathcal{U}}\langle Y,u\rangle\right] \\
&=\mathbb{E}\left[\max_{i,j}\left\langle Y,\pm\frac{1}{2}(e_i\pm e_j)\right\rangle\right] \\
&\le\mathbb{E}\left[\max_j\left|\left\langle Y,e_j\right\rangle\right|\right] \\
&\lesssim\max_j\sqrt{\Sigma^X_{(j)}\log(j+1)},
\end{aligned}
$$

where the final inequality follows from Lemma B.6. We have shown that with probability at least $1-e^{-t}$

$$\begin{aligned}\|\widehat{\Sigma}^0 - \Sigma^X\|_{\max} &\lesssim \left(\sqrt{\Sigma^X_{(1)} \max_j \frac{\Sigma^X_{(j)}\log(j+1)}{N}} \vee \max_j \frac{\Sigma^X_{(j)}\log(j+1)}{N} \vee \Sigma^X_{(1)}\sqrt{\frac{t}{N}} \vee \Sigma^X_{(1)}\frac{t}{N}\right) \\ &= \Sigma^X_{(1)}\left(\sqrt{\frac{r_\infty(\Sigma^X)}{N}} \vee \frac{r_\infty(\Sigma^X)}{N} \vee \sqrt{\frac{t}{N}} \vee \frac{t}{N}\right). \end{aligned} \tag{B1}$$

In the un-centered case, taking $X_n = Z_n + \mu^X$, we have $\widehat{\Sigma}^X = \widehat{\Sigma}^0 - \bar{Z}\bar{Z}^\top$, which follows that

$$\|\widehat{\Sigma}^X - \Sigma^X\|_{\max} \le \|\widehat{\Sigma}^0 - \Sigma^X\|_{\max} + \|\bar{Z}\bar{Z}^\top\|_{\max}.$$

By Lemma B.6, with probability at least $1 - ce^{-t}$

$$\|\bar{Z}\bar{Z}^\top\|_{\max} \le \|\bar{Z}\|^2_{\max} \le \frac{t}{N}\max_{j\le d}\Sigma^X_{(j)}\log(j+1) = t\Sigma^X_{(1)}\frac{r_\infty(\Sigma^X)}{N}. \tag{B2}$$

Denote the set on which (B1) occurs by E_1, and the set on which (B2) occurs by E_2. Then the intersection $E = E_1 \cap E_2$ has probability at least $1 - ce^{-t}$, and it holds on E that

$$\begin{aligned}\|\widehat{\Sigma}^X - \Sigma^X\|_{\max} &\lesssim \Sigma^X_{(1)}\left(\sqrt{\frac{r_\infty(\Sigma^X)}{N}} \vee \frac{r_\infty(\Sigma^X)}{N} \vee \sqrt{\frac{t}{N}} \vee \frac{t}{N} \vee \frac{tr_\infty(\Sigma^X)}{N}\right) \\ &= \Sigma^X_{(1)}\left(\sqrt{\frac{r_\infty(\Sigma^X)}{N}} \vee \sqrt{\frac{t}{N}} \vee \frac{t}{N} \vee \frac{tr_\infty(\Sigma^X)}{N}\right).\end{aligned}$$

□

LEMMA B.8. Let $X_1, \dots, X_N$ be d-dimensional i.i.d. sub-Gaussian random vectors with $\mathbb{E}[X_1] = \mu^X$ and $\mathrm{var}[X_1] = \Sigma^X$. Let $\widehat{\Sigma}^X = (N-1)^{-1}\sum_{n=1}^N (X_n - \mu^X)(X_n - \mu^X)^\top$. Then, for any $p \ge 1$,

$$\left[\mathbb{E}\|\widehat{\Sigma}^X - \Sigma^X\|^p_{\max}\right]^{1/p} \lesssim_p \Sigma^X_{(1)}\left(\sqrt{\frac{r_\infty(\Sigma^X)}{N}} \vee \frac{r_\infty(\Sigma^X)}{N}\right).$$

Proof. To ease notation, let $B \equiv \Sigma^X_{(1)}\left(\sqrt{\frac{r_\infty(\Sigma^X)}{N}} \vee \frac{r_\infty(\Sigma^X)}{N}\right)$, then using that for positive W, $\mathbb{E}[W^p] = p\int_0^\infty w^{p-1}\mathbb{P}(W > w)\,dw$ gives

$$\begin{aligned}
\left[\mathbb{E}\|\widehat{\Sigma} - \Sigma\|^p_{\max}\right] &= p\int_0^\infty x^{p-1}\mathbb{P}(\|\widehat{\Sigma}^X - \Sigma^X\|_{\max} > x)\,dx \\
&\le p\int_0^B x^{p-1}dx + \int_B^\infty x^{p-1}\mathbb{P}(\|\widehat{\Sigma}^X - \Sigma^X\|_{\max} > x)\,dx \\
&\lesssim B^p + p\int_0^\infty x^{p-1}\exp\left(-\min\left(\frac{Nx^2}{(\Sigma^X_{(1)})^2}, \frac{Nx}{\Sigma^X_{(1)}}, \frac{Nx}{r_\infty(\Sigma^X)\Sigma^X_{(1)}}\right)\right)dx \\
&= B^p + p\max\left(\frac{\Gamma(p/2)}{2}\left(\frac{(\Sigma^X_{(1)})^2}{N}\right)^{p/2}, \Gamma(p)\left(\frac{\Sigma^X_{(1)}}{N}\right)^p, \Gamma(p)\left(\frac{r_\infty(\Sigma^X)\Sigma^X_{(1)}}{N}\right)^p\right),
\end{aligned}$$

where the last line follows by direct integration. We therefore have

$$\begin{aligned}
\left[\mathbb{E}\|\widehat{\Sigma}^X - \Sigma^X\|^p_{\max}\right]^{1/p} &\lesssim B + c(p)\max\left(\frac{\Sigma^X_{(1)}}{\sqrt{N}}, \frac{\Sigma^X_{(1)}}{N}, \frac{r_\infty(\Sigma^X)\Sigma^X_{(1)}}{N}\right) \\
&\le c(p)\,\Sigma^X_{(1)}\left(\sqrt{\frac{r_\infty(\Sigma^X)}{N}} \vee \frac{r_\infty(\Sigma^X)}{N}\right),
\end{aligned}$$

where the final inequality holds due to the fact that $r_\infty(\Sigma^X) \gtrsim 1$. □

THEOREM B.9. (Covariance Estimation with Localized Sample Covariance—Operator-Norm Bound) Let $X_1, \ldots, X_N$ be d-dimensional i.i.d. sub-Gaussian random vectors with $\mathbb{E}[X_1] = \mu^X$ and $\mathrm{var}[X_1] = \Sigma^X$. Further, assume that $\Sigma^X \in \mathscr{U}_d(q, R_q)$ for some $q \in [0, 1)$ and $R_q > 0$. Let $\widehat{\Sigma}^X = (N-1)^{-1}\sum_{n=1}^N (X_n - \overline{X})(X_n - \overline{X})^\top$ and, for any $t \ge 1$, set

$$\rho_N \asymp \Sigma^X_{(1)}\left(\sqrt{\frac{r_\infty(\Sigma^X)}{N}} \vee \sqrt{\frac{t}{N}} \vee \frac{t}{N} \vee \frac{t r_\infty(\Sigma^X)}{N}\right)$$

and let $\widehat{\Sigma}^X_{\rho_N}$ be the localized sample covariance estimator. There exists a constant $c > 0$ such that, with probability at least $1 - ce^{-t}$, it holds that

$$\|\widehat{\Sigma}^X_{\rho_N} - \Sigma^X\| \lesssim R_q \rho_N^{1-q}.$$

Proof. The localized sample covariance matrix has elements

$$[\widehat{\Sigma}^X_{\rho_N}]_{ij} = \widehat{\Sigma}^X_{ij}\mathbf{1}_{|\widehat{\Sigma}^X_{ij}|\geq\rho_N}, \qquad 1 \leq i,j \leq d.$$

From Theorem B.7, it holds with probability at least $1 - ce^{-t}$ that

$$\|\widehat{\Sigma}^X - \Sigma^X\|_{\max} \lesssim \rho_N.$$

The remainder of the analysis is carried out conditional on this event, following the approach taken in [103, Theorem 6.27]. Define the set of indices of the ith row of Σ^X that exceed $\rho_N/2$ by

$$\mathcal{I}_i(\rho_N/2) \equiv \left(j \in (1,\ldots,d) : \left|\Sigma^X_{ij}\right| \geq \rho_N/2\right), \qquad i = 1,\ldots,d.$$

We then have

$$\begin{aligned}
\|\Sigma^X - \widehat{\Sigma}^X_{\rho_N}\| &\leq \|\Sigma^X - \widehat{\Sigma}^X_{\rho_N}\|_\infty \\
&= \max_{i=1,\ldots,d} \sum_{j=1}^{d} \left|\Sigma^X_{ij} - \widehat{\Sigma}^X_{ij}\mathbf{1}_{|\widehat{\Sigma}^X_{ij}|\geq\rho_N}\right| \\
&= \max_{i=1,\ldots,d} \left(\sum_{j\in\mathcal{I}_i(\rho_N/2)} \left|\Sigma^X_{ij} - \widehat{\Sigma}^X_{ij}\mathbf{1}_{|\widehat{\Sigma}^X_{ij}|\geq\rho_N}\right| + \sum_{j\notin\mathcal{I}_i(\rho_N/2)} \left|\Sigma^X_{ij} - \widehat{\Sigma}^X_{ij}\mathbf{1}_{|\widehat{\Sigma}^X_{ij}|\geq\rho_N}\right| \right),
\end{aligned}$$

where $\widehat{\Sigma}^X_{ij}$ is element (i,j) of $\widehat{\Sigma}^X$. For $j \in \mathcal{I}_i(\rho_N/2)$, it holds that $|\Sigma^X_{ij}| \geq \rho_N/2$ so that

$$\begin{aligned}
\sum_{j\in\mathcal{I}_i(\rho_N/2)} \left|\Sigma^X_{ij} - \widehat{\Sigma}^X_{ij}\mathbf{1}_{|\widehat{\Sigma}^X_{ij}|\geq\rho_N}\right| &\leq \sum_{j\in\mathcal{I}_i(\rho_N/2)} \left|\Sigma^X_{ij} - \widehat{\Sigma}^X_{ij}\right| + \left|\widehat{\Sigma}^X_{ij} - \widehat{\Sigma}^X_{ij}\mathbf{1}_{|\widehat{\Sigma}^X_{ij}|\geq\rho_N}\right| \\
&\leq \sum_{j\in\mathcal{I}_i(\rho_N/2)} \|\Sigma^X_{ij} - \widehat{\Sigma}^X_{ij}\|_{\max} + \left|\widehat{\Sigma}^X_{ij} - \widehat{\Sigma}^X_{ij}\mathbf{1}_{|\widehat{\Sigma}^X_{ij}|\geq\rho_N}\right| \\
&\leq \sum_{j\in\mathcal{I}_i(\rho_N/2)} \left(\frac{\rho_N}{2} + \rho_N\right) \\
&= |\mathcal{I}_i(\rho_N/2)|\frac{3\rho_N}{2},
\end{aligned}$$

where we have used the fact that

$$\left|\widehat{\Sigma}^X_{ij} - \widehat{\Sigma}^X_{ij}\mathbf{1}_{|\widehat{\Sigma}^X_{ij}|\geq\rho_N}\right| = 0 \times \mathbf{1}_{|\widehat{\Sigma}^X_{ij}|\geq\rho_N} + \widehat{\Sigma}^X_{ij} \times \mathbf{1}_{|\widehat{\Sigma}^X_{ij}|\leq\rho_N} \leq \rho_N.$$

Further, since

$$R_q \geq \sum_{j=1}^{d} |\Sigma_{ij}^X|^q \geq |\mathcal{I}_i(\rho_N/2)| \left(\frac{\rho_N}{2}\right)^q,$$

it follows that $|\mathcal{I}_i(\rho_N/2)| \leq 2^q \rho_N^{-q} R_q$, and so

$$\sum_{j \in \mathcal{I}_i(\rho_N/2)} \left| \Sigma_{ij}^X - \widehat{\Sigma}_{ij}^X \mathbf{1}_{|\widehat{\Sigma}_{ij}^X| \geq \rho_N} \right| \leq |\mathcal{I}_i(\rho_N/2)| \frac{3\rho_N}{2} \leq \frac{3}{2} 2^{-q} \rho_N^{1-q} R_q.$$

For $j \notin \mathcal{I}_i(\rho_N)$, then $|\Sigma_{ij}^X| \leq \rho_N/2$ and so

$$|\widehat{\Sigma}_{ij}^X| \leq |\widehat{\Sigma}_{ij}^X - \Sigma_{ij}^X| + |\Sigma_{ij}^X| \leq \|\widehat{\Sigma}^X - \Sigma^X\|_{\max} + |\Sigma_{ij}^X| \leq \frac{\rho_N}{2} + \frac{\rho_N}{2} = \rho_N.$$

This implies that $\widehat{\Sigma}_{ij}^X \mathbf{1}_{|\widehat{\Sigma}_{ij}^X| \geq \rho_N} = 0$, and, therefore, for $q \in [0, 1)$, since $|\Sigma_{ij}^X|/(\rho_N/2) \leq 1$, it holds that

$$\begin{aligned} \sum_{j \notin \mathcal{I}_i(\rho_N/2)} \left| \Sigma_{ij}^X - \widehat{\Sigma}_{ij}^X \mathbf{1}_{|\widehat{\Sigma}_{ij}^X| \geq \rho_N} \right| &\leq \sum_{j \notin \mathcal{I}_i(\rho_N/2)} |\Sigma_{ij}^X| = \frac{\rho_N}{2} \sum_{j \notin \mathcal{I}_i(\rho_N/2)} \frac{|\Sigma_{ij}^X|}{\frac{\rho_N}{2}} \\ &\leq \frac{\rho_N}{2} \sum_{j \notin \mathcal{I}_i(\rho_N/2)} \left(\frac{|\Sigma_{ij}^X|}{\rho_N/2} \right)^q \leq \rho_N^{1-q} R_q. \end{aligned}$$

Combining these two results gives

$$\|\Sigma^X - \widehat{\Sigma}_{\rho_N}^X\| \leq 4 \rho_N^{1-q} R_q.$$

□

Proof of Theorem 3.1 The result follows immediately from Theorem B.9. □

B.1.3 *Cross-covariance estimation under soft sparsity* This subsection contains the proof of Theorem 3.3. The presentation is parallel to that in Subsection B.1.2. We will use a max-norm cross-covariance estimation bound, analogous to Theorem B.7. The proof relies on a high probability bound for product function classes that was shown in [74, Theorem 1.13]. We present here a simplified version of that more general statement that suffices for our purposes.

THEOREM B.10. Let $(\mathcal{X}, \mathsf{A}, \mathbb{P})$ be a probability space and consider the random sample $X, X_1, \ldots, X_N \overset{\text{i.i.d.}}{\sim} \mathbb{P}$. Let $\mathcal{F}, \mathcal{G}$ be two classes of measurable functions on $(\mathcal{X}, \mathsf{A})$ such that $0 \in \mathcal{F}$ and $0 \in \mathcal{G}$. There exist

positive universal constants c_1, c_2, c_3 such that, for all $t \geq 1$, it holds with probability at least $1 - c_1 e^{-c_2 t}$ that

$$\sup_{f\in\mathcal{F},g\in\mathcal{G}} \left| \frac{1}{N}\sum_{n=1}^{N} f(X_n)g(X_n) - \mathbb{E}\big[f(X)g(X)\big] \right| \leq c_3 \left[\left(\frac{t}{N} \vee \sqrt{\frac{t}{N}} \right) \left(\sup_{f\in\mathcal{F}} \|f\|_{\psi_2} \gamma_2(\mathcal{G},\psi_2) \vee \sup_{g\in\mathcal{G}} \|g\|_{\psi_2} \gamma_2(\mathcal{F},\psi_2) \right) \vee \frac{\gamma_2(\mathcal{F},\psi_2)\gamma_2(\mathcal{G},\psi_2)}{N} \right].$$

Proof. For notational brevity, throughout this proof, we write $\gamma_2(\mathcal{F})$ instead of $\gamma_2(\mathcal{F},\psi_2)$ and $\mathsf{d}_{\psi_2}(\mathcal{F})$ instead of $\sup_{f\in\mathcal{F}}\|f\|$ and similarly for the class $\mathcal{G}$. The result follows by an application of [74, Theorem 1.13] and the ensuing remark, which deals with the case $\mathcal{F} = \mathcal{G}$, but is easily extended to the general case considered here. Together they imply that, for any $u \geq 1$, it holds with probability at least $1 - 2\exp\left(-cu^2\left(\frac{\gamma_2^2(\mathcal{F})}{\mathsf{d}_{\psi_2}^2(\mathcal{F})} \wedge \frac{\gamma_2^2(\mathcal{G})}{\mathsf{d}_{\psi_2}^2(\mathcal{G})}\right)\right)$ that

$$\sup_{f\in\mathcal{F},g\in\mathcal{G}} \left| \frac{1}{N}\sum_{n=1}^{N} f(X_n)g(X_n) - \mathbb{E}\big[f(X)g(X)\big] \right| \lesssim \frac{u^2}{N}\gamma_2(\mathcal{F})\gamma_2(\mathcal{G}) + \frac{u}{\sqrt{N}}\left(\gamma_2(\mathcal{F})\mathsf{d}_{\psi_2}(\mathcal{G}) + \gamma_2(\mathcal{G})\mathsf{d}_{\psi_2}(\mathcal{F})\right). \quad \text{(B3)}$$

We seek to rewrite (B3) so that all problem specific terms appear only in the upper bound. To this end, let

$$t \equiv u^2\left(\frac{\gamma_2^2(\mathcal{F})}{\mathsf{d}_{\psi_2}^2(\mathcal{F})} \wedge \frac{\gamma_2^2(\mathcal{G})}{\mathsf{d}_{\psi_2}^2(\mathcal{G})}\right) \implies u = \sqrt{t}\left(\frac{\mathsf{d}_{\psi_2}(\mathcal{F})}{\gamma_2(\mathcal{F})} \vee \frac{\mathsf{d}_{\psi_2}(\mathcal{G})}{\gamma_2(\mathcal{G})}\right)$$

and note that since $u \geq 1$, it must hold that $t \geq \left(\frac{\gamma_2^2(\mathcal{F})}{\mathsf{d}_{\psi_2}^2(\mathcal{F})} \wedge \frac{\gamma_2^2(\mathcal{G})}{\mathsf{d}_{\psi_2}^2(\mathcal{G})}\right)$. Therefore, for any $t \geq \left(\frac{\gamma_2^2(\mathcal{F})}{\mathsf{d}_{\psi_2}^2(\mathcal{F})} \wedge \frac{\gamma_2^2(\mathcal{G})}{\mathsf{d}_{\psi_2}^2(\mathcal{G})}\right)$, we have that with probability at least $1 - 2e^{-ct}$, the right-hand side of (B3) becomes

$$\frac{t}{N}\left(\frac{\mathsf{d}_{\psi_2}^2(\mathcal{F})}{\gamma_2^2(\mathcal{F})} \vee \frac{\mathsf{d}_{\psi_2}^2(\mathcal{G})}{\gamma_2^2(\mathcal{G})}\right)\gamma_2(\mathcal{F})\gamma_2(\mathcal{G}) + \frac{\sqrt{t}}{\sqrt{N}}\left(\frac{\mathsf{d}_{\psi_2}(\mathcal{F})}{\gamma_2(\mathcal{F})} \vee \frac{\mathsf{d}_{\psi_2}(\mathcal{G})}{\gamma_2(\mathcal{G})}\right)\left(\gamma_2(\mathcal{F})\mathsf{d}_{\psi_2}(\mathcal{G}) + \gamma_2(\mathcal{G})\mathsf{d}_{\psi_2}(\mathcal{F})\right).$$

The above implies that, for any $t \geq 1$, it holds with probability at least

$$1 - 2\exp\left(-c\left(t \vee \left(\frac{\gamma_2^2(\mathcal{F})}{\mathsf{d}_{\psi_2}^2(\mathcal{F})} \wedge \frac{\gamma_2^2(\mathcal{G})}{\mathsf{d}_{\psi_2}^2(\mathcal{G})}\right)\right)\right) \geq 1 - 2e^{-ct},$$

that

$$\frac{1}{N}\left(t \vee \left(\frac{\gamma_2^2(\mathcal{F})}{\mathsf{d}_{\psi_2}^2(\mathcal{F})} \wedge \frac{\gamma_2^2(\mathcal{G})}{\mathsf{d}_{\psi_2}^2(\mathcal{G})}\right)\right)\left(\frac{\mathsf{d}_{\psi_2}^2(\mathcal{F})}{\gamma_2^2(\mathcal{F})} \vee \frac{\mathsf{d}_{\psi_2}^2(\mathcal{G})}{\gamma_2^2(\mathcal{G})}\right)\gamma_2(\mathcal{F})\gamma_2(\mathcal{G}) \tag{B4}$$

$$+\frac{1}{\sqrt{N}}\left(\sqrt{t} \vee \left(\frac{\gamma_2(\mathcal{F})}{\mathsf{d}_{\psi_2}(\mathcal{F})} \wedge \frac{\gamma_2(\mathcal{G})}{\mathsf{d}_{\psi_2}(\mathcal{G})}\right)\right)\left(\frac{\mathsf{d}_{\psi_2}(\mathcal{F})}{\gamma_2(\mathcal{F})} \vee \frac{\mathsf{d}_{\psi_2}(\mathcal{G})}{\gamma_2(\mathcal{G})}\right)\left(\gamma_2(\mathcal{F})\mathsf{d}_{\psi_2}(\mathcal{G}) + \gamma_2(\mathcal{G})\mathsf{d}_{\psi_2}(\mathcal{F})\right). \tag{B5}$$

Straightforward calculations then show that the first of the two terms, (B4), is bounded above by

$$\frac{t}{N}\frac{\mathsf{d}_{\psi_2}^2(\mathcal{F})\gamma_2(\mathcal{G})}{\gamma_2(\mathcal{F})} \vee \frac{t}{N}\frac{\mathsf{d}_{\psi_2}^2(\mathcal{G})\gamma_2(\mathcal{F})}{\gamma_2(\mathcal{G})} \vee \frac{\gamma_2(\mathcal{F})\gamma_2(\mathcal{G})}{N},$$

and (B5) is similarly bounded above by

$$\sqrt{\frac{t}{N}}\frac{\mathsf{d}_{\psi_2}^2(\mathcal{F})\gamma_2(\mathcal{G})}{\gamma_2(\mathcal{F})} \vee \sqrt{\frac{t}{N}}\frac{\mathsf{d}_{\psi_2}^2(\mathcal{G})\gamma_2(\mathcal{F})}{\gamma_2(\mathcal{G})} \vee \frac{\mathsf{d}_{\psi_2}(\mathcal{G})\gamma_2(\mathcal{F})}{\sqrt{N}} \vee \frac{\mathsf{d}_{\psi_2}(\mathcal{F})\gamma_2(\mathcal{G})}{\sqrt{N}}.$$

Note then that since $0 \in \mathcal{F}$,

$$\mathsf{d}_{\psi_2}(\mathcal{F}) = \sup_{f\in\mathcal{F}} \|f\|_{\psi_2} \le \sup_{f_1,f_2\in\mathcal{F}} \|f_1 - f_2\|_{\psi_2} = \mathrm{diam}_{\psi_2}(\mathcal{F}) \le \gamma_2(\mathcal{F}),$$

where the final equality holds since $\gamma_2(\mathcal{F}) = \inf \sup_{f\in\mathcal{F}} \sum_{n=0}^{\infty} 2^{n/2}\mathrm{diam}_{\psi_2}(\Delta_n(f))$, and for $n = 0$, $\Delta_0 = \mathcal{F}$. Similarly, $\mathsf{d}_{\psi_2}(\mathcal{G}) \le \gamma_2(\mathcal{G})$, and so $\frac{\mathsf{d}_{\psi_2}^2(\mathcal{F})\gamma_2(\mathcal{G})}{\gamma_2(\mathcal{F})} \le d_{\psi_2}(\mathcal{F})\gamma_2(\mathcal{G})$ and $\frac{\mathsf{d}_{\psi_2}^2(\mathcal{G})\gamma_2(\mathcal{F})}{\gamma_2(\mathcal{G})} \le d_{\psi_2}(\mathcal{G})\gamma_2(\mathcal{F})$ which along with the fact that $t \ge 1$ completes the proof. □

THEOREM B.11. (Cross-Covariance Estimation—Max-Norm Bound) Let $X_1, \dots, X_N$ be d-dimensional i.i.d. sub-Gaussian random vectors with $\mathbb{E}[X_1] = \mu^X$ and $\mathrm{var}[X_1] = \Sigma^X$. Let $Y_1, \dots, Y_N$ be k-dimensional i.i.d. sub-Gaussian random vectors with $\mathbb{E}[Y_1] = \mu^Y$ and $\mathrm{var}[Y_1] = \Sigma^Y$. Define $\Sigma^{XY} = \mathbb{E}\left[(X - \mu^X)(Y - \mu^Y)^\top\right]$ and consider the cross-covariance estimator

$$\widehat{\Sigma}^{XY} = \frac{1}{N-1}\sum_{n=1}^{N}(X_n - \bar{X})(Y_n - \bar{Y})^\top.$$

Then there exist positive universal constants c_1, c_2 such that, for all $t \ge 1$, it holds with probability at least $1 - c_1 e^{-c_2 t}$ that

$$\|\widehat{\Sigma}^{XY} - \Sigma^{XY}\|_{\max} \lesssim (\Sigma^X_{(1)} \vee \Sigma^Y_{(1)})\left(\left(\frac{t}{N} \vee \sqrt{\frac{t}{N}}\right)\left(\sqrt{r_\infty(\Sigma^X)} \vee \sqrt{r_\infty(\Sigma^Y)}\right) \vee \sqrt{\frac{r_\infty(\Sigma^X)}{N}}\sqrt{\frac{r_\infty(\Sigma^Y)}{N}}\right).$$

Proof. Assume first that $\mu^X = \mu^Y = 0$. Let $Z_1, \ldots, Z_N$ be d-dimensional i.i.d. sub-Gaussian random vectors with zero mean and $\mathrm{var}[Z_1] = \Sigma^X$, and similarly let $V_1, \ldots, V_N$ be k-dimensional i.i.d. sub-Gaussian random vectors with zero mean and $\mathrm{var}[V_1] = \Sigma^Y$. Further, let $W_n \equiv [Z_n^\top, V_n^\top]^\top$ for $n = 1, \ldots, N$. We denote the distribution of W_1 by $\mathbb{P}$ and note that $\|\cdot\|_{\psi_2}$ and $\|\cdot\|_{L_2}$ are defined implicitly with respect to $\mathbb{P}$ throughout this proof. Define $\widehat{\Sigma}^0 = N^{-1} \sum_{n=1}^N Z_n V_n^\top$. Define the dilation operator: $\mathcal{H} : \mathbb{R}^{d\times k} \to \mathbb{R}^{(d+k)\times(d+k)}$ by

$$\mathcal{H}(A) = \begin{bmatrix} O & A \\ A^\top & O \end{bmatrix},$$

see, for example, [99, Section 2.1.16], and note that $\|A\|_{\max} = \|\mathcal{H}(A)\|_{\max}$. Let $\mathcal{B}^m$ be the space of standard basis vectors in m dimensions, i.e. any $b \in \mathcal{B}^m$ is an m-dimensional vector with 1 in a single coordinate and 0 otherwise. Then, for $e_i, e_j \in \mathcal{B}^{d+k}$, we have

$$\begin{aligned} \|\widehat{\Sigma}^0 - \Sigma^{XY}\|_{\max} = \|\mathcal{H}(\widehat{\Sigma}^0) - \mathcal{H}(\Sigma^{XY})\|_{\max} &= \max_{1\le i,j\le d+k} \left\langle (\mathcal{H}(\widehat{\Sigma}^0) - \mathcal{H}(\Sigma^{XY}))e_i, e_j \right\rangle \\ &\le 2 \sup_{u\in\mathcal{U}} \left| \left\langle (\mathcal{H}(\widehat{\Sigma}^0) - \mathcal{H}(\Sigma^{XY}))u, u \right\rangle \right|, \end{aligned}$$

where

$$\mathcal{U} \equiv \left\{ u \in \mathbb{R}^{d+k} : u = \pm\frac{1}{2}(e_i \pm e_j) \text{ and} e_i, e_j \in \mathcal{B}^{d+k} \right\}.$$

Writing $u = [u_1^\top, u_2^\top]^\top$ where $u_1 \in \mathbb{R}^d$ and $u_2 \in \mathbb{R}^k$, we have

$$\left\langle \mathcal{H}(\widehat{\Sigma}^0)u, u \right\rangle = \frac{2}{N} \sum_{n=1}^N \langle u_1, Z_n \rangle \langle u_2, V_n \rangle = \frac{2}{N} \sum_{n=1}^N f_u(W_n),$$

where $f_u(W_n) \equiv \langle \mathscr{A}_1 W_n, u_1 \rangle \langle \mathscr{A}_2 W_n, u_2 \rangle$ and $\mathscr{A}_1 \equiv [I_d, O_{d\times k}] \in \mathbb{R}^{d\times(d+k)}$ and $\mathscr{A}_2 \equiv [O_{k\times d}, I_k] \in \mathbb{R}^{k\times(d+k)}$ are the relevant selection matrices so that $\mathscr{A}_1 W_n = Z_n$ and $\mathscr{A}_2 W_n = V_n$. We define the class of functions

$$\mathcal{F}_\mathcal{U} \equiv \left\{ f_u(\cdot) = \langle \mathscr{A}_1 \cdot, u_1 \rangle \langle \mathscr{A}_2 \cdot, u_2 \rangle : u = [u_1^\top, u_2^\top]^\top \in \mathcal{U} \right\}.$$

It is clear then that $\mathcal{F}_{\mathcal{U}} \subset \mathcal{F}_1 \cdot \mathcal{F}_2$, where

$$\mathcal{U}_1 \equiv \left\{ u_1 \in \mathbb{R}^d : u_1 = \pm\frac{1}{2}(e_i \pm e_j) \text{ and} e_i, e_j \in \mathcal{B}_d \right\}, \quad \mathcal{F}_1 \equiv \{f(\cdot) = \langle \mathscr{A}_1 \cdot, u_1 \rangle : u_1 \in \mathcal{U}_1\},$$
$$\mathcal{U}_2 \equiv \left\{ u_2 \in \mathbb{R}^k : u_2 = \pm\frac{1}{2}(e_i \pm e_j) \text{ and} e_i, e_j \in \mathcal{B}_k \right\}, \quad \mathcal{F}_2 \equiv \{f(\cdot) = \langle \mathscr{A}_2 \cdot, u_2 \rangle : u_2 \in \mathcal{U}_2\},$$

and

$$\mathcal{F}_1 \cdot \mathcal{F}_2 \equiv \big\{ f(\cdot) = f_1(\cdot) f_2(\cdot) : f_1 \in \mathcal{F}_1, f_2 \in \mathcal{F}_2 \big\}.$$

We can then apply the product empirical process concentration bound of Theorem B.10, which implies that, with probability $1 - c_1 e^{-c_2 t}$,

$$\sup_{u \in \mathcal{U}} \left| \left\langle (\mathcal{H}(\widehat{\Sigma}^0) - \mathcal{H}(\Sigma^{XY}))u, u \right\rangle \right| = \sup_{f_u \in \mathcal{F}_{\mathcal{U}}} \left| \frac{1}{N} \sum_{n=1}^{N} f_u(W_n) - \mathbb{E}\big[f_u(W_n)\big] \right|$$
$$\lesssim \left(\frac{t}{N} \vee \sqrt{\frac{t}{N}} \right) \big(\mathsf{d}_{\psi_2}(\mathcal{F}_1)\gamma_2(\mathcal{F}_2) \vee \mathsf{d}_{\psi_2}(\mathcal{F}_2)\gamma_2(\mathcal{F}_1) \big) \vee \frac{\gamma_2(\mathcal{F}_1)\gamma_2(\mathcal{F}_2)}{N}, \quad \text{(B6)}$$

where we use the notational shorthand $\gamma_2(\mathcal{F}_1) = \gamma_2(\mathcal{F}_1, \psi_2)$ and $\mathsf{d}_{\psi_2}(\mathcal{F}_1) = \sup_{f \in \mathcal{F}_1} \|f\|_{\psi_2}$, and similarly for $\mathcal{F}_2$. Following a similar approach to the one taken in the proof of Theorem B.7, it follows by the equivalence of ψ_2 and L_2 norms for linear functionals that

$$d_{\psi_2}(\mathcal{F}_1) = \sup_{f_1 \in \mathcal{F}_1} \|f_1\|_{\psi_2} \lesssim \sup_{f_1 \in \mathcal{F}_1} \|f_1\|_{L_2} = \max_{u_1 \in \mathcal{U}_1} \sqrt{\langle u_1, \Sigma^X u_1 \rangle} \le \sqrt{\Sigma^X_{(1)}},$$

and similarly that $d_{\psi_2}(\mathcal{F}_2) \le \sqrt{\Sigma^Y_{(1)}}$. Further,

$$\gamma_2(\mathcal{F}_1) = \gamma_2(\mathcal{F}_1, \psi_2) \lesssim \gamma_2(\mathcal{F}_1, L_2) = \gamma_2(\mathcal{U}_1, \mathsf{d}_X),$$

where

$$\mathsf{d}_X(u, v) = \sqrt{\mathbb{E}\left[(\langle g_X, u \rangle - \langle g_X, v \rangle)^2 \right]}, \qquad g_X \sim \mathcal{N}(0, \Sigma^X).$$

From Theorem B.3 and Lemma B.6,

$$\gamma_2(\mathcal{U}_1, \mathsf{d}_X) \lesssim \mathbb{E}\left[\sup_{u_1 \in \mathcal{U}_1} \langle g_X, u_1 \rangle\right] = \mathbb{E}\left[\max_{i,j \le d} \left\langle g_X, \pm\frac{1}{2}(e_i \pm e_j)\right\rangle\right]$$
$$\le \mathbb{E}\left[\max_{i \le d} \langle g_X, e_i \rangle\right] \lesssim \max_{i \le d} \sqrt{\Sigma^X_{(i)} \log(i+1)}.$$

Similarly, $\gamma_2(\mathcal{F}_2) \lesssim \max_{j \le k} \sqrt{\Sigma^Y_{(j)} \log(j+1)}$. In summary, we have that

$$\|\widehat{\Sigma}^0 - \Sigma^{XY}\|_{\max} \lesssim \left(\frac{t}{N} \vee \sqrt{\frac{t}{N}}\right)\left(\sqrt{\Sigma^X_{(1)}} \max_{j \le k} \sqrt{\Sigma^Y_{(j)} \log(j+1)} \vee \sqrt{\Sigma^Y_{(1)}} \max_{i \le d} \sqrt{\Sigma^X_{(i)} \log(i+1)}\right)$$
$$\vee \frac{\max_{i \le d} \sqrt{\Sigma^X_{(i)} \log(i+1)} \max_{j \le k} \sqrt{\Sigma^Y_{(j)} \log(j+1)}}{N}$$
$$\lesssim \left(\frac{t}{N} \vee \sqrt{\frac{t}{N}}\right)\left(\Sigma^X_{(1)} \sqrt{r_\infty(\Sigma^X)} \vee \Sigma^Y_{(1)} \sqrt{r_\infty(\Sigma^Y)}\right) \vee \sqrt{\frac{\Sigma^X_{(1)} r_\infty(\Sigma^X)}{N}} \sqrt{\frac{\Sigma^Y_{(1)} r_\infty(\Sigma^Y)}{N}}$$
$$\lesssim (\Sigma^X_{(1)} \vee \Sigma^Y_{(1)}) \left(\left(\frac{t}{N} \vee \sqrt{\frac{t}{N}}\right)\left(\sqrt{r_\infty(\Sigma^X)} \vee \sqrt{r_\infty(\Sigma^Y)}\right) \vee \sqrt{\frac{r_\infty(\Sigma^X)}{N}} \sqrt{\frac{r_\infty(\Sigma^Y)}{N}}\right).$$

In the un-centered case, take $X_n = Z_n + \mu^X$ and $Y_n = V_n + \mu^Y$ for $n = 1, \ldots, N$, then $\widehat{\Sigma}^{XY} = \widehat{\Sigma}^0 - \bar{X}\bar{Y}^\top$, and so

$$\|\widehat{\Sigma}^{XY} - \Sigma^{XY}\|_{\max} \le \|\widehat{\Sigma}^0 - \Sigma^{XY}\|_{\max} + \|\bar{X}\bar{Y}^\top\|_{\max}.$$

The first term is controlled by appealing to the result in the centered case. For the second term, we note that from Lemma B.6,

$$\|\bar{X}\bar{Y}^\top\|_{\max} \le \|\bar{X}\|_{\max} \|\bar{Y}\|_{\max} \lesssim \frac{1}{N} \max_{i \le d} \sqrt{\Sigma^X_{(i)} \log(i+1)} \max_{j \le k} \sqrt{\Sigma^Y_{(j)} \log(j+1)}$$
$$\le \sqrt{\frac{\Sigma^X_{(1)} r_\infty(\Sigma^X)}{N}} \sqrt{\frac{\Sigma^Y_{(1)} r_\infty(\Sigma^Y)}{N}}.$$

□

THEOREM B.12. (Cross-Covariance Estimation with Localized Sample Cross-Covariance—Operator-Norm bound) Let $X_1, \ldots, X_N$ be d-dimensional i.i.d. sub-Gaussian random vectors with $\mathbb{E}[X_1] = \mu^X$ and $\mathrm{var}[X_1] = \Sigma^X$. Let $Y_1, \ldots, Y_N$ be k-dimensional i.i.d. sub-Gaussian random vectors with $\mathbb{E}[Y_1] = \mu^Y$

and $\text{var}[Y_1] = \Sigma^Y$. Define $\Sigma^{XY} = \mathbb{E}\big[(X - \mu^X)(Y - \mu^Y)^\top\big]$ and consider the estimator

$$\widehat{\Sigma}^{XY} = \frac{1}{N-1}\sum_{n=1}^{N}(X_n - \bar{X})(Y_n - \bar{Y})^\top.$$

Assume that $\Sigma^{XY} \in \mathcal{U}_{d,k}(q_1, R_{q_1})$ and $\Sigma^{YX} \in \mathcal{U}_{k,d}(q_2, R_{q_2})$ where $q_1, q_2 \in [0, 1)$ and R_{q_1}, R_{q_2} are positive constants. For any $t \geq 1$, set

$$\rho_N \asymp (\Sigma^X_{(1)} \vee \Sigma^Y_{(1)}) \left(\left(\frac{t}{N} \vee \sqrt{\frac{t}{N}}\right)\left(\sqrt{r_\infty(\Sigma^X)} \vee \sqrt{r_\infty(\Sigma^Y)}\right) \vee \sqrt{\frac{r_\infty(\Sigma^X)}{N}}\sqrt{\frac{r_\infty(\Sigma^Y)}{N}}\right),$$

and let $\widehat{\Sigma}^{XY}_{\rho_N}$ be the localized sample cross-covariance estimator. There exist positive universal constants c_1, c_2 such that, with probability at least $1 - c_1 e^{-c_2 t}$,

$$\|\widehat{\Sigma}^{XY}_{\rho_N} - \Sigma^{XY}\| \lesssim R_{q_1}\rho_N^{1-q_1} \vee R_{q_2}\rho_N^{1-q_2}.$$

Proof. (Proof of Theorem B.12) Let E denote the event on which $\|\widehat{\Sigma}^{XY} - \Sigma^{XY}\|_{\max} = \|\widehat{\Sigma}^{YX} - \Sigma^{YX}\|_{\max} \lesssim \rho_N$. From Theorem B.11, E holds with probability at least $1 - c_1 e^{-c_2 t}$. Conditional on E, and following an analysis identical to the one in the proof of Theorem B.9 with $\widehat{\Sigma}^{XY}(\widehat{\Sigma}^{YX})$ and $\Sigma^{XY}(\Sigma^{YX})$ in place of $\widehat{\Sigma}^X$ and Σ^X, respectively, it follows that

$$\|\widehat{\Sigma}^{XY}_{\rho_N} - \Sigma^{XY}\|_\infty \lesssim R_{q_1}\rho_N^{1-q_1},$$

and

$$\|\widehat{\Sigma}^{YX}_{\rho_N} - \Sigma^{YX}\|_\infty \lesssim R_{q_2}\rho_N^{1-q_2}.$$

The result then follows by noting that

$$\begin{aligned}\|\widehat{\Sigma}^{XY}_{\rho_N} - \Sigma^{XY}\| &= \|\mathcal{H}(\widehat{\Sigma}^{XY}_{\rho_N} - \Sigma^{XY})\| \leq \|\mathcal{H}(\widehat{\Sigma}^{XY}_{\rho_N} - \Sigma^{XY})\|_\infty \\ &= \|\widehat{\Sigma}^{XY}_{\rho_N} - \Sigma^{XY}\|_\infty \vee \|\widehat{\Sigma}^{YX}_{\rho_N} - \Sigma^{YX}\|_\infty \lesssim R_{q_1}\rho_N^{1-q_1} \vee R_{q_2}\rho_N^{1-q_2},\end{aligned}$$

where $\mathcal{H}$ is the dilation operator defined in the proof of Theorem B.11. □

Proof. (Proof of Theorem 3.3) The proof follows immediately from Theorem B.12: since $u_1, \ldots, u_N$ are i.i.d. Gaussian, they are sub-Gaussian. Moreover, since $\mathcal{G}$ is Lipschitz, by [102, Theorem 5.2.2],

$\|\mathcal{G}(u_1) - \mathbb{E}[\mathcal{G}(u_1)]\|_{\psi_2} \leq \|\mathcal{G}\|_{\mathrm{Lip}}\|C\|^{1/2} < \infty$, and so $\mathcal{G}(u_1), \ldots, \mathcal{G}(u_N)$ are i.i.d. sub-Gaussian random vectors. □

LEMMA B.13. (Stein's Lemma [90]) Let $u \sim \mathcal{N}(m, C)$ be a d-dimensional Gaussian vector. Let $h : \mathbb{R}^d \to \mathbb{R}$ such that $\partial_j h \equiv \partial h(u)/\partial u_j$ exists almost everywhere and $\mathbb{E}[|\partial_j h(u)|] < \infty, j = 1, \ldots, d$. Then

$$\mathrm{Cov}\big(u_j, h(u)\big) = \sum_{l=1}^{d} C_{jl}\mathbb{E}[\partial_l h(u)].$$

LEMMA B.14. (Soft-Sparsity of Cross-Covariance—Nonlinear Forward Map) Let u be a d-dimensional Gaussian random vector with $\mathbb{E}[u] = m$ and $\mathrm{var}[u] = C \in \mathscr{U}_d(q, c)$. Consider the function $\mathcal{G} : \mathbb{R}^d \to \mathbb{R}^k$ with coordinate functions $\mathcal{G}_1, \ldots, \mathcal{G}_k$. Assume that for each $i = 1, \ldots, d$ and $j = 1, \ldots, k$, $\mathcal{G}_j : \mathbb{R}^d \to \mathbb{R}$ for $j = 1, \ldots, k$, such that $\partial_i \mathcal{G}_j \equiv \partial \mathcal{G}_j(u)/\partial u_i$ exists almost everywhere, and $\mathbb{E}[|\partial_i \mathcal{G}_j|] < \infty$. Let $D\mathcal{G} \in \mathbb{R}^{k\times d}$ denote the Jacobian of $\mathcal{G}$, and assume that $\mathbb{E}\big[(D\mathcal{G})^\top\big] \in \mathscr{U}_{d,k}(q, a)$ for some $q \in [0, 1)$ and $a > 0$. Then,

$$C^{up} \in \mathscr{U}_{d,k}\big(q, ac\|\mathbb{E}[D\mathcal{G}]\|_{\max}^{1-q}\|C\|_{\max}^{1-q}\big).$$

Proof. By Stein's Lemma (Lemma B.13), the ith row sum of C^{up} is given by

$$\begin{aligned}
\sum_{j=1}^{k} C_{ij}^{up} = \sum_{j=1}^{k}\sum_{l=1}^{d} C_{il}\mathbb{E}\big[\partial_l \mathcal{G}_j(u)\big] &= \sum_{l=1}^{d} C_{il}\sum_{j=1}^{k}\mathbb{E}\big[\partial_l \mathcal{G}_j(u)\big] \\
&= \|\mathbb{E}[D\mathcal{G}]\|_{\max}\sum_{l=1}^{d} C_{il}\sum_{j=1}^{k}\frac{\mathbb{E}[\partial_l \mathcal{G}_j(u)]}{\|\mathbb{E}[D\mathcal{G}]\|_{\max}} \\
&\leq \|\mathbb{E}[D\mathcal{G}]\|_{\max}^{1-q}\sum_{l=1}^{d} C_{il}\sum_{j=1}^{k}\mathbb{E}[\partial_l \mathcal{G}_j(u)]^q \\
&\leq a\|\mathbb{E}[D\mathcal{G}]\|_{\max}^{1-q}\sum_{l=1}^{d} C_{il} \\
&\leq ac\|\mathbb{E}[D\mathcal{G}]\|_{\max}^{1-q}\|C\|_{\max}^{1-q},
\end{aligned}$$

where the first inequality holds since $q \in [0, 1)$ and $\mathbb{E}[\partial_l \mathcal{G}_j(u)] \leq \|\mathbb{E}[D\mathcal{G}]\|_{\max}$. □

LEMMA B.15. (Product of Two Soft-Sparse Matrices) Fix $q \in [0, 1)$ and let $S \in \mathscr{U}_d(q, s)$ and assume $S^\top = S$. Let $B \in \mathscr{U}_{k,d}(q, b)$. Then $BS \in \mathscr{U}_{k,d}(q, bs\|B\|_{\max}^{1-q}\|S\|_{\max}^{1-q})$.

Proof. The (i,j)th element of BS is given by $[BS]_{ij} = \sum_{l=1}^{d} B_{il}S_{lj}$, and so the sum of the ith row of BS satisfies

$$\begin{aligned}\sum_{j=1}^{d}[BS]_{ij} = \sum_{j=1}^{d}\sum_{l=1}^{d} B_{il}S_{lj} = \sum_{l=1}^{d} B_{il}\sum_{j=1}^{d} S_{lj} &= \|B\|_{\max}\|S\|_{\max}\sum_{l=1}^{d}\frac{B_{il}}{\|B\|_{\max}}\sum_{j=1}^{d}\frac{S_{lj}}{\|S\|_{\max}} \\ &\le \|B\|_{\max}\|S\|_{\max}\sum_{l=1}^{d}\left(\frac{B_{il}}{\|B\|_{\max}}\right)^{q}\sum_{j=1}^{d}\left(\frac{S_{lj}}{\|S\|_{\max}}\right)^{q} \\ &\le \|B\|_{\max}^{1-q}\|S\|_{\max}^{1-q} bs,\end{aligned}$$

where the first inequality holds since $q \in [0,1)$, and the second follows by the symmetry of S. □

LEMMA B.16. (Product of Three Soft-Sparse Matrices) Fix $q \in [0,1)$ and let $S \in \mathscr{U}_d(q,s)$ with $S^\top = S$. Let $B \in \mathscr{U}_{k,d}(q,b_1)$ and $B^\top \in \mathscr{U}_{d,k}(q,b_2)$, that is, B is both row and column sparse. Then $BSB^\top \in \mathscr{U}_{k,d}(q, b_1b_2s\|B\|_{\max}^{2(1-q)}\|S\|_{\max}^{1-q})$.

Proof. The (i,j)-th element of $BSB^\top$ is given by

$$[BSB^\top]_{ij} = \sum_{m=1}^{d}[BS]_{im}B^\top_{mj} = \sum_{m=1}^{d}[BS]_{im}B_{jm} = \sum_{m=1}^{d}\left(\sum_{l=1}^{d} B_{il}S_{lm}\right)B_{jm}.$$

Therefore, the sum of the ith row of $BSB^\top$ satisfies

$$\begin{aligned}\sum_{j=1}^{k}[BSB^\top]_{ij} = \sum_{j=1}^{k}\sum_{m=1}^{d}\sum_{l=1}^{d} B_{il}S_{lm}B_{jm} &= \sum_{m=1}^{d}\sum_{l=1}^{d} B_{il}S_{lm}\sum_{j=1}^{k} B_{jm} \\ &\le \|B\|_{\max}^{1-q} b_2 \sum_{m=1}^{d}\sum_{l=1}^{d} B_{il}S_{lm} \le b_1b_2s\|B\|_{\max}^{2(1-q)}\|S\|_{\max}^{1-q},\end{aligned}$$

where the final inequality follows from Lemma B.15. □

LEMMA B.17. (Sample Covariance Deviation) Let $X_1,\dots,X_N$ be d-dimensional i.i.d. sub-Gaussian random vectors with $\mathbb{E}[X_1] = \mu^X$ and $\mathrm{var}[X_1] = \Sigma^X$. Let $\widehat{\Sigma}_N^X = (N-1)^{-1}\sum_{n=1}^{N}(X_n-\mu^X)(X_n-\mu^X)^\top$. Then

$$\begin{aligned}\widehat{\Sigma}_N^X - \widehat{\Sigma}_{N-1}^X \asymp \frac{1}{N}X_NX_N^\top - \frac{1}{N}\widehat{\Sigma}_{N-1}^0 - \frac{1}{N^2}X_NX_N^\top - \left(\left(\frac{N-1}{N}\right)^2 - 1\right)\bar{X}_{N-1}\bar{X}_{N-1}^\top \\ - \left(\frac{N-1}{N^2}\right)\left(X_N\bar{X}_{N-1}^\top + \bar{X}_{N-1}X_N^\top\right),\end{aligned}$$

where $\widehat{\Sigma}_N^0 = N^{-1}\sum_{n=1}^{N} X_nX_n^\top$.

Proof. We work with the biased sample covariance estimator $\frac{1}{N}\sum_{n=1}^{n}(X_n - \bar{X}_N)(X_n - \bar{X}_N)^\top$, which is equivalent to the unbiased covariance estimator up to constants. Note then that

$$\widehat{\Sigma}_N^X \asymp \frac{1}{N}\sum_{n=1}^{N}(X_n - \bar{X}_N)(X_n - \bar{X}_N)^\top = \frac{1}{N}\sum_{n=1}^{N} X_n X_n^\top - \bar{X}_N \bar{X}_N^\top = \widehat{\Sigma}_N^0 - \bar{X}_N \bar{X}_N^\top.$$

We now seek to control the difference $\widehat{\Sigma}_N^X - \widehat{\Sigma}_{N-1}^X$. To that end, note that

$$\begin{aligned}\bar{X}_N \bar{X}_N^\top &= \left(\frac{1}{N}X_N + \frac{N-1}{N}\bar{X}_{N-1}\right)\left(\frac{1}{N}X_N + \frac{N-1}{N}\bar{X}_{N-1}\right)^\top \\ &= \frac{1}{N^2}X_N X_N^\top + \left(\frac{N-1}{N}\right)^2 \bar{X}_{N-1}\bar{X}_{N-1}^\top + \left(\frac{N-1}{N^2}\right)\left(X_N \bar{X}_{N-1}^\top + \bar{X}_{N-1}X_N^\top\right),\end{aligned}$$

and so

$$\begin{aligned}\bar{X}_N \bar{X}_N^\top - \bar{X}_{N-1}\bar{X}_{N-1}^\top = \frac{1}{N^2}X_N X_N^\top + \left(\left(\frac{N-1}{N}\right)^2 - 1\right)\bar{X}_{N-1}\bar{X}_{N-1}^\top \\ + \left(\frac{N-1}{N^2}\right)\left(X_N \bar{X}_{N-1}^\top + \bar{X}_{N-1}X_N^\top\right). \qquad \text{(B7)}\end{aligned}$$

Therefore,

$$\begin{aligned}\widehat{\Sigma}_N^X - \widehat{\Sigma}_{N-1}^X &\asymp \left(\frac{1}{N}\sum_{n=1}^{N} X_n X_n^\top - \bar{X}_N \bar{X}_N^\top\right) - \left(\frac{1}{N-1}\sum_{n=1}^{N-1} X_n X_n^\top - \bar{X}_{N-1}\bar{X}_{N-1}^\top\right) \\ &= \frac{1}{N}X_N X_N^\top + \left(\left(\frac{1}{N} - \frac{1}{N-1}\right)\sum_{n=1}^{N-1} X_n X_n^\top\right) + \left(\bar{X}_{N-1}\bar{X}_{N-1}^\top - \bar{X}_N \bar{X}_N^\top\right) \\ &= \frac{1}{N}X_N X_N^\top - \frac{1}{N}\widehat{\Sigma}_{N-1}^0 - \frac{1}{N^2}X_N X_N^\top - \left(\left(\frac{N-1}{N}\right)^2 - 1\right)\bar{X}_{N-1}\bar{X}_{N-1}^\top \\ &\quad - \left(\frac{N-1}{N^2}\right)\left(X_N \bar{X}_{N-1}^\top + \bar{X}_{N-1}X_N^\top\right), \qquad \text{(B8)}\end{aligned}$$

where the last equality follows by (B7). □

LEMMA B.18. (Sample Cross-Covariance Deviation) Let $X_1, \ldots, X_N$ be d-dimensional i.i.d. sub-Gaussian random vectors with $\mathbb{E}[X_1] = \mu^X$ and $\mathrm{var}[X_1] = \Sigma^X$. Let $Y_1, \ldots, Y_N$ be k-dimensional i.i.d.

sub-Gaussian random vectors with $\mathbb{E}[Y_1] = \mu^Y$ and $\mathrm{var}[Y_1] = \Sigma^Y$. Let $\widehat{\Sigma}_N^{XY} = (N-1)^{-1}\sum_{n=1}^N (X_n - \mu^X)(Y_n - \mu^Y)^\top$. Then

$$\widehat{\Sigma}_N^{XY} - \widehat{\Sigma}_{N-1}^{XY} \asymp \frac{1}{N} X_N Y_N^\top - \frac{1}{N}\widehat{\Sigma}_{N-1}^{0,XY} - \frac{1}{N^2} X_N Y_N^\top - \left(\left(\frac{N-1}{N}\right)^2 - 1\right)\bar{X}_{N-1}\bar{Y}_{N-1}^\top - \left(\frac{N-1}{N^2}\right)\left(X_N \bar{Y}_{N-1}^\top + \bar{X}_{N-1} Y_N^\top\right),$$

where $\widehat{\Sigma}_{N-1}^{0,XY} = N^{-1}\sum_{n=1}^N X_n Y_n$.

Proof. The result follows using the same approach utilized in the proof of Lemma B.17 and is omitted for brevity. □

LEMMA B.19. (Covariance Estimation with Known Particle—Operator-Norm Bound) Consider the set-up in Lemma B.17 and assume additionally that X_n is known for some $n \in \{1, \dots, N\}$. Then with probability at least $1 - ce^{-t}$

$$\|\widehat{\Sigma}_N^X - \Sigma^X\| \lesssim \frac{c(\|X_n\|_2, \|\mu^X\|_2)}{N} + \|\Sigma^X\|\left(\sqrt{\frac{r_2(\Sigma^X)}{N}} \vee \frac{r_2(\Sigma^X)}{N} \vee \sqrt{\frac{t}{N}} \vee \frac{t}{N}\right).$$

Proof. By symmetry, we can assume without loss of generality that $n = N$. Let E_1 denote the event on which $\|\bar{X}_{N-1} - \mu^X\|_2 \lesssim \sqrt{\|\Sigma^X\|\frac{r_2(\Sigma^X)\vee t}{N-1}} \asymp \sqrt{\|\Sigma^X\|\frac{r_2(\Sigma^X)\vee t}{N}}$. Then from Theorem A.1, $\mathbb{P}(E_1) \geq 1 - e^{-t}$ and on E_1 it holds that

$$\begin{aligned}
\|\widehat{\Sigma}_N^X - \widehat{\Sigma}_{N-1}^X\| &\lesssim \frac{1}{N}\|X_N X_N^\top\| + \frac{1}{N}\|\widehat{\Sigma}_{N-1}^0\| + \left(\left(\frac{N-1}{N}\right)^2 - 1\right)\|\bar{X}_{N-1}\bar{X}_{N-1}^\top\| \\
&\quad + \frac{N-1}{N^2}\left(\|X_N \bar{X}_{N-1}^\top\| + \|\bar{X}_{N-1} X_N^\top\|\right) \\
&\lesssim \frac{1}{N}\|X_N\|_2^2 + \frac{1}{N}\|\widehat{\Sigma}_{N-1}^0\| + \frac{1}{N}\|\bar{X}_{N-1}\|_2^2 + \frac{1}{N}\|X_N\|_2\|\bar{X}_{N-1}\|_2 \\
&\leq \frac{1}{N}\|X_N\|_2^2 + \frac{1}{N}\|\widehat{\Sigma}_{N-1}^0 - \Sigma^X\| + \frac{1}{N}\|\Sigma^X\| + \frac{1}{N}\|\bar{X}_{N-1} - \mu^X\|_2^2 + \frac{1}{N}\|\mu^X\|_2^2 \\
&\quad + \frac{1}{N}\|X_N\|_2\|\bar{X}_{N-1} - \mu^X\|_2 + \frac{1}{N}\|X_N\|_2\|\mu^X\|_2 \\
&\lesssim \frac{1}{N}\|X_N\|_2^2 + \frac{1}{N}\|\widehat{\Sigma}_{N-1}^0 - \Sigma^X\| + \frac{1}{N}\|\Sigma^X\| + \|\Sigma^X\|\frac{r_2(\Sigma^X)\vee t}{N^2} + \frac{1}{N}\|\mu^X\|_2^2 \\
&\quad + \|X_N\|_2\sqrt{\|\Sigma^X\|}\frac{\sqrt{r_2(\Sigma^X)\vee t}}{N^{3/2}} + \frac{1}{N}\|X_N\|_2\|\mu^X\|_2, \qquad \text{(B9)}
\end{aligned}$$

where the first line follows from Lemma B.17. Let E_2 denote the event on which

$$\|\widehat{\Sigma}^X_{N-1} - \Sigma^X\| \lesssim \|\Sigma^X\| \left(\sqrt{\frac{r_2(\Sigma^X)}{N-1}} \vee \frac{r_2(\Sigma^X)}{N-1} \vee \sqrt{\frac{t}{N-1}} \vee \frac{t}{N-1} \right)$$
$$\asymp \|\Sigma^X\| \left(\sqrt{\frac{r_2(\Sigma^X)}{N}} \vee \frac{r_2(\Sigma^X)}{N} \vee \sqrt{\frac{t}{N}} \vee \frac{t}{N} \right).$$

Then by Proposition 2.1, $\mathbb{P}(E_2) \geq 1 - ce^{-t}$. It holds on $E_1 \cap E_2$ that

$$\begin{aligned}
\|\widehat{\Sigma}^X_N - \Sigma^X\| &\leq \|\widehat{\Sigma}^X_N - \widehat{\Sigma}^X_{N-1}\| + \|\widehat{\Sigma}^X_{N-1} - \Sigma^X\| \\
&\lesssim \frac{1}{N}\|X_N\|_2^2 + \frac{1}{N}\|\widehat{\Sigma}^0_{N-1} - \Sigma^X\| + \frac{1}{N}\|\Sigma^X\| + \|\Sigma^X\|\frac{r_2(\Sigma^X) \vee t}{N^2} + \frac{1}{N}\|\mu^X\|_2^2 \\
&+ \|X_N\|_2\sqrt{\|\Sigma^X\|}\frac{\sqrt{r_2(\Sigma^X) \vee t}}{N^{3/2}} + \frac{1}{N}\|X_N\|_2\|\mu^X\|_2 \\
&+ \|\Sigma^X\| \left(\sqrt{\frac{r_2(\Sigma^X)}{N}} \vee \frac{r_2(\Sigma^X)}{N} \vee \sqrt{\frac{t}{N}} \vee \frac{t}{N} \right) \\
&\lesssim \frac{c(\|X_N\|_2, \|\mu^X\|_2)}{N} + \|\Sigma^X\| \left(\sqrt{\frac{r_2(\Sigma^X)}{N}} \vee \frac{r_2(\Sigma^X)}{N} \vee \sqrt{\frac{t}{N}} \vee \frac{t}{N} \right),
\end{aligned}$$

where the first inequality holds by (B9). The result follows by noting that $\mathbb{P}(E_1 \cap E_2) \geq 1 - ce^{-t}$. □

LEMMA B.20. (Covariance Estimation with Known Particle—Maximum-Norm Bound) Consider the set-up in Lemma B.17 and assume additionally that X_n is known for some $n \in \{1, \ldots, N\}$. Then with probability at least $1 - ce^{-t}$

$$\|\widehat{\Sigma}^X_N - \Sigma^X\|_{\max} \lesssim \frac{c(\|X_n\|_\infty, \|\mu^X\|_\infty)}{N} + \Sigma^X_{(1)} \left(\sqrt{\frac{r_\infty(\Sigma^X)}{N}} \vee \sqrt{\frac{t}{N}} \vee \frac{t}{N} \vee \frac{t r_\infty(\Sigma^X)}{N} \right).$$

Proof. As in the proof of Lemma B.19, we may assume that $n = N$. Let E_1 denote the event on which $\|\bar{X}_{N-1} - \mu^X\|_\infty \lesssim \sqrt{t\Sigma^X_{(1)}\frac{r_\infty(\Sigma^X)}{N-1}} \asymp \sqrt{t\Sigma^X_{(1)}\frac{r_\infty(\Sigma^X)}{N}}$. Then from Lemma B.6, $\mathbb{P}(E_1) \geq 1 - ce^{-ct}$ and on E_1, using similar calculations to those used to derive (B9), it holds that

$$\begin{aligned}
\|\widehat{\Sigma}^X_N - \widehat{\Sigma}^X_{N-1}\|_{\max} &\lesssim \frac{1}{N}\|X_N\|_\infty^2 + \frac{1}{N}\|\widehat{\Sigma}^0_{N-1} - \Sigma^X\|_{\max} + \frac{1}{N}\|\Sigma^X\|_{\max} + t\Sigma^X_{(1)}\frac{r_\infty(\Sigma^X)}{N^2} + \frac{1}{N}\|\mu^X\|_\infty^2 \\
&+ \|X_N\|_\infty\sqrt{t\Sigma^X_{(1)}}\frac{\sqrt{r_\infty(\Sigma^X)}}{N^{3/2}} + \frac{1}{N}\|X_N\|_\infty\|\mu^X\|_\infty. \qquad \text{(B10)}
\end{aligned}$$

Let E_2 be the event on which

$$\|\widehat{\Sigma}^X_{N-1} - \Sigma^X\|_{\max} \lesssim \Sigma^X_{(1)} \left(\sqrt{\frac{r_\infty(\Sigma^X)}{N}} \vee \sqrt{\frac{t}{N}} \vee \frac{t}{N} \vee \frac{tr_\infty(\Sigma^X)}{N} \right).$$

From Theorem B.7, $\mathbb{P}(E_2) \geq 1 - ce^{-t}$. Finally, note that the desired result holds on $E_1 \cap E_2$ and that $\mathbb{P}(E_1 \cap E_2) \geq 1 - ce^{-t}$, which completes the proof. □

LEMMA B.21. (Cross-Covariance Estimation with Known Particle—Maximum-Norm Bound) Consider the set-up in Lemma B.18 and assume additionally that (X_n, Y_n) is known for some $n \in \{1, \ldots, N\}$. Then with probability at least $1 - ce^{-t}$

$$\|\widehat{\Sigma}^{XY}_N - \Sigma^{XY}\|_{\max} \lesssim \frac{c(\|X_n\|_\infty, \|\mu^X\|_\infty, \|Y_n\|_\infty, \|\mu^Y\|_\infty)}{N} + (\Sigma^X_{(1)} \vee \Sigma^Y_{(1)}) \left(\left(\frac{t}{N} \vee \sqrt{\frac{t}{N}} \right) \left(\sqrt{r_\infty(\Sigma^X)} \vee \sqrt{r_\infty(\Sigma^Y)} \right) \vee \sqrt{\frac{r_\infty(\Sigma^X)}{N}} \sqrt{\frac{r_\infty(\Sigma^Y)}{N}} \right).$$

Proof. The result follows using the same approach utilized in the proof of Lemma B.17 and utilizing the statements of Lemma B.18 and Theorem B.11. We omit the details for brevity. □

LEMMA B.22. (Covariance Estimation with Localized Sample Covariance and with Known Particle—Operator-Norm Bound) Consider the set-up in Theorem B.9 and assume additionally that X_n is known for some $n \in \{1, \ldots, N\}$. For any $t \geq 1$, set

$$\rho_N \asymp \frac{c(\|X_n\|_\infty, \|\mu^X\|_\infty)}{N} + \Sigma^X_{(1)} \left(\sqrt{\frac{r_\infty(\Sigma^X)}{N}} \vee \sqrt{\frac{t}{N}} \vee \frac{t}{N} \vee \frac{tr_\infty(\Sigma^X)}{N} \right)$$

and let $\widehat{\Sigma}^X_{\rho_N}$ be the localized sample covariance estimator. There exists a constant $c > 0$ such that, with probability at least $1 - ce^{-t}$, it holds that

$$\|\widehat{\Sigma}^X_{\rho_N} - \Sigma^X\| \lesssim R_q \rho_N^{1-q}.$$

Proof. The proof follows in identical fashion to that of Theorem B.9, except that we now use the max-norm bound established in Lemma B.20 in place of Theorem B.7. □

LEMMA B.23. (Cross-Covariance Estimation with Localized Sample Covariance and with Known Particle—Operator-Norm Bound) Consider the set-up in Theorem B.12 and assume additionally that

(X_n, Y_n) is known for some $n \in \{1, \dots, N\}$. For any $t \geq 1$, set

$$\rho_N \asymp \frac{c(\|X_n\|_\infty, \|\mu^X\|_\infty, \|Y_n\|_\infty, \|\mu^Y\|_\infty)}{N} + (\Sigma^X_{(1)} \vee \Sigma^Y_{(1)}) \left(\left(\frac{t}{N} \vee \sqrt{\frac{t}{N}} \right) \left(\sqrt{r_\infty(\Sigma^X)} \vee \sqrt{r_\infty(\Sigma^Y)} \right) \vee \sqrt{\frac{r_\infty(\Sigma^X)}{N}} \sqrt{\frac{r_\infty(\Sigma^Y)}{N}} \right).$$

There exists positive universal constants c_1, c_2 such that, with probability at least $1 - c_1 e^{-c_2 t}$,

$$\|\widehat{\Sigma}^{XY}_{\rho_N} - \Sigma^{XY}\| \lesssim R_{q_1} \rho_N^{1-q_1} \vee R_{q_2} \rho_N^{1-q_2}.$$

Proof. The proof follows in identical fashion to that of Theorem B.9, except that we now use the max-norm bound established in Lemma B.21 in place of Theorem B.7. □

B.2 *Proof of main results in Section 3*

Proof of Theorem 3.5 First, we may write

$$\begin{aligned} \|\upsilon_n - \upsilon_n^*\|_2 &= \big\| \big(y - \mathcal{G}(u_n) - \eta_n\big) \big(\mathscr{P}(\widehat{C}^{up}, \widehat{C}^{pp}) - \mathscr{P}(C^{up}, C^{pp})\big) \big\|_2 \\ &\leq \|y - \mathcal{G}(u_n) - \eta_n\|_2 \|\mathscr{P}(\widehat{C}^{up}, \widehat{C}^{pp}) - \mathscr{P}(C^{up}, C^{pp})\|_2. \end{aligned} \tag{B11}$$

For the second term in (B11), it follows from Lemma A.7 that

$$\|\mathscr{P}(\widehat{C}^{up}, \widehat{C}^{pp}) - \mathscr{P}(C^{up}, C^{pp}))\|_2 \leq \|\Gamma^{-1}\| \|\widehat{C}^{up} - C^{up}\| + \|\Gamma^{-1}\|^2 \|C^{up}\| \|\widehat{C}^{pp} - C^{pp}\|.$$

In order to control the two deviation terms, we write $W_i \equiv [u_i^\top, \mathcal{G}^\top(u_i)]^\top$ for $1 \leq i \leq N$. Further, let

$$\widehat{C}^W = \frac{1}{N-1} \sum_{i=1}^N (W_i - \overline{W}_N)(W_i - \overline{W}_N)^\top, \qquad C^W = \begin{bmatrix} C & C^{up} \\ C^{pu} & C^{pp} \end{bmatrix}.$$

with $\overline{W}_N = [\widehat{m}^\top, \overline{\mathcal{G}}^\top]^\top$ and $\overline{\mathcal{G}}$ the sample mean of $\{\mathcal{G}(u_n)\}_{n=1}^N$. Since $u \sim \mathcal{N}(m, C)$ and $\mathcal{G}$ is Lipschitz, by Gaussian concentration [102, Theorem 5.2.2], it holds that $\|\mathcal{G}(u) - \mathbb{E}[\mathcal{G}(u)]\|_{\psi_2} \leq \|\mathcal{G}\|_{\mathrm{Lip}} \|C\|^{1/2}$ and

we can apply Lemma B.19. Letting E_1 be the event on which

$$\|\widehat{C}^{up} - C^{up}\| \vee \|\widehat{C}^{pp} - C^{pp}\| \leq \|\widehat{C}^{W} - C^{W}\| \\ \lesssim \frac{c(\|W_n\|, \|\mathbb{E}[W_n]\|)}{N} + \|C^W\|\left(\sqrt{\frac{r_2(C^W)}{N}} \vee \frac{r_2(C^W)}{N} \vee \sqrt{\frac{t}{N}} \vee \frac{t}{N}\right),$$

then Lemma B.19 ensures that $\mathbb{P}(E_1) \geq 1 - c_1 e^{-c_2 t}$. It follows that on the event E_1, we also have

$$\|\mathscr{P}(\widehat{C}^{up}, \widehat{C}^{pp}) - \mathscr{P}(C^{up}, C^{pp}))\|_2 \lesssim \|\Gamma^{-1}\|(1 \vee \|C^{up}\|)\|C^W\|\left(\sqrt{\frac{r_2(C^W)}{N}} \vee \frac{r_2(C^W)}{N} \vee \sqrt{\frac{t}{N}} \vee \frac{t}{N}\right).$$

The expression can be simplified by noting that since $C^W \succeq 0$, $\|C^W\| \leq \|C\| + \|C^{pp}\|$ and further since $\mathrm{Tr}(C^W) = \mathrm{Tr}(C) + \mathrm{Tr}(C^{pp})$

$$\|C^W\|\left(\sqrt{\frac{r_2(C^W)}{N}} \vee \frac{r_2(C^W)}{N} \vee \sqrt{\frac{t}{N}} \vee \frac{t}{N}\right) \\ \lesssim (\|C\| \vee \|C^{pp}\|)\left(\sqrt{\frac{\mathrm{Tr}(C) + \mathrm{Tr}(C^{pp})}{N(\|C\| \vee \|C^{pp}\|)}} \vee \frac{\mathrm{Tr}(C) + \mathrm{Tr}(C^{pp})}{N(\|C\| \vee \|C^{pp}\|)} \vee \sqrt{\frac{t}{N}} \vee \frac{t}{N}\right) \\ \lesssim (\|C\| \vee \|C^{pp}\|)\left(\sqrt{\frac{r_2(C)}{N}} \vee \frac{r_2(C)}{N} \vee \sqrt{\frac{r_2(C^{pp})}{N}} \vee \frac{r_2(C^{pp})}{N} \vee \sqrt{\frac{t}{N}} \vee \frac{t}{N}\right),$$

where the last inequality follows by similar reasoning to that used in the proof of Lemma A.3. □

Proof of Theorem 3.7 As in the proof of Theorem 3.5, we have that

$$\|\upsilon_n^\rho - \upsilon_n^*\|_2 \leq \|y - \mathcal{G}(u_n) - \eta_n\|_2 \|\mathscr{P}(\widehat{C}^{up}_{\rho_N}, \widehat{C}^{pp}_{\rho_N}) - \mathscr{P}(C^{up}, C^{pp})\|_2.$$

Further, from Lemma A.7,

$$\|\mathscr{P}(\widehat{C}^{up}_{\rho_N}, \widehat{C}^{pp}_{\rho_N}) - \mathscr{P}(C^{up}, C^{pp})\|_2 \leq (\|\Gamma^{-1}\| \vee \|\Gamma^{-1}\|^2)(1 \vee \|C^{up}\|)(\|\widehat{C}^{up}_{\rho_N} - C^{up}\| + \|\widehat{C}^{pp}_{\rho_N} - C^{pp}\|).$$

Let E_1 denote the event on which

$$\|\widehat{C}^{up}_{\rho_N} - C^{up}\| \lesssim R_{q_1} \rho_{N,1}^{1-q_1} \vee R_{q_2} \rho_{N,2}^{1-q_2}.$$

By Lemma B.23 , E_1 has probability at least $1 - c_1 e^{-c_2 t}$. Let E_2 be the event on which

$$\|\widehat{C}^{pp}_{\rho_N} - C^{pp}\| \lesssim R_{q_3} \rho_{N,3}^{1-q_3}.$$

By Lemma B.22, E_2 has probability at least $1 - c_1 e^{-c_2 t}$. Therefore, $E = E_1 \cap E_2$ has probability at least $1 - c_1 e^{-c_2 t}$, and on E it holds that

$$\|\mathscr{P}(\widehat{C}^{up}_{\rho_N}, \widehat{C}^{pp}_{\rho_N}) - \mathscr{P}(C^{up}, C^{pp})\|_2 \lesssim (\|\Gamma^{-1}\| \vee \|\Gamma^{-1}\|^2)(1 \vee \|C^{up}\|)(R_{q_1}\rho_{N,1}^{1-q_1} + R_{q_2}\rho_{N,2}^{1-q_2} + R_{q_3}\rho_{N,3}^{1-q_3}).$$

□

C. Proofs: Section 4

This appendix contains the proofs of the auxiliary results discussed in Section 4.

LEMMA C.1. (Kalman Gain Deviation with Localization) Let $u_1, \dots, u_N$ be d-dimensional i.i.d. sub-Gaussian random vectors with $\mathbb{E}[u_1] = m$ and $\mathbb{E}\big[(u_1 - m)(u_1 - m)^\top\big] = C$. Assume further that $C \in \mathscr{U}_d(q, R_q)$ for some $q \in [0, 1)$ and $R_q > 0$. For any $t \geq 1$, set

$$\rho_N \asymp C_{(1)} \left(\sqrt{\frac{r_\infty(C)}{N}} \vee \sqrt{\frac{t}{N}} \vee \frac{t}{N} \vee \frac{t r_\infty(C)}{N} \right)$$

and let $\widehat{C}_{\rho_N}$ be the localized sample covariance estimator. There exists a positive universal constant c such that, with probability at least $1 - ce^{-t}$,

$$\|\mathscr{K}(\widehat{C}_{\rho_N}) - \mathscr{K}(\widehat{C})\| \lesssim \|A\|\|\Gamma^{-1}\| R_q (1 + \|A\|^2 \|\Gamma^{-1}\|\|C\|) \rho_N^{1-q}.$$

Proof. By Lemma A.4 and Theorem 3.1, it follows immediately that

$$\begin{aligned} \|\mathscr{K}(\widehat{C}_{\rho_N}) - \mathscr{K}(\widehat{C})\| &\leq \|A\|\|\Gamma^{-1}\| \|\widehat{C}_{\rho_N} - C\| (1 + \|A\|^2 \|\Gamma^{-1}\|\|C\|) \\ &\lesssim \|A\|\|\Gamma^{-1}\| R_q \rho_N^{1-q} (1 + \|A\|^2 \|\Gamma^{-1}\|\|C\|). \end{aligned}$$

□

THEOREM C.2. (Square Root Ensemble Kalman Covariance Deviation with Localization) Consider the localized SR ensemble Kalman update given by (4.2), leading to an estimate $\widehat{\Sigma}$ of the posterior covariance Σ defined in (1.2). Assume that $C \in \mathscr{U}_d(q, R_q)$ for $q \in [0, 1)$ and $R_q > 0$. For any $t \geq 1$, set

$$\rho_N \asymp C_{(1)} \left(\sqrt{\frac{r_\infty(C)}{N}} \vee \sqrt{\frac{t}{N}} \vee \frac{t}{N} \vee \frac{t r_\infty(C)}{N} \right).$$

There exists a positive universal constant c such that, with probability at least $1 - ce^{-t}$,

$$\|\widehat{\Sigma} - \Sigma\| \lesssim R_q \rho_N^{1-q} \left(1 + \|A\|^2 \|\Gamma^{-1}\| \left(2\|C\| + R_q \rho_N^{1-q}\right) + \|A\|^4 \|\Gamma^{-1}\|^2 \|C\| (\|C\| + R_q \rho_N^{1-q})\right).$$

Proof. For the localized SR update we have $\widehat{\Sigma} = \mathscr{C}(\widehat{C}_{\rho_N})$. From Lemma A.6, the continuity of $\mathscr{C}$ implies that

$$\|\mathscr{C}(\widehat{C}_{\rho_N}) - \mathscr{C}(C)\| \leq \|\widehat{C}_{\rho_N} - C\| \Big(1 + \|A\|^2 \|\Gamma^{-1}\| \big(\|\widehat{C}_{\rho_N}\| + \|C\|\big) + \|A\|^4 \|\Gamma^{-1}\|^2 \|\widehat{C}_{\rho_N}\| \|C\|\Big).$$

Let E denote the event on which

$$\|\widehat{C}_{\rho_N} - C\| \lesssim R_q \rho_N^{1-q}.$$

From Theorem B.12, E has probability at least $1 - ce^{-t}$. It also holds on E that

$$\|\widehat{C}_{\rho_N}\| \leq \|C\| + \|\widehat{C}_{\rho_N} - C\| \lesssim \|C\| + R_q \rho_N^{1-q}.$$

Therefore, it holds on E that

$$\|\widehat{\Sigma} - \Sigma\| \lesssim R_q \rho_N^{1-q} \left(1 + \|A\|^2 \|\Gamma^{-1}\| \left(2\|C\| + R_q \rho_N^{1-q}\right) + \|A\|^4 \|\Gamma^{-1}\|^2 \|C\| (\|C\| + R_q \rho_N^{1-q})\right).$$

□