SIAM J. NUMER. ANAL.
Vol. 64, No. 2, pp. 391–429

ACCURACY OF THE ENSEMBLE KALMAN FILTER IN THE NEAR-LINEAR SETTING*

E. CALVELLO†, P. MONMARCHÉ‡, A. M. STUART†, AND U. VAES§

Abstract. The filtering distribution captures the statistics of the state of a possibly stochastic dynamical system from partial and noisy observations. Classical particle filters provably approximate this distribution in quite general settings; however, they behave poorly for high dimensional problems, suffering weight collapse. This issue is circumvented by the ensemble Kalman filter, which is an equal-weights interacting particle system. However, this finite particle system is only proven to approximate the true filter in the linear Gaussian case. In practice, however, it is applied in much broader settings; as a result, establishing its approximation properties more generally is important. There has been recent progress in the theoretical analysis of the algorithm in discrete time, establishing stability and error estimates, in relation to the true filter, in non-Gaussian settings; but the assumptions on the dynamics and observation models rule out the unbounded vector fields that arise in practice, and the analysis applies only to the mean field limit of the discrete time ensemble Kalman filter. The present work establishes error bounds between the filtering distribution and the finite particle discrete time ensemble Kalman filter when the dynamics and observation vector fields may be unbounded, allowing linear growth.

Key words. ensemble Kalman filter, stochastic filtering, weighted total variation metric, stability estimates, accuracy estimates, near-linear setting

MSC codes. 60G35, 62F15, 65C35, 70F45, 93E11

DOI. 10.1137/25M1732544

1. Introduction.

This paper describes a setting in which the ensemble Kalman filter (EnKF) accurately approximates the filtering distribution of discrete time, partially observed, stochastic dynamical systems. In subsection 1.1, we set the work in context by providing a literature review. Subsection 1.2 details our contributions and overviews the paper structure. In subsection 1.3 we establish notation used throughout the paper. Subsection 1.4 contains a precise statement of the filtering distribution which we approximate using the EnKF.

1.1. Literature review.

Algorithms for filtering employ noisy observations, arising from a (possibly stochastic) dynamical system, estimate the distribution of the system state conditional on the observations. The Kalman filter [20] determines the filtering distribution exactly for linear Gaussian dynamics and observations.

*Received by the editors February 7, 2025; accepted for publication (in revised form) October 24, 2025; published electronically March 17, 2026.
https://doi.org/10.1137/25M1732544
Funding: The research of PM is supported by the projects SWIDIMS (ANR-20-CE40-0022) and CONVIVIALITY (ANR-23-CE40-0003) of the French National Research Agency. The work of AMS is supported by a Department of Defense Vannevar Bush Faculty Fellowship and by the SciAI Center, funded by the Office of Naval Research (ONR), Grant Number N00014-23-1-2729. The work of EC was supported by the Resnick Sustainability Institute, the SciAI Center and the Vannevar Bush Faculty Fellowship held by AMS. UV is partially supported by the European Research Council (ERC) under the EU Horizon 2020 program (grant agreement No 810367) and by the Agence Nationale de la Recherche under grants ANR-21-CE40-0006 (SINEQ) and ANR-23-CE40-0027 (IPSO).
†Department of Computing and Mathematical Sciences, Caltech, USA (e.calvello@caltech.edu, astuart@caltech.edu).
‡LAMA, Université Gustave Eiffel, Université Paris Est Créteil, CNRS, and Institut Universitaire de France, France (pierre.monmarche@univ-eiffel.fr).
§MATHERIALS, Inria Paris, and CERMICS, École des Ponts, France (urbain.vaes@inria.fr).

The extended Kalman filter generalizes the Kalman filtering methodology to nonlinear problems and is based on a linearization approximation; see [19] and [2]. The linearization approximation leads to an inexact distribution for nonlinear problems and furthermore requires evaluation of covariance matrices, making the methodology impractical for large-scale geophysical applications [18]. Particle filters or sequential Monte Carlo methods [15, 10] offer an alternative methodology for nonlinear filtering problems, allowing recovery of the exact filtering distribution in the large particle limit. However, this class of methods scales poorly with dimension and, in particular, suffers weight collapse, making its application to geophysical problems prohibitive; see, for example, [28, 5, 26, 1].

For this reason, the EnKF has emerged as a popular robust methodology for filtering, especially in high dimensional problems. We refer to [9] for a detailed discussion of the literature in areas aligned with, or adjacent to, the results we establish about the EnKF. In this literature review, we confine ourselves primarily to overviewing a narrow subset of the work extensively reviewed in [9] with a focus on several of the open questions detailed in our closing remarks. The EnKF was introduced in the seminal paper [16]. Its success stems from the low-rank approximation of large covariances, cheaply computed using an ensemble of particles, which allows the methodology to be deployed in geophysical applications. Quantifying the accuracy of the ensemble Kalman filter has been of great practical and theoretical interest. In paper [21], accuracy of the EnKF, viewed as a state estimator and not in terms of the filtering distribution, is established. This work is undertaken in both discrete and continuous time. This continuous time analysis of state estimation accuracy was improved upon in [11], where a variance inflation condition from [21] was dispensed with. The continuous time model used in [21, 11], however, was not rigorously justified as a continuous time limit of the discrete time EnKF until [22].

For nonlinear non-Gaussian settings, which are of particular practical interest, analysis of the accuracy of the EnKF in relation to the true filtering distribution, rather than just state estimation, remains in its infancy. The papers [24, 25] studied this issue in the linear Gaussian setting where the mean field limit of the Kalman filter is exact; they demonstrated that the ensemble Kalman filter may be viewed as an interacting particle system approximation of the mean field limit and established Monte Carlo type error bounds. The recent article [8] overviewed the formulation of ensemble Kalman methods using mean field dynamical systems and provided a platform from which the analyses of [24, 25] may be generalized beyond the linear Gaussian setting. We refer the reader to [8] for a comprehensive overview of mean-field limits of ensemble Kalman filters. In the recent paper [9], the authors established stability properties of the discrete time mean field ensemble Kalman filter and used them to prove accuracy of the filter in a near-Gaussian setting.[1] However, this paper does not consider finite particle approximations of the mean field, and the conditions on the dynamics-observation model require boundedness of the vector fields arising. In this paper, we address both these issues, establishing error bounds between the finite particle discrete time ensemble Kalman filter and the true filtering distribution in settings where the dynamics and observation vector fields grow linearly.

[1] We note that in [9], the authors use "near-Gaussian" to characterize the joint distribution of state and data, which leads to a near-Gaussian filtering distribution. However, in their context, the additional assumption of bounded dynamics and observation model vector fields limits their results to the setting of near-constant vector fields. In this work, we are more general and use "near-linear" to characterize the dynamics and observation model vector fields: this also leads to near-Gaussian filtering distributions.

We note that extending our work in this paper, concerning filter accuracy, to the continuous time setting established in [22] remains open. Likewise, an important open question stemming from our work is to establish estimates valid on the infinite time horizon in either the discrete or the continuous time setting. In the continuous time setting, this is studied for ensemble Kalman approximation of the Kalman-Bucy filter when the true filter is Gaussian, in [13]; the paper [6] comprehensively reviews this line of work. The paper [12] contains long time stability estimates for the sample means and sample variances of the discrete ensemble Kalman filter in nonlinear, non-Gaussian settings but is confined to one dimension. Stability estimates of this kind will be needed in any generalization of our work to the infinite time horizon.

1.2. Contributions and outline. We make the following contributions.

1. Theorem 2.1 is a stability result for the discrete time mean field ensemble Kalman filter in the setting of dynamical models and Lipschitz obervation operators that grow at most linearly at infinity.
2. Theorem 2.3 quantifies the error between the discrete time mean field ensemble Kalman filter and the true filter in the setting of dynamical models and Lipschitz observation operators that are near-linear.
3. Theorem 2.4 quantifies the error between the finite particle discrete time ensemble Kalman filter (found as an interacting particle system approximation of the mean field) and the true filter in the setting of near-linear, Lipschitz dynamical models, and linear observation operators.

When going beyond the work in [9], the current paper simultaneously addresses a more applicable problem class by allowing linear growth in the dynamics and observation operators and confronts the substantial technical challenges which arise from doing so. Indeed, allowing for unbounded vector fields enables the application of this analysis to the setting of the Kalman filter; in addition, it enables the integration of this mean field analysis with finite-particle error estimates developed in the literature, such as in [24]. Bounds on moments of the filtering distribution and mean field ensemble Kalman filter must be established; these bounds grow in the number of iterations, which is in contrast to [9], where the L^∞ bounds on model and observation operators ensure a uniform bound on moments. A further challenge is presented by the need to establish stability bounds for the conditioning map, giving rise to the true filter, and the transport map giving rise to the ensemble Kalman filter. These results exhibit dependence on moments in the stability constants and require control of the growth given by the dynamics; this is again in contrast to [9], where the L^∞ bounds allow for a uniform control.

After discussing notation that will be used throughout the paper in subsection 1.3, we introduce the filtering problem in subsection 1.4. In section 2, we outline the main results of the paper concerning the ensemble Kalman filter. In subsection 2.1, we formulate the ensemble Kalman filter that we consider in this paper, along with the relevant assumptions we will use in the analysis. In subsection 2.2, we state a stability theorem for the mean field formulation of the ensemble Kalman filter, hence Contribution 1. We leverage this result in subsection 2.3 to derive a theorem quantifying the error between the mean field ensemble Kalman filter and the true filter, Contribution 2. Finally, in subsection 2.4, we make use of the results of the previous subsections to state a theorem quantifying the error between the (finite particle) ensemble Kalman filter itself and the true filter, yielding Contribution 3. Various technical results, used in the proof of our three main theorems, may be found in the appendices. We conclude with closing remarks in section 3.

1.3. Notation. We denote by $|\bullet|$ the Euclidean norm on $\mathbb{R}^n$, while the induced operator norm on matrices is denoted by $\|\bullet\|$. For symmetric positive definite matrix $S \in \mathbb{R}^{n\times n}$, we denote by $|\bullet|_S$ the weighted norm $|S^{-1/2}\bullet|$. Given a function $f\colon \mathbb{R}^n \to \mathbb{R}$ and $r \geqslant 0$, we let $B_{L^\infty}(f,r)$ denote the L^∞ ball of radius r centered at f. Similarly, for functions f,g and $r \geqslant 0$, we denote by $B_{L^\infty}((f,g),r)$ the L^∞ ball of radius r centered at (f,g) on the product space, i.e., the set of all $(\mathfrak{f},\mathfrak{g}) \in L^\infty(\mathbb{R}^n) \times L^\infty(\mathbb{R}^d)$ satisfying

$$\|\mathfrak{f} - f\|_{L^\infty(\mathbb{R}^n)} \leqslant r, \quad \|\mathfrak{g} - g\|_{L^\infty(\mathbb{R}^d)} \leqslant r.$$

We write $|\bullet|_{C^{0,1}}$ for the $C^{0,1}$ seminorm, referring in particular to the Lipschitz constant.

We apply the notation $\perp\!\!\!\perp$ to denote independence of two random variables. We write $\mathcal{N}(m,C)$ to denote the Gaussian distribution with mean $m \in \mathbb{R}^n$ and covariance $C \in \mathbb{R}^{n\times n}$. We use $\mathcal{P}(\mathbb{R}^n)$ to denote the space of probability measures over $\mathbb{R}^n$, while we write $\mathcal{G}(\mathbb{R}^n)$ for the space of Gaussian probability measures over $\mathbb{R}^n$. We will denote by $\mathcal{P}^p(\mathbb{R}^n)$ the set of probability measures over $\mathbb{R}^n$ with finite moments up to order p. With the exception of empirical measures formed from finite ensembles, this manuscript primarily deals with probability measures that have a Lebesgue density because of the assumptions concerning the noise structure in the dynamics model and the data acquisition model. In this setting, we abuse notation by using the same symbols for probability measures and their densities. For $\mu \in \mathcal{P}(\mathbb{R}^n)$, the notation $\mu(x)$ for $x \in \mathbb{R}^n$ refers to the Lebesgue density of μ evaluated at x, while $\mu[f]$ for a function $f\colon \mathbb{R}^n \to \mathbb{R}$ will be notation for $\int_{\mathbb{R}^n} f\,\mathrm{d}\mu$. For function $f : \mathbb{R}^n \to \mathbb{R}^m$ and measure $\mu \in \mathcal{P}(\mathbb{R}^n)$, we denote by $f_\sharp\mu \in \mathcal{P}(\mathbb{R}^m)$ the measure given by the pushforward, under f, of μ.

For a probability measure $\mu \in \mathcal{P}(\mathbb{R}^n)$, the notation $\mathcal{M}_q(\mu)$ denotes the qth polynomial moment under the measure μ, defined as

$$\mathcal{M}_q(\mu) := \int_{\mathbb{R}^n} |x|^q \mu(\mathrm{d}x). \tag{1.1}$$

We denote by $\mathcal{M}(\mu)$ and $\mathcal{C}(\mu)$ the mean and covariance under μ, respectively, as follows:

$$\mathcal{M}(\mu) = \mu[x], \qquad \mathcal{C}(\mu) = \mu\Big[\big(x - \mathcal{M}(\mu)\big) \otimes \big(x - \mathcal{M}(\mu)\big)\Big].$$

We use $\mathcal{P}_R(\mathbb{R}^n)$ with $R \geqslant 1$ to denote the subset of $\mathcal{P}(\mathbb{R}^n)$ of probability measures with mean and covariance satisfying the bounds

$$|\mathcal{M}(\mu)| \leqslant R, \qquad \frac{1}{R^2} I_n \preccurlyeq \mathcal{C}(\mu) \preccurlyeq R^2 I_n, \tag{1.2}$$

where I_n denotes the identity matrix in $\mathbb{R}^{n\times n}$, and $\preccurlyeq$ is the partial ordering defined by the convex cone of positive semidefinite matrices. Additionally, $\mathcal{G}_R(\mathbb{R}^n)$ is the subset of $\mathcal{G}(\mathbb{R}^n)$ of probability measures satisfying (1.2).

For $\pi \in \mathcal{P}(\mathbb{R}^{d_u} \times \mathbb{R}^{d_y})$ defined on the joint space of state and data associated with random variable $(u,y) \in \mathbb{R}^{d_u} \times \mathbb{R}^{d_y}$ we denote by $\mathcal{M}^u(\pi)$, $\mathcal{M}^y(\pi)$ the means of the marginal distributions and by $\mathcal{C}^{uu}(\pi)$, $\mathcal{C}^{uy}(\pi)$ and $\mathcal{C}^{yy}(\pi)$ the blocks of the covariance matrix $\mathcal{C}(\pi)$. In particular, we write

$$\mathcal{M}(\pi) = \begin{pmatrix} \mathcal{M}^u(\pi) \\ \mathcal{M}^y(\pi) \end{pmatrix}, \qquad \mathcal{C}(\pi) = \begin{pmatrix} \mathcal{C}^{uu}(\pi) & \mathcal{C}^{uy}(\pi) \\ \mathcal{C}^{uy}(\pi)^\mathsf{T} & \mathcal{C}^{yy}(\pi) \end{pmatrix}. \tag{1.3}$$

For $h\colon \mathbb{R}^{d_u} \to \mathbb{R}^{d_y}$ we also define $\mathcal{C}^{hh}(\pi)$ to be the covariance of vector $h(u)$ for u distributed according to the marginal of π on u, and $\mathcal{C}^{uh}(\pi)$ to be the covariance between u and $h(u)$. Given this notation, we also introduce the set $\mathcal{P}^2_{\succ 0}(\mathbb{R}^{d_u \times d_y})$ of probability measures π with finite second moment satisfying $\mathcal{C}^{yy}(\pi) \succ 0$.

Throughout, we will denote operators acting on the space of probability measures via the mathsf font. We introduce the Gaussian projection operator $\mathsf{G}\colon \mathcal{P}^2(\mathbb{R}^n) \to \mathcal{G}(\mathbb{R}^n)$ defined by $\mathsf{G}\mu = \mathcal{N}\big(\mathcal{M}(\mu), \mathcal{C}(\mu)\big)$. We refer to G as a projection because $\mathsf{G}\mu$ is the Gaussian distribution closest to μ with respect to $\mathrm{KL}(\mu\|\bullet)$ [7], where $\mathrm{KL}(\mu\|\nu)$ denotes the Kullback–Leibler (KL) divergence of μ from ν. We note that G defines a nonlinear mapping. Throughout the paper, we make use of the following weighted total variation distance as in [9].

Definition 1.1. *Let $g\colon \mathbb{R}^n \to [1,\infty)$ define the function $g(v) := 1 + |v|^2$. The weighted total variation metric $d_g\colon \mathcal{P}(\mathbb{R}^n) \times \mathcal{P}(\mathbb{R}^n) \to \mathbb{R}$ is defined by*

$$d_g(\mu_1,\mu_2) = \sup_{|f|\leqslant g} \big|\mu_1[f] - \mu_2[f]\big|,$$

where the supremum is taken over all functions $f\colon \mathbb{R}^n \to \mathbb{R}$ bounded from above by g pointwise and in absolute value.

Remark 1.2. The following remarks concern the distance d_g:

- If μ_1, μ_2 have Lebesgue densities ρ_1, ρ_2, then

$$d_g(\mu_1,\mu_2) = \int g(v)|\rho_1(v) - \rho_2(v)|\,\mathrm{d}v.$$

- Unlike the usual total variation distance, the weighted metric in Definition 1.1 enables control of the differences $|\mathcal{M}(\mu_1) - \mathcal{M}(\mu_2)|$ and $\|\mathcal{C}(\mu_1) - \mathcal{C}(\mu_2)\|$. This is the content of Lemma B.6, stated in Appendix B.1, which is used in the proof of a key auxiliary result (Lemma B.11).

1.4. Filtering distribution. Here, we introduce the hidden Markov model that gives rise to the filtering distribution. We employ notation similar to that established in [8, 9]. We consider $\{u_j\}_{j\in[\![0,J]\!]} \subset \mathbb{R}^{d_u}$ to be unknown states to be determined from associated observations $\{y_j\}_{j\in[\![1,J]\!]} \subset \mathbb{R}^{d_y}$. We assume the states and observations to be governed by the following stochastic dynamics and data model:

$$u_{j+1} = \Psi(u_j) + \xi_j, \qquad \xi_j \sim \mathcal{N}(0,\Sigma), \tag{1.4a}$$

$$y_{j+1} = h(u_{j+1}) + \eta_{j+1}, \qquad \eta_{j+1} \sim \mathcal{N}(0,\Gamma). \tag{1.4b}$$

We assume that the initial state is a Gaussian random variable $u_0 \sim \mathcal{N}(m_0, C_0)$ and that the following independence condition is satisfied:

$$u_0 \perp\!\!\!\perp \xi_0 \perp\!\!\!\perp \cdots \perp\!\!\!\perp \xi_{J-1} \perp\!\!\!\perp \eta_1 \perp\!\!\!\perp \cdots \perp\!\!\!\perp \eta_J. \tag{1.5}$$

We define the conditional distribution of the state u_j given a realization $Y_j^\dagger := \{y_1^\dagger, \dots, y_j^\dagger\}$ as the *filtering distribution,* which we denote by μ_j. We also refer to μ_j as the *true filter.* In the context of this paper, data $Y_j^\dagger$ are assumed to be a collection of realizations from (1.4). As formulated in [23, 27], the evolution of the filtering distribution may be written as

$$\mu_{j+1} = \mathsf{L}_j \mathsf{P} \mu_j, \tag{1.6}$$

where P and L_j are maps from $\mathcal{P}(\mathbb{R}^{d_u})$ into itself and which effect what are referred to in the data assimilation community as the *prediction* and *analysis* steps, respectively [3]. The prediction operator P is linear and is determined by the Markov kernel arising from (1.4a). On the other hand, the operator L_j is nonlinear and encodes the incorporation of the new data point $y^\dagger_{j+1}$ using Bayes' theorem. These operators are defined via action on probability measures with Lebesgue density μ as

$$\mathsf{P}\mu(u) = \frac{1}{\sqrt{(2\pi)^{d_u}\det\Sigma}}\int_{\mathbb{R}^{d_u}} \exp\left(-\frac{1}{2}\left|u-\Psi(v)\right|^2_\Sigma\right)\mu(v)\,\mathrm{d}v, \tag{1.7a}$$

$$\mathsf{L}_j\mu(u) = \frac{\exp\left(-\frac{1}{2}\left|y^\dagger_{j+1}-h(u)\right|^2_\Gamma\right)\mu(u)}{\int_{\mathbb{R}^{d_u}}\exp\left(-\frac{1}{2}\left|y^\dagger_{j+1}-h(U)\right|^2_\Gamma\right)\mu(U)\,\mathrm{d}U}. \tag{1.7b}$$

We may rewrite the analysis map L_j as the composition $\mathsf{B}_j\mathsf{Q}$, where the operators $\mathsf{Q}\colon \mathcal{P}(\mathbb{R}^{d_u})\to\mathcal{P}(\mathbb{R}^{d_u}\times\mathbb{R}^{d_y})$ and $\mathsf{B}_j\colon \mathcal{P}(\mathbb{R}^{d_u}\times\mathbb{R}^{d_y})\to\mathcal{P}(\mathbb{R}^{d_u})$ are defined by

$$\mathsf{Q}\mu(u,y) = \frac{1}{\sqrt{(2\pi)^{d_y}\det\Gamma}}\exp\left(-\frac{1}{2}\left|y-h(u)\right|^2_\Gamma\right)\mu(u), \tag{1.8a}$$

$$\mathsf{B}_j\mu(u) = \frac{\mu(u,y^\dagger_{j+1})}{\int_{\mathbb{R}^{d_u}}\mu(U,y^\dagger_{j+1})\,\mathrm{d}U}. \tag{1.8b}$$

Linear operator Q maps a probability measure with density μ into the law of random variable $(U, h(U)+\eta)$, with $U\sim\mu$ independent of $\eta\sim\mathcal{N}(0,\Gamma)$. Nonlinear operator B_j effects conditioning on the datum $y^\dagger_{j+1}$. Hence, we may reformulate (1.6) as

$$\mu_{j+1} = \mathsf{B}_j\mathsf{Q}\mathsf{P}\mu_j. \tag{1.9}$$

2. The ensemble Kalman filter.

We begin in subsection 2.1 by introducing the specific form of the mean field ensemble Kalman filter that we consider in this paper; other versions leading to implementable algorithms and for which similar analysis may be developed can be found in [8]. In this subsection, we also outline the various assumptions that will be employed for the results of the paper. We note that the subsections that follow proceed with increasingly restrictive assumptions on dynamics and observation operator. However, all of the theorems allow for linear growth of the vector fields defining the dynamics and the observation processes. In subsection 2.2, we state and prove a stability theorem, which shows that the error between the true filter and the mean field ensemble Kalman filter may be controlled by the error between the true filter and its Gaussian projection on the joint space of state and observations. In subsection 2.3, we establish an error estimate which shows that the mean field ensemble Kalman filter provides an accurate approximation of the true filtering distribution for dynamics and observation operators that are close-to-linear. In subsection 2.4, we deploy the results of the two preceding subsections to state a theorem quantifying the error between the finite particle ensemble Kalman filter itself and the true filtering distribution.

2.1. The algorithm.

The ensemble Kalman filter as implemented in practice may be derived as a particle approximation of various mean field dynamics [8]. The particular mean field ensemble Kalman filter that we consider in this paper may be written as

$$\widehat{u}_{j+1} = \Psi(u_j) + \xi_j, \qquad \xi_j \sim \mathcal{N}(0,\Sigma), \tag{2.1a}$$
$$\widehat{y}_{j+1} = h(\widehat{u}_{j+1}) + \eta_{j+1}, \qquad \eta_{j+1} \sim \mathcal{N}(0,\Gamma), \tag{2.1b}$$
$$u_{j+1} = \widehat{u}_{j+1} + \mathcal{C}^{uy}\left(\widehat{\pi}^{\mathrm{EK}}_{j+1}\right)\mathcal{C}^{yy}\left(\widehat{\pi}^{\mathrm{EK}}_{j+1}\right)^{-1}\left(y^{\dagger}_{j+1} - \widehat{y}_{j+1}\right), \tag{2.1c}$$

where $\widehat{\pi}^{\mathrm{EK}}_{j+1} = \mathrm{Law}(\widehat{u}_{j+1}, \widehat{y}_{j+1})$ and where the independence condition (1.5) holds. We refer the reader back to subsection 1.3 for the definition of the covariance matrices appearing in (2.1c). We denote by μ^{EK}_j the law of u_j. We aim to formulate the evolution of μ^{EK}_j in terms of maps on probability measures; hence, we introduce, for a given $y^{\dagger}_{j+1}$, the map $\mathsf{T}_j : \mathcal{P}^2_{\succ 0}(\mathbb{R}^{d_u} \times \mathbb{R}^{d_y}) \to \mathcal{P}(\mathbb{R}^{d_u})$ defined by

$$\mathsf{T}_j \pi = \mathscr{T}(\bullet, \bullet; \pi, y^{\dagger}_{j+1})_{\sharp}\pi. \tag{2.2}$$

Here, for given $\pi \in \mathcal{P}^2_{\succ 0}(\mathbb{R}^{d_u} \times \mathbb{R}^{d_y})$, $z \in \mathbb{R}^{d_y}$, the transport map $\mathscr{T}$ is defined as

$$\mathscr{T}(\bullet, \bullet; \pi, z) : \mathbb{R}^{d_u} \times \mathbb{R}^{d_y} \to \mathbb{R}^{d_u}; \tag{2.3a}$$
$$(u, y) \mapsto u + \mathcal{C}^{uy}(\pi)\mathcal{C}^{yy}(\pi)^{-1}\big(z - y\big). \tag{2.3b}$$

Evolution of the probability measure μ^{EK}_j may then be written (see [8]) as

$$\mu^{\mathrm{EK}}_{j+1} = \mathsf{T}_j \mathsf{Q}\mathsf{P}\mu^{\mathrm{EK}}_j. \tag{2.4}$$

We note that a specific instance of map $\mathscr{T}$ defined in (2.3) appears in (2.1c) and is determined by probability measure $\pi := \widehat{\pi}^{\mathrm{EK}}_{j+1}$ (here, equal to $\mathsf{Q}\mathsf{P}\mu^{\mathrm{EK}}_j$) and data $z := y^{\dagger}_{j+1}$. We refer to [9] for a step-by-step argument detailing why the law of u_j in (2.1) evolves according to (2.4). The operator T_j is equivalent, over the Gaussians $\mathcal{G}(\mathbb{R}^{d_u} \times \mathbb{R}^{d_y}) \subset \mathcal{P}(\mathbb{R}^{d_u} \times \mathbb{R}^{d_y})$, to the conditioning operator B_j; see Lemma B.7. We recall that in the particular case where μ_0 is Gaussian, which we assume throughout the paper, and the operators Ψ and h are linear, that the mean field ensemble Kalman filter (2.4) coincides with the filtering distribution (1.9). However, this paper focuses on the aim of analyzing the accuracy of ensemble Kalman methods when Ψ and h are not assumed to be linear.

The ensemble Kalman filter as implemented in practice may be derived as a particle approximation of the mean field dynamics as defined by sample paths in (2.1) or as an evolution on measures in (2.4). To this end, we observe that for any π in the range of Q, $\mathcal{C}^{yy}(\pi) = \mathcal{C}^{hh}(\pi) + \Gamma$ and $\mathcal{C}^{uy}(\pi) = \mathcal{C}^{uh}(\pi)$ using the notation introduced in, and following, (1.3). This is since the noise in the observation component defining π is then independent of the state component defining π. Using this observation, the particle approximation of (2.1) takes the form

$$\widehat{u}^{(i)}_{j+1} = \Psi\big(u^{(i)}_j\big) + \xi^{(i)}_j, \qquad \xi^{(i)}_j \sim \mathcal{N}(0,\Sigma), \tag{2.5a}$$
$$\widehat{y}^{(i)}_{j+1} = h\big(\widehat{u}^{(i)}_{j+1}\big) + \eta^{(i)}_{j+1}, \qquad \eta^{(i)}_{j+1} \sim \mathcal{N}(0,\Gamma), \tag{2.5b}$$
$$u^{(i)}_{j+1} = \widehat{u}^{(i)}_{j+1} + \mathcal{C}^{uh}\left(\widehat{\pi}^{\mathrm{EK},N}_{j+1}\right)\left(\mathcal{C}^{hh}\left(\widehat{\pi}^{\mathrm{EK},N}_{j+1}\right) + \Gamma\right)^{-1}\big(y^{\dagger}_{j+1} - \widehat{y}^{(i)}_{j+1}\big), \tag{2.5c}$$

where $\widehat{\pi}_{j+1}^{\mathrm{EK},N}$ is the empirical measure

$$\widehat{\pi}_{j+1}^{\mathrm{EK},N} = \frac{1}{N}\sum_{i=1}^{N} \delta_{\left(\widehat{u}_{j+1}^{(i)}, \widehat{y}_{j+1}^{(i)}\right)},$$

and $\xi_j^{(i)} \sim \mathcal{N}(0,\Sigma)$ i.i.d. in both i and j and $\eta_{j+1}^{(i)} \sim \mathcal{N}(0,\Gamma)$ i.i.d. in both i and j; furthermore, the set of $\{\xi_j^{(i)}\}$ is independent from the set of $\{\eta_{j+1}^{(i)}\}$. Choosing to express the particle approximation of the covariance in observation space through $\mathcal{C}^{hh}$ and Γ ensures invertibility, provided Γ is invertible. From the particles evolving according to the dynamics in (2.5), we define the empirical measure

$$\mu_{j+1}^{\mathrm{EK},N} = \frac{1}{N}\sum_{i=1}^{N} \delta_{u_{j+1}^{(i)}}, \tag{2.6}$$

whose evolution describes the ensemble Kalman filter.

Our theorems in subsections 2.2 and 2.3 regard the relationship between the true filter (1.9) and the mean field ensemble Kalman filter (2.4). In subsection 2.2, we consider the setting in which Ψ and h exhibit linear growth but are not assumed to be linear; hence, the true filter is not Gaussian. In subsection 2.3, we then consider small perturbations of the Gaussian setting that arise when the vector fields Ψ and h are close to affine. The theorem in subsection 2.4 concerns the relationship between the true filter (1.9) and the ensemble Kalman filter (2.6). In this subsection, we combine existing analysis on the convergence of the ensemble Kalman filter to the mean field ensemble Kalman filter with the results from the previous subsections to state and prove an error estimate between the ensemble Kalman filter and the true filter in the nonlinear setting. Specifically, we study the case of a vector field Ψ that is a bounded perturbation away from affine and an affine vector field h. To state our theorems, we will use the following set of assumptions.

Assumption H. There exist positive constants $\kappa_y, \kappa_\Psi, \kappa_h, \ell_h, \sigma$, and γ such that the data $\{y_j^\dagger\}$, the vector fields (Ψ, h), and the covariances (Σ, Γ) satisfy

(H1) data $Y^\dagger = \{y_j^\dagger\}_{j=1}^J$ lies in set $B_y \subset \mathbb{R}^{KJ}$ defined by

$$B_y := \left\{ Y^\dagger \in \mathbb{R}^{KJ} : \max_{j\in[\![1,J]\!]} |y_j^\dagger| \leqslant \kappa_y \right\};$$

(H2) covariance matrices Σ and Γ are positive definite: $\Sigma \succcurlyeq \sigma I_{d_u}$ and $\Gamma \succcurlyeq \gamma I_{d_y}$ for positive σ and γ;

(H3) function $\Psi\colon \mathbb{R}^{d_u} \to \mathbb{R}^{d_u}$ satisfies $\left|\Psi(u)\right| \leqslant \kappa_\Psi(1+|u|)$ for all $u \in \mathbb{R}^{d_u}$;

(H4) function $h\colon \mathbb{R}^{d_u} \to \mathbb{R}^{d_y}$ satisfies $\left|h(u)\right| \leqslant \kappa_h(1+|u|)$ for all $u \in \mathbb{R}^{d_u}$;

(H5) function $h\colon \mathbb{R}^{d_u} \to \mathbb{R}^{d_y}$ satisfies $|h|_{C^{0,1}} \leqslant \ell_h < \infty$.

Assumption V. The vector fields (Ψ, h) are affine; i.e., they satisfy the following:

(V1) the function $\Psi\colon \mathbb{R}^{d_u} \to \mathbb{R}^{d_u}$ satisfies $\Psi(u) := Mu + b$ for some $M \in \mathbb{R}^{d_u \times d_u}$ and $b \in \mathbb{R}^{d_u}$;

(V2) the function $h\colon \mathbb{R}^{d_u} \to \mathbb{R}^{d_y}$ satisfies $h(u) := Hu + w$ for some $H \in \mathbb{R}^{d_y \times d_u}$ and $w \in \mathbb{R}^{d_y}$.

2.2. Stability theorem: Mean field ensemble Kalman filter. In this section, we consider the setting of dynamic and observation operators satisfying the linear growth assumptions contained in Assumption H. Informally, the result of Theorem 2.1 shows that, given these assumptions and if the true filtering distributions

$(\mu_j)_{j\in[\![0,J-1]\!]}$ are close to Gaussian after lifting to the joint state and data space, then the distribution μ_j^{EK} given by the mean field ensemble Kalman filter (2.4) is close to the filtering distribution μ_j given by (1.9) for all $j\in[\![0,J]\!]$.

THEOREM 2.1 (Stability: mean field ensemble Kalman filter). *Assume that the probability measures* $(\mu_j)_{j\in[\![0,J]\!]}$ *and* $(\mu_j^{\mathrm{EK}})_{j\in[\![0,J]\!]}$ *are given respectively by the dynamical systems* (1.9) *and* (2.4), *initialized at the Gaussian probability measure* $\mu_0=\mu_0^{\mathrm{EK}}\in\mathcal{G}(\mathbb{R}^{d_u})$. *That is,*

$$\mu_{j+1}=\mathsf{B}_j\mathsf{Q}\mathsf{P}\mu_j,\qquad \mu_{j+1}^{\mathrm{EK}}=\mathsf{T}_j\mathsf{Q}\mathsf{P}\mu_j^{\mathrm{EK}}.$$

If Assumption H *holds, then there exists* $C=C(\mathcal{M}_{\max\{3+d_u,4+d_y\}}(\mu_0),\kappa_y,\kappa_\Psi,\kappa_h,\ell_h,\Sigma,\Gamma,J)$ *such that*

$$d_g(\mu_J^{\mathrm{EK}},\mu_J)\leqslant C\max_{j\in[\![0,J-1]\!]} d_g(\mathsf{Q}\mathsf{P}\mu_j,\mathsf{G}\mathsf{Q}\mathsf{P}\mu_j).$$

Proof of Theorem 2.1 below relies on the following auxiliary results, all proved in Appendix B.

1. For any probability measure μ with finite first and second order polynomial moments $\mathcal{M}_1(\mu)$ and $\mathcal{M}_2(\mu)$, the means of $\mathsf{P}\mu$ and $\mathsf{Q}\mathsf{P}\mu$ are bounded from above, and their covariances are bounded from both above and from below. The constants in these bounds depend only on the parameters κ_Ψ, κ_h, Σ,Γ and on $\mathcal{M}_1(\mu)$ and $\mathcal{M}_2(\mu)$. See Lemmas B.1 and B.2.
2. Let $(\mu_j)_{j\in[\![1,J]\!]}$ and $(\mu_j^{\mathrm{EK}})_{j\in[\![1,J]\!]}$ denote the probability measures obtained respectively from the dynamical systems (1.9) and (2.4), initialized at the same Gaussian measure $\mu_0=\mu_0^{\mathrm{EK}}\in\mathcal{G}(\mathbb{R}^{d_u})$. Then, for any integer $q\geqslant 2$, there exist constants $M_q,M_q^{\mathrm{EK}}<\infty$ depending on $\mathcal{M}_q(\mu_0),\kappa_y,\kappa_\Psi,\kappa_h,\Sigma,\Gamma,J$ so that
$$\max_{j\in[\![0,J]\!]}\mathcal{M}_q(\mu_j)\leqslant M_q\quad\text{and}\quad\max_{j\in[\![0,J]\!]}\mathcal{M}_q\big(\mu_j^{\mathrm{EK}}\big)\leqslant M_q^{\mathrm{EK}}.$$
See Lemmas B.3 and B.5. This will facilitate use of the stability results from Items 6 and 7.
3. For any Gaussian measure $\mu\in\mathcal{G}(\mathbb{R}^{d_u}\times\mathbb{R}^{d_y})$, we have that $\mathsf{B}_j\mu=\mathsf{T}_j\mu$. See Lemma B.7 and [8, 9].
4. The map P is Lipschitz on $\mathcal{P}(\mathbb{R}^{d_u})$ for the metric d_g. The Lipschitz constant L_{P} depends only on the parameters κ_Ψ and Σ. See Lemma B.8.
5. The map Q is Lipschitz on $\mathcal{P}(\mathbb{R}^{d_u})$ for the metric d_g. The Lipschitz constant L_{Q} depends only on the parameters κ_h and Γ. See Lemma B.9.
6. The map B_j satisfies, for all $\pi\in\{\mathsf{Q}\mathsf{P}\mu:\mu\in\mathcal{P}(\mathbb{R}^{d_u})\text{ and }\mathcal{M}_2(\mu)<\infty\}\subset\mathcal{P}(\mathbb{R}^{d_u}\times\mathbb{R}^{d_y})$, it holds that
$$\forall j\in[\![0,J]\!],\qquad d_g(\mathsf{B}_j\mathsf{G}\pi,\mathsf{B}_j\pi)\leqslant C_{\mathsf{B}}d_g(\mathsf{G}\pi,\pi),$$
where $C_{\mathsf{B}}=C_{\mathsf{B}}(\mathcal{M}_2(\mu),\kappa_y,\kappa_\Psi,\kappa_h,\Sigma,\Gamma)$. Namely, this item regards the stability of the operator B_j between a probability measure and its Gaussian projection. See Lemma B.10.
7. The map T_j satisfies the following bound: for all $R\geqslant 1$, it holds for all $\pi\in\mathcal{P}_R(\mathbb{R}^{d_u}\times\mathbb{R}^{d_y})$ and $p\in\{\mathsf{Q}\mathsf{P}\mu:\mu\in\mathcal{P}(\mathbb{R}^{d_u})\text{ and }\mathcal{M}_{2\cdot\max\{3+d_u,4+d_y\}}(\mu)<\infty\}\subset\mathcal{P}(\mathbb{R}^{d_u}\times\mathbb{R}^{d_y})$ that
$$\forall j\in[\![0,J]\!],\qquad d_g(\mathsf{T}_j\pi,\mathsf{T}_jp)\leqslant L_{\mathsf{T}}\,d_g(\pi,p),$$
for a constant $L_{\mathsf{T}}=L_{\mathsf{T}}(R,\mathcal{M}_{2\cdot\max\{3+d_u,4+d_y\}}(\mu),\kappa_y,\kappa_\Psi,\kappa_h,\ell_h,\Sigma,\Gamma)$. Namely, this item may be regarded as a local Lipschitz continuity result. See Lemma B.11.

Proof of Theorem 2.1. Within the following argument, we make reference to the list of items outlined above. We begin by defining for ease of exposition the following measure of difference between the true filter and its Gaussian projection:

$$\varepsilon := \max_{j\in[\![0,J-1]\!]} d_g(\mathsf{QP}\mu_j, \mathsf{GQP}\mu_j). \tag{2.7}$$

We assume throughout the argument that $j \in [\![0, J-1]\!]$. The proof rests upon the following application of the triangle inequality:

$$\begin{aligned} d_g(\mu_{j+1}^{\mathrm{EK}}, \mu_{j+1}) &= d_g\left(\mathsf{T}_j\mathsf{QP}\mu_j^{\mathrm{EK}}, \mathsf{B}_j\mathsf{QP}\mu_j\right) \\ &\leqslant d_g\left(\mathsf{T}_j\mathsf{QP}\mu_j^{\mathrm{EK}}, \mathsf{T}_j\mathsf{QP}\mu_j\right) + d_g\left(\mathsf{T}_j\mathsf{QP}\mu_j, \mathsf{T}_j\mathsf{GQP}\mu_j\right) \\ &\quad + d_g\left(\mathsf{B}_j\mathsf{GQP}\mu_j, \mathsf{B}_j\mathsf{QP}\mu_j\right). \end{aligned} \tag{2.8a}$$

Here, we have used the fact that $\mathsf{T}_j\mathsf{GQP}\mu_j = \mathsf{B}_j\mathsf{GQP}\mu_j$ by Item 3 (Lemma B.7). Item 2 (Lemmas B.3 and B.5) shows that, for any integer $q \geqslant 2$, there exist constants $M_q, M_q^{\mathrm{EK}} < \infty$ depending on $\mathcal{M}_q(\mu_0), \kappa_y, \kappa_\Psi, \kappa_h, \Sigma, \Gamma, J$ so that

$$\max_{j\in[\![0,J]\!]} \mathcal{M}_q(\mu_j) \leqslant M_q \quad \text{and} \quad \max_{j\in[\![0,J]\!]} \mathcal{M}_q\left(\mu_j^{\mathrm{EK}}\right) \leqslant M_q^{\mathrm{EK}}.$$

Therefore, by Item 1 (Lemma B.2), there is a constant $R \geqslant 1$ depending on the covariances Σ, Γ, the bounds κ_Ψ and κ_h from Assumption H, and the moment bounds M_2 and M_2^{EK}, such that for any $j \in [\![0, J-1]\!]$, it holds that $\mathsf{QP}\mu_j, \mathsf{QP}\mu_j^{\mathrm{EK}} \in \mathcal{P}_R(\mathbb{R}^{d_u} \times \mathbb{R}^{d_y})$. In view of Items 4, 5, and 7 (Lemmas B.8, B.9 and B.11), the first term in (2.8a) satisfies

$$d_g\left(\mathsf{T}_j\mathsf{QP}\mu_j^{\mathrm{EK}}, \mathsf{T}_j\mathsf{QP}\mu_j\right) \leqslant L_{\mathsf{T}}\big(R, M_{2\cdot\max\{3+d_u,4+d_y\}}\big) L_{\mathsf{Q}} L_{\mathsf{P}} d_g\left(\mu_j^{\mathrm{EK}}, \mu_j\right), \tag{2.9}$$

where, for conciseness, we have omitted the dependence of the constants in the bound on $\kappa_y, \kappa_\Psi, \kappa_h, \Sigma, \Gamma$. By definition of R, it holds that $\mathsf{GQP}\mu_j \in \mathcal{G}_R(\mathbb{R}^{d_u} \times \mathbb{R}^{d_y})$, hence the second term in (2.8a) may be bounded using Item 7 (Lemma B.11) and the definition in of ε in (2.7). Indeed, it holds that

$$\begin{aligned} d_g\left(\mathsf{T}_j\mathsf{QP}\mu_j, \mathsf{T}_j\mathsf{GQP}\mu_j\right) &\leqslant L_{\mathsf{T}}\big(R, M_{2\cdot\max\{3+d_u,4+d_y\}}\big) d_g\left(\mathsf{QP}\mu_j, \mathsf{GQP}\mu_j\right) \\ &\leqslant L_{\mathsf{T}}\big(R, M_{2\cdot\max\{3+d_u,4+d_y\}}\big)\varepsilon. \end{aligned}$$

Using Item 6 (Lemma B.10) and the definition of ε in (2.7), we establish the following bound on the third term in (2.8a):

$$d_g\left(\mathsf{B}_j\mathsf{GQP}\mu_j, \mathsf{B}_j\mathsf{QP}\mu_j\right) \leqslant C_{\mathsf{B}} d_g\left(\mathsf{GQP}\mu_j, \mathsf{QP}\mu_j\right) \leqslant C_{\mathsf{B}}\varepsilon.$$

Therefore, setting $\ell = L_{\mathsf{T}}\big(R, M_{2\cdot\max\{3+d_u,4+d_y\}}\big) L_{\mathsf{Q}} L_{\mathsf{P}}$, we have that

$$d_g(\mu_{j+1}^{\mathrm{EK}}, \mu_{j+1}) \leqslant \ell d_g(\mu_j^{\mathrm{EK}}, \mu_j) + \big(L_{\mathsf{T}}\big(R, M_{\max\{3+d_u,4+d_y\}}\big) + C_{\mathsf{B}}\big)\varepsilon.$$

The conclusion then follows from the discrete Grönwall lemma using the fact that $\mu_0^{\mathrm{EK}} = \mu_0$. ∎

2.3. Error estimate: Mean field ensemble Kalman filter. It is possible to deduce from the result of Theorem 2.1 that the error between the true filter and the mean field ensemble Kalman filter can be arbitrarily small if the true filter is arbitrarily close to its Gaussian projection in state-observation space. This condition on the true filter, which we refer to as "closeness to Gaussian," can be satisfied in the setting of unbounded vector fields by considering small perturbations of affine vector fields because we prove in Proposition 2.2. Combining Proposition 2.2 with Theorem 2.1 gives an error estimate for the mean field ensemble Kalman filter, yielding Theorem 2.3.

PROPOSITION 2.2 (Approximation Result). *Suppose that the data* $Y^\dagger = \{y_j^\dagger\}_{j=1}^J$ *and the matrices* (Σ,Γ) *satisfy Assumptions* (H1) *and* (H2). *Fix* $\kappa_\Psi, \kappa_h > 0$ *and assume that* $\Psi_0 \colon \mathbb{R}^{d_u} \to \mathbb{R}^{d_u}$ *and* $h_0 \colon \mathbb{R}^{d_u} \to \mathbb{R}^{d_y}$ *are affine functions and, hence, satisfy* (V1) *and* (V2), *respectively, while also satisfying Assumptions* (H3) *and* (H4) *with* κ_Ψ, κ_h. *Let* $(\mu_j)_{j\in[\![0,J]\!]}$ *denote the true filtering distribution associated with functions* (Ψ, h), *initialized at the Gaussian probability measure* $\mu_0 = \mathcal{N}(m_0, C_0) \in \mathcal{G}(\mathbb{R}^{d_u})$. *Then, for any* $J \in \mathbb{Z}^+$, *there is* $C = C(m_0, C_0, \kappa_y, \kappa_\Psi, \kappa_h, \Sigma, \Gamma, J) > 0$ *such that for all* $\varepsilon \in [0,1]$ *and all* $(\Psi, h) \in B_{L^\infty}((\Psi_0, h_0), \varepsilon)$, *it holds that*

$$\max_{j\in[\![0,J-1]\!]} d_g(\mathsf{GQP}\mu_j, \mathsf{QP}\mu_j) \leqslant C\varepsilon. \tag{2.10}$$

Proof. In what follows, $(\mu_j^0)_{j\in[\![0,J]\!]}$ and $(\mu_j)_{j\in[\![0,J]\!]}$ denote the true filtering distributions associated with functions (Ψ_0, h_0) and (Ψ, h), respectively, initialized at the same Gaussian measure $\mathcal{N}(m_0, C_0)$. Furthermore, we let P_0 and Q_0 denote the kernel integral operators (1.7a) and (1.8a) defined by the specific vector fields (Ψ_0, h_0). By Lemma B.3, the filtering distributions have bounded second moments. Let

$$\mathcal{R} = \max_{j\in[\![0,J-1]\!]} \left(\left| y_{j+1}^\dagger \right|^2, 1 + \mathcal{M}_2(\mu_j^0), 1 + \mathcal{M}_2(\mu_j) \right).$$

Throughout this proof, C denotes a constant whose value is irrelevant in the context, depends only on the constants $m_0, C_0, \kappa_y, \kappa_\Psi, \kappa_h, \Sigma, \Gamma, j$ (in particular, it does not depend on ε, Ψ, h), and may change from line to line. Fix $j \in [\![0, J-1]\!]$. Note that the filtering distribution defined by (Ψ_0, h_0) is Gaussian. Then, using the triangle inequality and Gaussianity of $\mathsf{Q}_0\mathsf{P}_0\mu_j^0$, we obtain

$$\begin{aligned} d_g(\mathsf{GQP}\mu_j, \mathsf{QP}\mu_j) &\leqslant d_g(\mathsf{GQP}\mu_j, \mathsf{GQ}_0\mathsf{P}_0\mu_j^0) + d_g(\mathsf{GQ}_0\mathsf{P}_0\mu_j^0, \mathsf{QP}\mu_j) \\ &= d_g(\mathsf{GQP}\mu_j, \mathsf{GQ}_0\mathsf{P}_0\mu_j^0) + d_g(\mathsf{Q}_0\mathsf{P}_0\mu_j^0, \mathsf{QP}\mu_j). \end{aligned}$$

We note that since the filters have bounded first and second order polynomial moments, by Lemmas B.1 and B.2, we may deduce that there exists $R \geqslant 1$ such that

$$\forall j \in [\![0, J-1]\!], \qquad \mathsf{P}_0\mu_j^0, \in \mathcal{P}_R(\mathbb{R}^{d_u}), \qquad \mathsf{QP}\mu_j, \mathsf{Q}_0\mathsf{P}_0\mu_j^0 \in \mathcal{P}_R(\mathbb{R}^{d_u} \times \mathbb{R}^{d_y}). \tag{2.11}$$

The second part of this display allows direct application of Lemma C.1, which concerns the local Lipschitz continuity result for the Gaussian projection operator G, to obtain

$$\begin{aligned} d_g(\mathsf{GQP}\mu_j, \mathsf{QP}\mu_j) &\leqslant C d_g(\mathsf{Q}_0\mathsf{P}_0\mu_j^0, \mathsf{QP}\mu_j) \\ &\leqslant C d_g(\mathsf{Q}_0\mathsf{P}_0\mu_j^0, \mathsf{QP}_0\mu_j^0) + C d_g(\mathsf{QP}_0\mu_j^0, \mathsf{QP}\mu_j). \end{aligned}$$

Using (C.2) from Lemma C.2, noting that since $\mathsf{P}_0\mu^0 \in \mathcal{P}_R(\mathbb{R}^{d_u})$ by (2.11), it holds that $\mathcal{M}_2(\mathsf{P}_0\mu_j^0) \leqslant d_u R$, and using the Lipschitz continuity of Q (Lemma B.9), we deduce that

$$\begin{aligned} d_g(\mathsf{GQP}\mu_j, \mathsf{QP}\mu_j) &\leqslant C\varepsilon(1 + d_u R) + C d_g(\mathsf{P}_0\mu_j^0, \mathsf{P}\mu_j) \\ &\leqslant C\varepsilon(1 + d_u R) + C d_g\big(\mathsf{P}_0\mu_j^0, \mathsf{P}\mu_j^0\big) + C d_g\big(\mathsf{P}\mu_j^0, \mathsf{P}\mu_j\big) \\ &\leqslant C\varepsilon(1 + d_u R) + C\varepsilon\mathcal{R} + C d_g(\mu_j^0, \mu_j), \end{aligned}$$

where the second inequality follows by the triangle inequality, and the third inequality follows from bounding $d_g(\mathsf{P}_0\mu_j^0, \mathsf{P}\mu_j)$ using (C.1) from Lemma C.2 and from the Lipschitz continuity of P (Lemma B.9). The statement then follows from Lemma C.3. ∎

THEOREM 2.3 (Error Estimate: mean field ensemble Kalman filter). *Assume that the probability measures* $(\mu_j)_{j\in[\![0,J]\!]}$ *and* $(\mu_j^{\mathrm{EK}})_{j\in[\![0,J]\!]}$ *are given respectively by the dynamical systems* (1.9) *and* (2.4) *initialized at the Gaussian measure* $\mu_0 = \mu_0^{\mathrm{EK}} \in \mathcal{G}(\mathbb{R}^{d_u})$. *That is,*

$$\mu_{j+1} = \mathsf{B}_j \mathsf{QP}\mu_j, \qquad \mu_{j+1}^{\mathrm{EK}} = \mathsf{T}_j \mathsf{QP}\mu_j^{\mathrm{EK}}.$$

Suppose that the data $Y^\dagger = \{y_j^\dagger\}_{j=1}^J$ *and the matrices* (Σ,Γ) *satisfy Assumptions* (H1) *and* (H2). *Let the vector fields* Ψ_0, h_0 *be affine functions satisfying Assumptions* (V1) *and* (V2), *respectively, while also satisfying Assumptions* (H3) *and* (H4) *with some constants* $\kappa_\Psi, \kappa_h > 0$. *Additionally, let* $\ell_\Psi > 0$ *be a constant. Then, there exists a constant* $C = C(\mathcal{M}_q(\mu_0), \kappa_y, \kappa_\Psi, \kappa_h, \ell_h, \Sigma, \Gamma, J)$, *where* $q := 2\cdot \max\{3+d_u, 4+d_y\}$, *such that for all* $\varepsilon \in [0,1]$ *and all* $(\Psi, h) \in B_{L^\infty}((\Psi_0, h_0), \varepsilon)$ *with* h *satisfying Assumption* (H5), *it holds that*

$$d_g(\mu_J^{\mathrm{EK}}, \mu_J) \leqslant C\varepsilon.$$

Proof. Since Assumption H is satisfied, it is possible to apply the result of Theorem 2.1 to deduce that there exists $C = C(\mathcal{M}_{2\cdot\max\{3+d_u,4+d_y\}}(\mu_0), J, \kappa_y, \kappa_\Psi, \kappa_h, \ell_h, \Sigma, \Gamma)$ such that

$$d_g(\mu_J^{\mathrm{EK}}, \mu_J) \leqslant C \max_{j\in[\![0,J-1]\!]} d_g(\mathsf{QP}\mu_j, \mathsf{GQP}\mu_j).$$

Additionally, since $(\Psi, h) \in B_{L^\infty}((\Psi_0, h_0), \varepsilon)$ for Ψ_0, h_0 satisfying Assumptions (V1) and (V2) and moreover Assumptions (H3) and (H4), we may apply Proposition 2.2 to deduce the result. ∎

2.4. Error estimate: Finite particle ensemble Kalman filter. In this subsection, we combine the results from the work in [24] with stability Theorem 2.1, together with approximation Theorem 2.3, to derive a quantitative error estimate between the finite particle ensemble Kalman filter and the true filter in the nonlinear setting. In order to define an appropriate metric, we introduce the following class of vector fields.

Assumption P1. The vector field $\phi : \mathbb{R}^{d_u} \to \mathbb{R}$ satisfies for any $u, v \in \mathbb{R}^{d_u}$, the condition

$$|\phi(u) - \phi(v)| \leqslant L_\phi |u - v| \big(1 + |u|^\varsigma + |v|^\varsigma\big),$$

for some $\varsigma \geqslant 0$ and for some $L_\phi > 0$. We note that for any such ϕ, there exists $R_\phi > 0$ which depends on L_ϕ so that $|\phi(u)| \leqslant R_\phi(1 + |u|^{\varsigma+1})$ for any $u \in \mathbb{R}^{d_u}$.

We will prove various technical lemmas under Assumption P1; these may be useful beyond the confines of this paper. However, for our theorems, we use the more specific Assumption P2, which enables the control of first and second moments.

Assumption P2. The vector field $\phi : \mathbb{R}^{d_u} \to \mathbb{R}$ satisfies for any $u, v \in \mathbb{R}^{d_u}$ the condition

$$|\phi(u) - \phi(v)| \leqslant L_\phi |u - v| \big(1 + |u| + |v|\big)$$

for some $L_\phi > 0$. Note that for any such ϕ, there exists $R_\phi > 0$ depending on L_ϕ so that $|\phi(u)| \leqslant R_\phi(1 + |u|^2)$ for any $u \in \mathbb{R}^{d_u}$.

THEOREM 2.4 (Error Estimate: finite particle ensemble Kalman filter). *Assume that the probability measures* $(\mu_j)_{j\in[\![0,J]\!]}$ *and* $(\mu_j^{\mathrm{EK},N})_{j\in[\![0,J]\!]}$ *for finite* $J \in \mathbb{N}$ *are obtained respectively from the dynamical systems* (1.9) *and* (2.6) *initialized at the*

Gaussian probability measure $\mu_0 \in \mathcal{G}(\mathbb{R}^{d_u})$ *and at the empirical measure* $\mu_0^{\mathrm{EK},N} = \frac{1}{N}\sum_{i=1}^N \delta_{u_0^{(i)}}$ *for* $u_0^{(i)} \sim \mu_0$ *i.i.d. samples, respectively. That is,*

$$\mu_{j+1} = \mathsf{B}_j \mathsf{Q}\mathsf{P}\mu_j, \qquad \mu_{j+1}^{\mathrm{EK},N} = \frac{1}{N}\sum_{i=1}^N \delta_{u_{j+1}^{(i)}},$$

where $u_{j+1}^{(i)}$ *evolve according to the iteration in* (2.5)*. Suppose that the data* $Y^\dagger = \{y_j^\dagger\}_{j=1}^J$ *and the matrices* (Σ,Γ) *satisfy Assumptions* (H1) *and* (H2)*. Assume that the vector field* h *is linear, and let* κ_h, ℓ_h *be positive constants such that Assumptions* (H4) *and* (H5) *are satisfied. Furthermore, let the vector field* Ψ_0 *be an affine function satisfying Assumption* (V1)*, as well as Assumption* (H3)*, with* $\kappa_\Psi > 0$*. Additionally, let* $\ell_\Psi > 0$ *be a constant. Then, there exists a constant* $C = C(\mathcal{M}_q(\mu_0), R_\phi, L_\phi, \kappa_y, \kappa_\Psi, \kappa_h, \ell_h, \ell_\Psi, \Sigma, \Gamma, J)$*, where* $q := 2\cdot\max\{3+d_u, 4+d_y, 2\cdot 4^J\}$*, such that for all* $\varepsilon \in [0,1]$ *and all* $\Psi \in B_{L^\infty}(\Psi_0,\varepsilon)$ *satisfying* $|\Psi|_{C^{0,1}} \leqslant \ell_\Psi < \infty$*, the following bound holds for any* ϕ *satisfying Assumption* P2*:*

$$\left(\mathbb{E}\left|\mu_J^{\mathrm{EK},N}[\phi] - \mu_J[\phi]\right|^2\right)^{1/2} \leqslant C\left(\frac{1}{\sqrt{N}} + \varepsilon\right).$$

We note that the expectation appearing in the bound from Theorem 2.4 is with respect to the law of the particles $u_j^{(i)}$ which evolve according to (2.5) and which define the empirical measure $\mu_{j+1}^{\mathrm{EK},N}$. The proof presented hereafter relies on the following elements.

1. We apply the triangle inequality in order to employ two distinct results concerning the mean field ensemble Kalman filter. The proof thus involves quantifying the error between finite particle ensemble Kalman filter and the mean field ensemble Kalman filter, as discussed in Item 2, and the error between the mean field ensemble Kalman filter and the true filter, as discussed in Item 3.
2. The work in [24] establishes a Monte Carlo error estimate, with rate of $1/\sqrt{N}$, between the empirical measure $\mu^{\mathrm{EK},N}$, representing the particle ensemble Kalman filter, and the measure μ^{EK} describing the evolution of the mean field ensemble Kalman filter. In particular, this holds under Assumptions (H1) and (H2) on the data and covariances of the noise processes, the linearity of the vector field h, and Assumption (H3) on Ψ with the additional assumption that $|\Psi|_{C^{0,1}} \leqslant \ell_\Psi < \infty$. In Lemma D.1, we provide a self-contained proof of [24, Theorem 5.2] to gain insight into the dependence of the constant prefactor multiplying $1/\sqrt{N}$ on the parameters of the Gaussian initial condition and the number of steps J.
3. We assume that the data $Y^\dagger = \{y_j^\dagger\}_{j=1}^J$ and the matrices (Σ,Γ) satisfy Assumptions (H1) and (H2) and assume that h satisfies Assumption (V2). Furthermore, we assume Ψ satisfies Assumption (H3) and $\Psi \in B_{L^\infty}(\Psi_0,\varepsilon)$ with Ψ_0 satisfying Assumption (V1). These assumptions allow us to apply the result from Theorem 2.3.

Proof. Recall that μ_j^{EK} is the mean field ensemble Kalman filter, here initialized at the same Gaussian μ_0 as the true filter. We fix a function ϕ satisfying Assumption P2 and apply the triangle inequality to deduce that

$$\left(\mathbb{E}\left|\mu_J^{\mathrm{EK},N}[\phi] - \mu_J[\phi]\right|^2\right)^{1/2} \leqslant \left(\mathbb{E}\left|\mu_J^{\mathrm{EK},N}[\phi] - \mu_J^{\mathrm{EK}}[\phi]\right|^2\right)^{1/2} + \left(\mathbb{E}\left|\mu_J^{\mathrm{EK}}[\phi] - \mu_J[\phi]\right|^2\right)^{1/2}. \tag{2.12}$$

Since μ_J^{EK} and μ_J are deterministic probability measures, it holds that

$$\left(\mathbb{E}\left|\mu_J^{\mathrm{EK}}[\phi]-\mu_J[\phi]\right|^2\right)^{1/2}=\left|\mu_J^{\mathrm{EK}}[\phi]-\mu_J[\phi]\right|.$$

Since ϕ satisfies Assumption P2, it follows that

$$\left|\mu_J^{\mathrm{EK}}[\phi]-\mu_J[\phi]\right|\leqslant\sup_{|\phi|\leqslant R_\phi(1+|u|^2)}\left|\mu_J^{\mathrm{EK}}[\phi]-\mu_J[\phi]\right|=R_\phi\cdot d_g\big(\mu_J^{\mathrm{EK}},\mu_J\big).$$

We note that Assumption H holds as h is assumed to be affine; since we additionally assume that $\Psi\in B_{L^\infty}(\Psi_0,\varepsilon)$, where Ψ_0 is an affine function, we may apply the result of Theorem 2.3 to deduce that

$$d_g\big(\mu_J^{\mathrm{EK}},\mu_J\big)\leqslant C\varepsilon, \tag{2.13}$$

where C is a constant depending on $\mathcal{M}_{2\cdot\max\{3+d_u,4+d_y\}}(\mu_0),J,\kappa_y,\kappa_\Psi,\kappa_h,\ell_h,\Sigma,\Gamma$.

The fact that h is assumed to be linear and the additional assumption that $|\Psi|_{C^{0,1}}\leqslant\ell_\Psi<\infty$ allows us to apply the result of Lemma D.1, so that

$$\left(\mathbb{E}\left|\mu_J^{\mathrm{EK},N}[\phi]-\mu_J^{\mathrm{EK}}[\phi]\right|^2\right)^{1/2}\leqslant\frac{C}{\sqrt{N}}, \tag{2.14}$$

where C is a constant depending on $\mathcal{M}_{4^{J+1}}(\mu_0),J,R_\phi,L_\phi,\kappa_y,\kappa_\Psi,\kappa_h,\Sigma,\Gamma$. In view of (2.12), combining (2.13) and (2.14) yields the desired result. ∎

3. Discussion and future directions. In this paper, we have presented the first analysis in the setting of a filtering problem defined by nonlinear state space dynamics, quantifying the error between the empirical measure obtained by the finite particle discrete time ensemble Kalman filter and the true filtering distribution. The analysis for the EnKF outlined is based on the proof methodology developed for particle filters in [26] and developed for the mean field ensemble Kalman filter in [9]. Our work goes beyond [9] by considering finite particle filters and by allowing for dynamical models and observation operators that are unbounded. The work builds substantially on the new avenue for analysis of the ensemble Kalman methodology that was introduced in [9] but leaves open numerous avenues of investigation for further work; we detail these.

1. As surveyed in [8], the ensemble Kalman filtering methodology may be used for solving inverse problems and for sampling; it is of interest to extend the analysis in our paper to the ensemble Kalman based inversion algorithms outlined in that paper.
2. There is a substantial body of literature that studied the continuous time limits of ensemble Kalman methods; see [8] for a review. Performing analysis analogous to that presented here, but in the continuous time setting, would be of interest.
3. For large scale applications, there has been recent wide interest in replacing the dynamical model Ψ, representing the solution operator obtained via a high fidelity numerical solver, with a cheap to evaluate surrogate. Multifidelity ensemble methods [4] allow the use of a small number of particles evolved according to the high fidelity solver and a large number evolved according to the surrogate. Extending our error analysis to incorporate the effect of model error would be of interest in this context.

4. We highlight that there may be other "near-Gaussian problems" arising in the small noise or large data volume limits that would be interesting to study using Bernstein-von-Mises theorems; however, this will require new analysis that makes observability assumptions and controls behavior of the constants in the analysis with respect to noise and number of observations.
5. We have derived results in a conditional setting arising from assuming a uniform bound on the set of considered observations; removing this assumption and understanding the probabilistic nature of error bounds with respect to observational noise would be of interest.
6. Our error estimate comparing the EnKF with the true filter blows up exponentially over time. In the presence of stability assumptions about the true filter, it is natural to ask whether our results can be extended to the infinite time horizon.

Appendix A. Auxiliary results. We first state and prove a standard result that will be used throughout the paper.

LEMMA A.1. *Let X be a random vector in $\mathbb{R}^{d_u}$ with finite second moment. Then, it holds that*

$$\mathbb{E}\Big[\big(X - \mathbb{E}[X]\big)\big(X - \mathbb{E}[X]\big)^{\mathsf{T}}\Big] = \mathbb{E}\big[XX^{\mathsf{T}}\big] - \mathbb{E}[X]\mathbb{E}[X]^{\mathsf{T}} \tag{A.1}$$

and

$$\forall a \in \mathbb{R}^{d_u}, \qquad \mathbb{E}\Big[\big(X - a\big)\big(X - a\big)^{\mathsf{T}}\Big] \succcurlyeq \mathbb{E}\Big[\big(X - \mathbb{E}[X]\big)\big(X - \mathbb{E}[X]\big)^{\mathsf{T}}\Big]. \tag{A.2}$$

Proof. The statements follow from the following equalities:

$$\begin{aligned}\mathbb{E}\Big[(X-a)(X-a)^{\mathsf{T}}\Big] &= \mathbb{E}\Big[\Big(\big(X - \mathbb{E}[X]\big) + \big(\mathbb{E}[X] - a\big)\Big)\Big(\big(X - \mathbb{E}[X]\big) + \big(\mathbb{E}[X] - a\big)\Big)^{\mathsf{T}}\Big] \\ &= \mathbb{E}\Big[\big(X - \mathbb{E}[X]\big)\big(X - \mathbb{E}[X]\big)^{\mathsf{T}}\Big] + \big(\mathbb{E}[X] - a\big)\big(\mathbb{E}[X] - a\big)^{\mathsf{T}}. \qquad \square\end{aligned}$$

LEMMA A.2. *Let $g(\bullet; m, S)$ be the Lebesgue density of $\mathcal{N}(m,S)$ and $\mathcal{S}^n_\alpha$ the set of symmetric $n \times n$ matrices M satisfying*

$$\frac{1}{\tau} I_n \preccurlyeq M \preccurlyeq \tau I_n.$$

Then, for any $n \in \mathbb{N}^+$ and $\tau \geqslant 1$, there is $L_{n,\tau} > 0$ such that for all parameters $(c_1, m_1, S_1) \in \mathbb{R} \times \mathbb{R}^n \times \mathcal{S}^n_\tau$ and $(c_2, m_2, S_2) \in \mathbb{R} \times \mathbb{R}^n \times \mathcal{S}^n_\tau$, it holds that

$$\|h\|_\infty \leqslant L_{n,\tau} \|h\|_1, \qquad h(y) = c_1 g(y; m_1, S_1) - c_2 g(y; m_2, S_2).$$

Proof. The lemma as stated may be found in [9, Lemma A.4], where a complete proof is given. $\square$

LEMMA A.3. *Let P and Q denote the operators on probability measures given respectively in* (1.7a) *and* (1.8a). *Let Assumptions* (H2) *to* (H5) *be satisfied and suppose that $\mu \in \mathcal{P}(\mathbb{R}^{d_u})$ satisfies, for some $q > 0$, the moment bound*

$$\mathcal{M}_q(\mu) = \int_{\mathbb{R}^{d_u}} |x|^q \, \mu(\mathrm{d}x) < \infty. \tag{A.3}$$

Then, there is $L = L(\mathcal{M}_q(\mu), \kappa_\Psi, \kappa_h, \ell_h, \Sigma, \Gamma) > 0$ such that for all $(u_1, u_2, y) \in \mathbb{R}^{d_u} \times \mathbb{R}^{d_u} \times \mathbb{R}^{d_y}$, the probability density of $p = \mathsf{QP}\mu$ satisfies

$$\big|p(u_1, y) - p(u_2, y)\big| \leqslant L|u_1 - u_2| \min\left\{\max\left\{\frac{1}{1+|u_1|^q}, \frac{1}{1+|u_2|^q}\right\}, \frac{1}{1+|y|^q}\right\}. \tag{A.4}$$

Proof. Throughout this proof, C denotes a constant whose value is irrelevant in the context, depends only on $\mathcal{M}_q(\mu), \kappa_\Psi, \kappa_h, \ell_h, \Sigma, \Gamma$, and may change from line to line. Sometimes, we write the dependence explicitly to indicate which parameters are involved.

Step 1. *Bounding the density of* $\mathsf{P}\mu$. We first rewrite

$$\big(1+|u|^q\big)\mathsf{P}\mu(u) = C(\Sigma)\int_{\mathbb{R}^{d_u}} \exp\left(-\frac{1}{2}|u-\Psi(v)|_\Sigma^2\right)\frac{1+|u|^q}{1+|v|^q}\big(1+|v|^q\big)\mu(\mathrm{d}v).$$

We note that

$$S(\Sigma,\kappa_\Psi) := \sup_{(u,v)\in\mathbb{R}^{d_u}\times\mathbb{R}^{d_u}} \exp\left(-\frac{1}{2}\left|u-\Psi(v)\right|_\Sigma^2\right)\frac{1+|u|^q}{1+|v|^q} < \infty;$$

this may be seen observing that

$$\begin{aligned}\text{(A.5)}\qquad &\exp\left(-\frac{1}{2}\left|u-\Psi(v)\right|_\Sigma^2\right)\frac{1+|u|^q}{1+|v|^q}\\ &\leqslant \exp\left(-\frac{1}{2}\left|u-\Psi(v)\right|_\Sigma^2\right)\frac{(1+2^{q-1}|u-\Psi(v)|^q)+2^{q-1}|\Psi(v)|^q}{1+|v|^q}.\end{aligned}$$

The first term on the right-hand side of (A.5) may be bounded by noting that $1+2^{q-1}e^{-x^2}x^q$ is uniformly bounded in $x\in\mathbb{R}$; on the other hand, the second term may be bounded by applying Assumption (H3). Now, using (A.3), we obtain that

$$\text{(A.6)}\qquad \mathsf{P}\mu(u) \leqslant \frac{C\big(\Sigma,\kappa_\Psi,\mathcal{M}_q(\mu)\big)}{1+|u|^q}.$$

Step 2. *Establishing Lipschitz continuity of* $u\mapsto\mathsf{P}\mu(u)$. Since $g(x) := \mathrm{e}^{-x^2}$ has derivative $2x\,\mathrm{e}^{-x^2}$ and $|x\,\mathrm{e}^{-x^2}| \leqslant \mathrm{e}^{-\frac{x^2}{2}}$ for all $x\in\mathbb{R}$, it holds for some ξ between $|a|$ and $|b|$ that

$$\text{(A.7)}\qquad \forall(a,b)\in\mathbb{R}\times\mathbb{R},\qquad \left|\mathrm{e}^{-a^2}-\mathrm{e}^{-b^2}\right| = |b-a|\,|g'(\xi)| \leqslant 2|b-a|\left(\mathrm{e}^{-\frac{a^2}{2}}+\mathrm{e}^{-\frac{b^2}{2}}\right).$$

Using this inequality with $a^2 = \frac{1}{2}|u_1-\Psi(v)|_\Sigma^2$ and $b^2 = \frac{1}{2}|u_2-\Psi(v)|_\Sigma^2$, the triangle inequality, and equivalence of norms, we deduce that, for all $(u_1,u_2,v)\in\mathbb{R}^{d_u}\times\mathbb{R}^{d_u}\times\mathbb{R}^{d_u}$,

$$\begin{aligned}&\left|\exp\left(-\frac{1}{2}\left|u_1-\Psi(v)\right|_\Sigma^2\right)-\exp\left(-\frac{1}{2}\left|u_2-\Psi(v)\right|_\Sigma^2\right)\right|\\ &\quad\leqslant C|u_2-u_1|\left(\exp\left(-\frac{1}{4}\left|u_1-\Psi(v)\right|_\Sigma^2\right)+\exp\left(-\frac{1}{4}\left|u_2-\Psi(v)\right|_\Sigma^2\right)\right)\end{aligned}$$

for constant $C=C(\Sigma)$. Integrating out the v variable with respect to μ, we obtain that

$$\begin{aligned}|\mathsf{P}\mu(u_1)-\mathsf{P}\mu(u_2)| \leqslant{}& C|u_1-u_2|\int_{\mathbb{R}^{d_u}}\exp\left(-\frac{1}{4}\left|u_1-\Psi(v)\right|_\Sigma^2\right)\mu(\mathrm{d}v)\\ &+ C|u_1-u_2|\int_{\mathbb{R}^{d_u}}\exp\left(-\frac{1}{4}\left|u_2-\Psi(v)\right|_\Sigma^2\right)\mu(\mathrm{d}v).\end{aligned}$$

The integrals on the right-hand side can be bounded as in the first step, which leads to the inequality

$$\forall (u_1,u_2)\in\mathbb{R}^{d_u}\times\mathbb{R}^{d_u},\quad |\mathsf{P}\mu(u_1)-\mathsf{P}\mu(u_2)|\leqslant C|u_1-u_2|\max\left\{\frac{1}{1+|u_1|^q},\frac{1}{1+|u_2|^q}\right\}. \tag{A.8}$$

Step 3. *Obtaining a coarse estimate.* In view of the elementary inequality (A.7) and the assumed Lipschitz continuity of h, it holds that, for constant $C=C(\Gamma,\ell_h)$,

$$\begin{aligned}\Big|\mathcal{N}\big(h(u_1),\Gamma\big)(y)-\mathcal{N}\big(h(u_2),\Gamma\big)(y)\Big| \\ \leqslant C|u_1-u_2|\exp\left(-\frac{1}{4}|y-h(u_1)|_\Gamma^2\right)+C|u_1-u_2|\exp\left(-\frac{1}{4}|y-h(u_2)|_\Gamma^2\right).\end{aligned} \tag{A.9}$$

Using the decomposition

$$\begin{aligned}p(u_1,y)-p(u_2,y)&=\mathsf{P}\mu(u_1)\mathcal{N}\big(h(u_1),\Gamma\big)(y)-\mathsf{P}\mu(u_2)\mathcal{N}\big(h(u_2),\Gamma\big)(y)\\&=\big(\mathsf{P}\mu(u_1)-\mathsf{P}\mu(u_2)\big)\mathcal{N}\big(h(u_1),\Gamma\big)(y)\\&\quad+\mathsf{P}\mu(u_2)\Big(\mathcal{N}\big(h(u_1),\Gamma\big)(y)-\mathcal{N}\big(h(u_2),\Gamma\big)(y)\Big).\end{aligned}$$

and employing (A.6), (A.8), and (A.9), we deduce that

$$\begin{aligned}|p(u_1,y)-p(u_2,y)|\leqslant C|u_1-u_2|\max\left\{\frac{1}{1+|u_1|^q},\frac{1}{1+|u_2|^q}\right\}\\ \times\max\left\{\exp\left(-\frac{1}{4}|y-h(u_1)|_\Gamma^2\right),\exp\left(-\frac{1}{4}|y-h(u_2)|_\Gamma^2\right)\right\}.\end{aligned} \tag{A.10}$$

Note that, for the function $h(u)=u$, the quantity multiplying $|u_1-u_2|$ on the right-hand side does not tend to 0 along the sequence $(u_1^{(n)},u_2^{(n)},y^{(n)})=(0,n,n)$. For our purposes in this work, we need the finer estimate (A.4); establishing this bound is the aim of the next two steps.

Step 4. *Bounding the density* $p(u,y)$. Recall that $p(u,y)=\mathsf{P}\mu(u)\mathcal{N}\big(h(u),\Gamma\big)(y)$. We prove in this step the inequality

$$p(u,y)\leqslant C\min\left\{\frac{1}{1+|u|^q},\frac{1}{1+|y|^q}\right\}, \tag{A.11}$$

or equivalently

$$\sup_{(u,y)\in\mathbb{R}^{d_u}\times\mathbb{R}^{d_y}} p(u,y)\max\{1+|u|^q,1+|y|^q\}<\infty.$$

To this end, note that

$$\begin{aligned}p(u,y)\max\{|u|^q,|y|^q\}&=\mathsf{P}\mu(u)\mathcal{N}\big(h(u),\Gamma\big)(y)\max\{|u|^q,|y|^q\}\\&\leqslant\mathsf{P}\mu(u)\mathcal{N}\big(h(u),\Gamma\big)(y)|u|^q+\mathsf{P}\mu(u)\mathcal{N}\big(h(u),\Gamma\big)(y)|y|^q.\end{aligned}$$

The first term is bounded uniformly by **Step 1**, estimate (A.6). For the second term, we use that

$$|y|^q\leqslant 2^{q-1}\big|h(u)\big|^q+2^{q-1}\big|y-h(u)\big|^q \tag{A.12}$$

to obtain

$$\mathsf{P}\mu(u)\mathcal{N}\big(h(u),\Gamma\big)(y)|y|^q \leqslant C\mathsf{P}\mu(u)\mathcal{N}\big(h(u),\Gamma\big)(y)\Big(|h(u)|^q + |y-h(u)|^q\Big).$$

The first term on the right-hand side is bounded uniformly, again by **Step 1**, estimate (A.6), and using Assumption (H4). The second term is also bounded uniformly because the function $x \mapsto \mathcal{N}(0,\Gamma)(x)|x|^q$ is uniformly bounded in x, by a value depending only on Γ and q.

Step 5. *Obtaining the estimate* (A.4). The claimed inequality is equivalent to

$$\sup_{u_1,u_2,y} \frac{\big|p(u_1,y)-p(u_2,y)\big|}{|u_1-u_2|} \max\Big\{\min\{1+|u_1|^q, 1+|u_2|^q\}, 1+|y|^q\Big\} < \infty,$$

where the supremum is over $\mathbb{R}^{d_u} \times \mathbb{R}^{d_u} \times \mathbb{R}^{d_y}$. By (A.10), it holds that

$$\sup_{u_1,u_2,y} \frac{\big|p(u_1,y)-p(u_2,y)\big|}{|u_1-u_2|} \min\{1+|u_1|^q, 1+|u_2|^q\} < \infty,$$

so it remains to show that

$$\sup_{u_1,u_2,y} \frac{\big|p(u_1,y)-p(u_2,y)\big|}{|u_1-u_2|} |y|^q < \infty.$$

By (A.11), it is clear that the supremum is uniformly bounded if restricted to the set $|u_1-u_2| \geqslant 1$, so it suffices to show that

$$\sup_{(u,\delta,y)\in\mathbb{R}^{d_u}\times B(0,1)\times\mathbb{R}^{d_y}} \frac{\big|p(u,y)-p(u+\delta,y)\big|}{|\delta|} |y|^q < \infty,$$

where $B(0,1)$ is the open ball of radius 1 centered at the origin in $\mathbb{R}^{d_u}$. We use again (A.12) in order to bound

$$\begin{aligned}\frac{\big|p(u,y)-p(u+\delta,y)\big|}{|\delta|}|y|^q \leqslant C\frac{\big|p(u,y)-p(u+\delta,y)\big|}{|\delta|}\big|h(u)\big|^q \\ + C\frac{\big|p(u,y)-p(u+\delta,y)\big|}{|\delta|}|y-h(u)|^q.\end{aligned}$$

The first term is bounded uniformly by (A.10) and the assumption that $\big|h(u)\big| \leqslant \kappa_h(1+|u|)$. To conclude the proof, it remains to show that

$$\sup_{(u,\delta,y)\in\mathbb{R}^{d_u}\times B(0,1)\times\mathbb{R}^{d_y}} C\frac{\big|p(u,y)-p(u+\delta,y)\big|}{|\delta|}|y-h(u)|^q < \infty. \tag{A.13}$$

By (A.10), it holds that

$$\begin{aligned}&\big|p(u,y)-p(u+\delta,y)\big| \\ &\quad\leqslant \frac{C|\delta|}{1+|u|^q}\max\left\{\exp\left(-\frac{1}{4}|y-h(u)|_\Gamma^2\right), \exp\left(-\frac{1}{4}|y-h(u+\delta)|_\Gamma^2\right)\right\}. \\ &\quad\leqslant \frac{C|\delta|}{1+|u|^q}\exp\left(-\frac{1}{8}|y-h(u)|_\Gamma^2 + \frac{1}{4}\,|h(u)-h(u+\delta)|_\Gamma^2\right).\end{aligned}$$

In the last line, we used the inequality $|a+b|^2 \geqslant \frac{1}{2}|a|^2 - |b|^2$, which follows from Young's inequality. Since the function $x \mapsto \mathrm{e}^{-\frac{x^2}{8}+C} x^q$ is bounded uniformly in $x \in \mathbb{R}$ and by Assumption (H5), the bound (A.13) easily follows, concluding the proof. ∎

Appendix B. Technical Results for Theorem 2.1. We establish moment bounds in Appendix B.1, we recall that on Gaussian measures the action of the conditioning map and the Kalman transport map are equivalent in Appendix B.2, and we prove stability results in Appendix B.3.

B.1. Moment Bounds.

LEMMA B.1 (Moment Bounds). *Let μ be a probability measure on $\mathbb{R}^{d_u}$ with bounded first and second order polynomial moments $\mathcal{M}_1(\mu), \mathcal{M}_2(\mu) < \infty$. Given Assumptions* (H2) *and* (H3)*, it holds that*

$$\text{(B.1)} \qquad \left|\mathcal{M}(\mathsf{P}\mu)\right| \leqslant \kappa_\Psi\bigl(1+\mathcal{M}_1(\mu)\bigr), \qquad \Sigma \preccurlyeq \mathcal{C}(\mathsf{P}\mu) \preccurlyeq \Sigma + 2\kappa_\Psi^2\bigl(1+\mathcal{M}_2(\mu)\bigr) I_{d_u}.$$

Proof. The proof of this lemma follows the steps of the proof of [9, Lemma B.1]; however, here, different bounds reflecting the assumptions on Ψ and h in this paper are required. Using the definition of P in (1.7a), it holds that

$$\begin{aligned}\mathcal{M}(\mathsf{P}\mu) = \int_{\mathbb{R}^{d_u}} u\,\mathsf{P}\mu(u)\,\mathrm{d}u &= \frac{1}{\sqrt{(2\pi)^{d_u}\det\Sigma}} \int_{\mathbb{R}^{d_u}}\int_{\mathbb{R}^{d_u}} u \exp\left(-\frac{1}{2}\left|u-\Psi(v)\right|_\Sigma^2\right)\mu(\mathrm{d}v)\,\mathrm{d}u \\ &= \int_{\mathbb{R}^{d_u}} \Psi(v)\,\mu(\mathrm{d}v),\end{aligned}$$

where application of Fubini's theorem yields the last equality. The first inequality in (B.1) then follows from Assumption (H3). Obtaining the lower bound of the second inequality in (B.1) may be done identically to the proof of [9, Lemma B.1]. We turn our attention to obtaining the upper bound. Using Lemma A.1 and noting that $ww^\top \preccurlyeq (w^\top w) I_{d_u}$ for any vector $w \in \mathbb{R}^{d_u}$ by the Cauchy–Schwarz inequality, we deduce that

$$\text{(B.2)} \qquad \begin{aligned}\mathcal{C}(\mathsf{P}\mu) &\preccurlyeq \int_{\mathbb{R}^{d_u}} u\otimes u\,\mathsf{P}\mu(u)\,\mathrm{d}u \\ &= \frac{1}{\sqrt{(2\pi)^{d_u}\det\Sigma}} \int_{\mathbb{R}^{d_u}}\int_{\mathbb{R}^{d_u}} u\otimes u \exp\left(-\frac{1}{2}\left|u-\Psi(v)\right|_\Sigma^2\right)\mu(\mathrm{d}v)\,\mathrm{d}u \\ &= \int_{\mathbb{R}^{d_u}} \bigl(\Psi(v)\otimes\Psi(v)+\Sigma\bigr)\,\mu(\mathrm{d}v) \\ &\preccurlyeq \Sigma + \left(\int_{\mathbb{R}^{d_u}} |\Psi(v)|^2\mu(\mathrm{d}v)\right) I_{d_u} \preccurlyeq \Sigma + 2\kappa_\Psi^2\bigl(1+\mathcal{M}_2(\mu)\bigr) I_{d_u},\end{aligned}$$

which yields the desired result. ∎

LEMMA B.2. *Let μ be a probability measure on $\mathbb{R}^{d_u}$ with bounded first and second order polynomial moments $\mathcal{M}_1(\mu), \mathcal{M}_2(\mu) < \infty$. Given Assumptions* (H2) *to* (H4)*, it holds that*

$$\text{(B.3a)} \qquad \left|\mathcal{M}^u(\mathsf{QP}\mu)\right| \leqslant \kappa_\Psi\bigl(1+\mathcal{M}_1(\mu)\bigr),$$

$$\text{(B.3b)} \qquad \left|\mathcal{M}^y(\mathsf{QP}\mu)\right| \leqslant \kappa_h\sqrt{2\Bigl(1+\operatorname{tr}(\Sigma)+2\kappa_\Psi^2\bigl(1+\mathcal{M}_2(\mu)\bigr)\Bigr)}.$$

Furthermore, it holds that

(B.4a)

$$\mathcal{C}(\mathsf{QP}\mu) \preccurlyeq \begin{pmatrix} 4\kappa_\Psi^2\big(1+\mathcal{M}_2(\mu)\big)I_{d_u} + 2\Sigma & 0_{d_u\times d_y} \\ 0_{d_y\times d_u} & 4\kappa_h^2\big(1+\operatorname{tr}(\Sigma)+2\kappa_\Psi^2\big(1+\mathcal{M}_2(\mu)\big)\big)I_{d_y}+\Gamma \end{pmatrix},$$

(B.4b)

$$\mathcal{C}(\mathsf{QP}\mu) \succcurlyeq \frac{\gamma\cdot\min\Big\{2\sigma, \gamma+4\kappa_h^2\Big(1+\operatorname{tr}(\Sigma)+2\kappa_\Psi^2\big(1+\mathcal{M}_2(\mu)\big)\Big)\Big\}}{2\gamma+8\kappa_h^2\Big(1+\operatorname{tr}(\Sigma)+2\kappa_\Psi^2\big(1+\mathcal{M}_2(\mu)\big)\Big)} I_{d_u+d_y}.$$

Proof. The inequalities (B.3a) and (B.3b) follow from the assumptions and the fact that

$$\mathcal{M}(\mathsf{QP}\mu) = \begin{pmatrix} \mathcal{M}(\mathsf{P}\mu) \\ \mathsf{P}\mu[h] \end{pmatrix}. \tag{B.5}$$

Indeed, from Lemma B.1, we know that $\big|\mathcal{M}(\mathsf{P}\mu)\big| \leqslant \kappa_\Psi\big(1+\mathcal{M}_1(\mu)\big)$, which leads by Jensen's inequality to (B.3a). To deduce (B.3b), we note that

$$\begin{aligned} \big|P\mu[h]\big|^2 &\leqslant \int_{\mathbb{R}^{d_u}} |h(u)|^2\,\mathsf{P}\mu(\mathrm{d}u) \\ &\leqslant 2\kappa_h^2 + 2\kappa_h^2 \int_{\mathbb{R}^{d_u}} |u|^2\,\mathsf{P}\mu(\mathrm{d}u) \qquad \text{(B.6a)} \\ &= 2\kappa_h^2 + 2\kappa_h^2 \int_{\mathbb{R}^{d_u}} \Big(|\Psi(v)|^2 + \operatorname{tr}(\Sigma)\Big)\,\mu(\mathrm{d}v), \qquad \text{(B.6b)} \end{aligned}$$

where (B.6a) follows by applying Assumption (H4) and Young's inequality, and (B.6b) follows by applying the properties of the Gaussian transition density. The result then follows by applying Assumption (H3).

To obtain the covariance bounds, we proceed using analogous steps to the proof of [9, Lemma B.2]; however, here, different bounds reflecting the assumptions on Ψ and h in this paper are required. We begin by establishing inequality (B.4a). To this end, letting $\phi\colon \mathbb{R}^{d_u} \to \mathbb{R}^{d_u+d_y}$ define the map $\phi(u) = (u, h(u))$, it holds that

$$\mathcal{C}(\mathsf{QP}\mu) = \mathcal{C}(\phi_\sharp\mathsf{P}\mu) + \begin{pmatrix} 0_{d_u\times d_u} & 0_{d_u\times d_y} \\ 0_{d_y\times d_u} & \Gamma \end{pmatrix} = \begin{pmatrix} \mathcal{C}^{uu}(\phi_\sharp\mathsf{P}\mu) & \mathcal{C}^{uy}(\phi_\sharp\mathsf{P}\mu) \\ \mathcal{C}^{yu}(\phi_\sharp\mathsf{P}\mu) & \mathcal{C}^{yy}(\phi_\sharp\mathsf{P}\mu)+\Gamma \end{pmatrix}. \tag{B.7}$$

As shown in the proof of [9, Lemma B.2] we have, for all $(a,b) \in \mathbb{R}^{d_u}\times\mathbb{R}^{d_y}$,

$$\begin{pmatrix} a \\ b \end{pmatrix} \otimes \begin{pmatrix} a \\ b \end{pmatrix} \preccurlyeq 2\begin{pmatrix} aa^\mathsf{T} & 0_{d_u\times d_y} \\ 0_{d_y\times d_u} & bb^\mathsf{T} \end{pmatrix},$$

from which we establish, using Lemmas A.1 and B.1 and Assumptions (H2) to (H4), that

(B.8)

$$\begin{aligned} \mathcal{C}(\phi_\sharp\mathsf{P}\mu) &\preccurlyeq \int_{\mathbb{R}^{d_u}} \begin{pmatrix} u \\ h(u) \end{pmatrix} \otimes \begin{pmatrix} u \\ h(u) \end{pmatrix} \mathsf{P}\mu(u)\,\mathrm{d}u \\ &\preccurlyeq 2\int_{\mathbb{R}^{d_u}} \begin{pmatrix} uu^\mathsf{T} & 0_{d_u\times d_y} \\ 0_{d_y\times d_u} & h(u)h(u)^\mathsf{T} \end{pmatrix} \mathsf{P}\mu(u)\,\mathrm{d}u \\ &\preccurlyeq 2\begin{pmatrix} \Sigma + 2\kappa_\Psi^2\big(1+\mathcal{M}_2(\mu)\big)I_{d_u} & 0_{d_u\times d_y} \\ 0_{d_y\times d_u} & 2\kappa_h^2\big(1+\operatorname{tr}(\Sigma)+2\kappa_\Psi^2\big(1+\mathcal{M}_2(\mu)\big)\big)I_{d_y} \end{pmatrix}, \end{aligned}$$

where we applied (B.2) and the calculation resulting in (B.6a) in the last inequality. Noting this inequality in combination with (B.7) yields the upper bound (B.4a). We now show that

$$\text{(B.9)} \qquad \begin{pmatrix} a \\ b \end{pmatrix}^{\mathsf{T}} \mathcal{C}(\mathsf{QP}\mu) \begin{pmatrix} a \\ b \end{pmatrix} \geqslant (1-\varepsilon)\sigma\, |a|^2 + \left(\gamma - \left(\frac{1}{\varepsilon} - 1\right) \mathsf{P}\mu\big[|h|^2\big]\right) |b|^2 .$$

in order to establish the lower bound (B.4b). The argument is similar to that in [9, Lemma B.2] but is explicit in certain constants whose dependencies on moments need to be controlled in this paper. We first note that for any $\pi \in \mathcal{P}(\mathbb{R}^{d_u} \times \mathbb{R}^{d_y})$ and all $(a,b) \in \mathbb{R}^{d_u} \times \mathbb{R}^{d_y}$, it holds that

$$\begin{aligned} \left|a^{\mathsf{T}} \mathcal{C}^{uy}(\pi) b\right| &= \int_{\mathbb{R}^{d_u} \times \mathbb{R}^{d_y}} \Big(a^{\mathsf{T}}\big(u - \mathcal{M}^u(\pi)\big)\Big)\Big(b^{\mathsf{T}}\big(y - \mathcal{M}^y(\pi)\big)\Big)\pi(\mathrm{d}u\mathrm{d}y) \\ &\leqslant \sqrt{a^{\mathsf{T}} \mathcal{C}^{uu}(\pi) a}\,\sqrt{b^{\mathsf{T}} \mathcal{C}^{yy}(\pi) b} \end{aligned}$$

by the Cauchy–Schwarz inequality. Hence, we deduce that for all $\varepsilon \in (0,1)$ and for all $(a,b) \in \mathbb{R}^{d_u} \times \mathbb{R}^{d_y}$,

$$\begin{aligned} \begin{pmatrix} a \\ b \end{pmatrix}^{\mathsf{T}} \mathcal{C}(\phi_\sharp \mathsf{P}\mu) \begin{pmatrix} a \\ b \end{pmatrix} &\geqslant (1-\varepsilon) a^{\mathsf{T}} \mathcal{C}^{uu}(\phi_\sharp \mathsf{P}\mu) a - \left(\frac{1}{\varepsilon} - 1\right) b^{\mathsf{T}} \mathcal{C}^{yy}(\phi_\sharp \mathsf{P}\mu) b \\ &\geqslant (1-\varepsilon) a^{\mathsf{T}} \Sigma a - \left(\frac{1}{\varepsilon} - 1\right) \mathsf{P}\mu\big[|h|^2\big], \end{aligned}$$

where we applied Young's inequality for the bound in the first line and (B.1) and the bound $\mathcal{C}^{yy}(\phi_\sharp \mathsf{P}\mu) \preccurlyeq \mathsf{P}\mu\big[|h|^2\big] I_{d_y}$ for the bound in the second line. We then apply (B.7) to find

$$\begin{aligned} \begin{pmatrix} a \\ b \end{pmatrix}^{\mathsf{T}} \mathcal{C}(\mathsf{QP}\mu) \begin{pmatrix} a \\ b \end{pmatrix} &\geqslant (1-\varepsilon) a^{\mathsf{T}} \Sigma a - \left(\frac{1}{\varepsilon} - 1\right) \mathsf{P}\mu\big[|h|^2\big] + b^{\mathsf{T}} \Gamma b \\ &\geqslant (1-\varepsilon)\sigma\, |a|^2 + \left(\gamma - \left(\frac{1}{\varepsilon} - 1\right) \mathsf{P}\mu\big[|h|^2\big]\right) |b|^2. \end{aligned}$$

Choosing $\varepsilon = \frac{2\mathsf{P}\mu[|h|^2]}{\gamma + 2\mathsf{P}\mu[|h|^2]}$, so that the coefficient of the $|b|^2$ term is $\frac{\gamma}{2}$, we obtain

$$\text{(B.10)} \qquad \begin{aligned} \begin{pmatrix} a \\ b \end{pmatrix}^{\mathsf{T}} \mathcal{C}(\mathsf{QP}\mu) \begin{pmatrix} a \\ b \end{pmatrix} &\geqslant \frac{\gamma\sigma}{\gamma + 2\mathsf{P}\mu\big[|h|^2\big]} |a|^2 + \frac{\gamma}{2} |b|^2 \\ &\geqslant \min\left\{\frac{\gamma\sigma}{\gamma + 2\mathsf{P}\mu\big[|h|^2\big]}, \frac{\gamma}{2}\right\} \left(|a|^2 + |b|^2\right). \end{aligned}$$

Since this is true for any $(a,b) \in \mathbb{R}^{d_u} \times \mathbb{R}^{d_y}$, it follows that

$$\mathcal{C}(\mathsf{QP}\mu) \succcurlyeq \gamma \frac{\min\left\{2\sigma, \gamma + 2\mathsf{P}\mu\big[|h|^2\big]\right\}}{2\gamma + 4\mathsf{P}\mu\big[|h|^2\big]} I_{d_u + d_y}.$$

In order to deduce (B.4b), we use (B.6b) to find that

$$\mathsf{P}\mu\big[|h|^2\big] \leqslant 2\kappa_h^2 \Big(1 + \operatorname{tr}(\Sigma) + 2\kappa_\Psi^2 \big(1 + \mathcal{M}_2(\mu)\big)\Big),$$

from which we conclude the desired result. ∎

Lemma B.3 (Moment Bound for the True Filtering Distribution). *Let $q \geqslant 2$ be an integer. Assume that the probability measures $(\mu_j)_{j\in[\![0,J]\!]}$ are obtained from the dynamical system* (1.9) *initialized at the probability measure $\mu_0 \in \mathcal{P}(\mathbb{R}^{d_u})$ with bounded qth order polynomial moment $\mathcal{M}_q(\mu_0) < \infty$. If Assumptions* (H1) *to* (H4) *hold, then there exists $C = C(\mathcal{M}_q(\mu_0), J, \kappa_y, \kappa_\Psi, \kappa_h, \Sigma, \Gamma)$ such that*

$$\max_{j\in[\![0,J]\!]} \mathcal{M}_q(\mu_j) \leqslant C.$$

Proof. We have $\mu_{j+1} = \mathsf{B}_j\mathsf{Q}\mathsf{P}\mu_j$, for $j \in [\![0, J-1]\!]$. Equivalently, using the notation from (1.6), it holds that $\mu_{j+1} = \mathsf{L}_j\mathsf{P}\mu_j$ for each $j \in [\![0, J-1]\!]$. Hence, we note that

$$\mu_{j+1} = \frac{\exp\left(-\frac{1}{2}\left|y^\dagger_{j+1} - h(u)\right|^2_\Gamma\right)\mathsf{P}\mu_j(u)}{\int_{\mathbb{R}^{d_u}} \exp\left(-\frac{1}{2}\left|y^\dagger_{j+1} - h(U)\right|^2_\Gamma\right)\mathsf{P}\mu_j(U)\,\mathrm{d}U}. \tag{B.11}$$

It is readily observed that

$$\begin{aligned}\mathcal{M}_q(\mu_{j+1}) &= \frac{\int_{\mathbb{R}^{d_u}} |u|^q \exp\left(-\frac{1}{2}\left|y^\dagger_{j+1} - h(u)\right|^2_\Gamma\right)\mathsf{P}\mu_j(u)\mathrm{d}u}{\int_{\mathbb{R}^{d_u}} \exp\left(-\frac{1}{2}\left|y^\dagger_{j+1} - h(U)\right|^2_\Gamma\right)\mathsf{P}\mu_j(U)\,\mathrm{d}U} \\ &\leqslant \frac{\int_{\mathbb{R}^{d_u}} |u|^q\,\mathsf{P}\mu_j(u)\mathrm{d}u}{\int_{\mathbb{R}^{d_u}} \exp\left(-\frac{1}{2}\left|y^\dagger_{j+1} - h(U)\right|^2_\Gamma\right)\mathsf{P}\mu_j(U)\,\mathrm{d}U}.\end{aligned} \tag{B.12}$$

We first bound from above the numerator of (B.12); indeed, note that

$$\begin{aligned}\int_{\mathbb{R}^{d_u}} |u|^q\mathsf{P}\mu_j(u)\mathrm{d}u &= \int_{\mathbb{R}^{d_u}} |u|^q \left(\int_{\mathbb{R}^{d_u}} \exp\left(-\frac{1}{2}\left|u - \Psi(v)\right|^2_\Sigma\right)\mu_j(\mathrm{d}v)\right)\mathrm{d}u \\ &= \int_{\mathbb{R}^{d_u}} \left(\int_{\mathbb{R}^{d_u}} |u|^q \exp\left(-\frac{1}{2}\left|u - \Psi(v)\right|^2_\Sigma\right)\mathrm{d}u\right)\mu_j(\mathrm{d}v) && \text{(B.13a)} \\ &\leqslant C\int_{\mathbb{R}^{d_u}} \bigl(1 + |\Psi(v)|^q\bigr)\mu_j(\mathrm{d}v) && \text{(B.13b)} \\ &\leqslant C\Bigl(1 + \mathcal{M}_q(\mu_j)\Bigr), && \text{(B.13c)}\end{aligned}$$

where (B.13a) follows from Fubini's theorem, the inequality (B.13b) from properties of Gaussians and Assumption (H2), and (B.13c) from application of Assumption (H3). We note that in (B.13c), the constant C depends on κ_Ψ, Σ. Now, to obtain a lower bound on the denominator of (B.12), we observe that

$$\begin{aligned}\int_{\mathbb{R}^{d_u}} \exp\left(-\frac{1}{2}\left|y^\dagger_{j+1} - h(u)\right|^2_\Gamma\right)\mathsf{P}\mu_j(u)\,\mathrm{d}u &\geqslant \exp\left(-\frac{1}{2}\int_{\mathbb{R}^{d_u}} \left|y^\dagger_{j+1} - h(u)\right|^2_\Gamma\mathsf{P}\mu_j(u)\,\mathrm{d}u\right) \\ &\geqslant C\exp\left(-\left\|\Gamma^{-1}\right\|\int_{\mathbb{R}^{d_u}} |h(u)|^2\mathsf{P}\mu_j(u)\,\mathrm{d}u\right) \\ &\geqslant C\exp\Bigl(-4\kappa_h^2\kappa_\Psi^2\left\|\Gamma^{-1}\right\|\mathcal{M}_2(\mu_j)\Bigr),\end{aligned} \tag{B.14}$$

where the first inequality follows by application of Jensen's inequality, and (B.14) follows from the calculation leading to (B.6b). We note that the C in (B.14) is a constant depending on $\kappa_y, \kappa_\Psi, \kappa_h, \Sigma, \Gamma$. Therefore, by combining (B.14) with (B.13c), it is possible to deduce from (B.12) that

$$\mathcal{M}_q(\mu_{j+1}) \leqslant C \exp\Big(4\kappa_h^2 \kappa_\Psi^2 \left\|\Gamma^{-1}\right\| \mathcal{M}_2(\mu_j)\Big)\Big(1 + \mathcal{M}_q(\mu_j)\Big), \tag{B.15}$$

where C is a constant depending on $\kappa_h, \kappa_\Psi, \kappa_y, \Sigma, \Gamma$. ∎

Remark B.4.

- In some situations, the bound (B.15) is overly pessimistic. For example, if h satisfies Assumption (H4) as well as the inequality $\left|h(u) - y^\dagger_{j+1}\right| \geqslant c_\ell(|u| - 1)$ for all $u \in \mathbb{R}^{d_u}$ for some positive c_ℓ, then by [17, Proposition A.3] for all $q > 0$, there is $C = C(c_\ell, q)$ such that

$$\forall \mu \in \mathcal{P}(\mathbb{R}^{d_u}), \qquad \mathsf{L}_j\mu\Big[|u|^q\Big] \leqslant C\mu\Big[|u|^q\Big]. \tag{B.16}$$

  In this setting, better control of the moments can be achieved than in (B.15).
- In the absence of any additional assumption on h beyond Assumption (H4), there may not exist a finite constant $C > 0$ such that the bound (B.16) holds for an arbitrary probability measure μ. To illustrate this, consider the case where $d_u = 2$, $d_y = 1$, $\Gamma = \frac{1}{4}$, $y^\dagger_{j+1} = 0$, and

$$h(u) = u_1, \qquad u = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}, \qquad \mu_R = \frac{x_R^2 \delta_{(R,0)} + \delta_{(0,x_R)}}{x_R^2 + 1}, \qquad x_R = \mathrm{e}^{\frac{R^2}{2}}.$$

  Then, as $R \to \infty$, $\mu_R[|u|^2] \sim R^2$, while

$$\begin{aligned} \mathsf{L}_j\mu_R\left[|u|^2\right] &= \frac{\int_{\mathbb{R}^2} (u_1^2 + u_2^2) \exp\left(-2u_1^2\right)\, \mu_R(\mathrm{d}u_1 \mathrm{d}u_2)}{\int_{\mathbb{R}^2} \exp\left(-2u_1^2\right)\, \mu_R(\mathrm{d}u_1 \mathrm{d}u_2)} \\ &= \frac{x_R^2 \,\mathrm{e}^{-2R^2}\, R^2 + x_R^2}{x_R^2 \,\mathrm{e}^{-2R^2} + 1} \sim x_R^2 = \mathrm{e}^{R^2}. \end{aligned}$$

  Thus, it holds that

$$\lim_{R \to \infty} \frac{\mathsf{L}_j\mu_R\left[|u|^2\right]}{\mu_R\left[|u|^2\right]} = +\infty,$$

  which shows that (B.16) fails to hold in this case.

Lemma B.5 (Moment Bound for the Approximate Filtering Distribution). *Let $q \geqslant 2$ be an integer. Assume that the probability measures $(\mu_j^{\mathrm{EK}})_{j \in [\![0,J]\!]}$ are obtained from the dynamical system* (2.4) *initialized at the probability measure $\mu_0^{\mathrm{EK}} \in \mathcal{P}(\mathbb{R}^{d_u})$ with bounded qth polynomial order moment $\mathcal{M}_q(\mu_0^{\mathrm{EK}}) < \infty$. If Assumptions* (H1) *to* (H4) *hold, then there exists $C = C(\mathcal{M}_q(\mu_0^{\mathrm{EK}}), J, \kappa_y, \kappa_\Psi, \kappa_h, \Sigma, \Gamma)$ such that*

$$\max_{j \in [\![0,J]\!]} \mathcal{M}_q\big(\mu_j^{\mathrm{EK}}\big) \leqslant C.$$

Proof. We begin by noting that $\mu_{j+1}^{\mathrm{EK}} = \mathsf{T}_j \mathsf{QP} \mu_j^{\mathrm{EK}}$, for $j \in [\![0, J-1]\!]$. Thus,

(B.17)

$$\begin{aligned}
\mathcal{M}_q\big(\mu_{j+1}^{\mathrm{EK}}\big) &= \int_{\mathbb{R}^{d_u}} |u|^q \mathsf{T}_j \mathsf{QP}\mu_j^{\mathrm{EK}}(\mathrm{d}u) \\
&= \int_{\mathbb{R}^{d_u}\times\mathbb{R}^{d_y}} \left|\mathscr{T}(u, y; \mathsf{QP}\mu_j^{\mathrm{EK}}, y_{j+1}^{\dagger})\right|^q \mathsf{QP}\mu_j^{\mathrm{EK}}(u, y)\,\mathrm{d}y\,\mathrm{d}u \\
&= \int_{\mathbb{R}^{d_u}\times\mathbb{R}^{d_y}} \left|u + \mathcal{C}^{uy}(\mathsf{QP}\mu_j^{\mathrm{EK}})\mathcal{C}^{yy}(\mathsf{QP}\mu_j^{\mathrm{EK}})^{-1}\big(y_{j+1}^{\dagger} - y\big)\right|^q \mathsf{QP}\mu_j^{\mathrm{EK}}(u, y)\,\mathrm{d}y\,\mathrm{d}u \\
&\leqslant C \int_{\mathbb{R}^{d_u}\times\mathbb{R}^{d_y}} |u|^q \,\mathsf{QP}\mu_j^{\mathrm{EK}}(u, y)\,\mathrm{d}y\,\mathrm{d}u \\
&\quad + C\Big(1 + \mathcal{M}_2\big(\mu_j^{\mathrm{EK}}\big)\Big)^{2q} \cdot \left(1 + \int_{\mathbb{R}^{d_u}\times\mathbb{R}^{d_y}} |y|^q \,\mathsf{QP}\mu_j^{\mathrm{EK}}(u, y)\,\mathrm{d}y\,\mathrm{d}u\right),
\end{aligned}$$

where in (B.17), the constant C depends on $\kappa_y, \kappa_\Psi, \kappa_h, \Sigma, \Gamma$ and where the dependence on the second moment of μ_j^{EK} in the second term is derived from (B.4a) and (B.4b). By using the definitions of Q and P and by applying reasoning analogous to (B.13c), we deduce that

$$\begin{aligned}
\mathcal{M}_q\big(\mu_{j+1}^{\mathrm{EK}}\big) &\leqslant C\Big(1 + \mathcal{M}_2\big(\mu_j^{\mathrm{EK}}\big)\Big)^{2q} \cdot \Big(1 + \mathcal{M}_q\big(\mu_j^{\mathrm{EK}}\big)\Big), \\
&\leqslant C\Big(1 + \mathcal{M}_q\big(\mu_j^{\mathrm{EK}}\big)\Big)^{2q+1}
\end{aligned}$$

where C is a constant depending on $\kappa_y, \kappa_\Psi, \kappa_h, \Sigma, \Gamma$. Iteration gives the desired result. ∎

LEMMA B.6. *Let $\mu_1, \mu_2 \in \mathcal{P}(\mathbb{R}^n)$ have finite second moments. Then, the following bounds hold:*

$$\begin{aligned}
\big|\mathcal{M}(\mu_1) - \mathcal{M}(\mu_2)\big| &\leqslant \frac{1}{2} d_g(\mu_1, \mu_2), \\
\big\|\mathcal{C}(\mu_1) - \mathcal{C}(\mu_2)\big\| &\leqslant \left(1 + \frac{1}{2}|\mathcal{M}(\mu_1) + \mathcal{M}(\mu_2)|\right) d_g(\mu_1, \mu_2).
\end{aligned}$$

Proof. The lemma as stated may be found in [9, Lemma B.4], where a complete proof is given. ∎

B.2. Action of T_j on Gaussians.

LEMMA B.7 ($\mathsf{B}_j\mathsf{G} = \mathsf{T}_j\mathsf{G}$). *Fix $y_{j+1}^{\dagger} \in \mathbb{R}^{d_y}$ and let π be a Gaussian measure over $\mathbb{R}^{d_u} \times \mathbb{R}^{d_y}$. Then, the probability measure $\mathsf{B}_j\pi$, with B_j defined in* (1.8b) *is equivalent to the probability measure $\mathsf{T}_j\pi = \mathscr{T}(\bullet, \bullet; \pi, y_{j+1}^{\dagger})_\sharp \pi$, where $\mathscr{T}$ is defined in* (2.3b).

Proof. The lemma as stated may be found in [9, Lemma B.5], where a complete proof is given. ∎

B.3. Stability Results.

LEMMA B.8 (Map P is Lischitz). *Suppose that Σ and Ψ satisfy Assumptions* (H2) *and* (H3), *respectively. Then, it holds that*

$$\forall (\mu, \nu) \in \mathcal{P}(\mathbb{R}^{d_u}) \times \mathcal{P}(\mathbb{R}^{d_u}), \qquad d_g(\mathsf{P}\mu, \mathsf{P}\nu) \leqslant \Big(1 + 2\kappa_\Psi^2 + \mathrm{tr}(\Sigma)\Big)\, d_g(\mu, \nu).$$

Proof. By definition of P, it holds that

$$\mathsf{P}\mu(\mathrm{d}u) = \int_{\mathbb{R}^{d_u}} p(v,\mathrm{d}u)\,\mu(\mathrm{d}v), \qquad \text{where } p(v,\mathrm{d}u) := \frac{\exp\left(-\frac{1}{2}\left|u-\Psi(v)\right|_\Sigma^2\right)}{\sqrt{(2\pi)^{d_u}\det\Sigma}}\,\mathrm{d}u.$$

Let $g(u) = 1 + |u|^2$, as in Definition 1.1, and take any function $f\colon \mathbb{R}^{d_u} \to \mathbb{R}$ such that $|f| \leqslant g$. Assumption (H3) implies that

$$\begin{aligned}\left|\int_{\mathbb{R}^{d_u}} f(u)\,p(v,\mathrm{d}u)\right| &\leqslant \int_{\mathbb{R}^{d_u}} g(u)\,p(v,\mathrm{d}u) = 1 + |\Psi(v)|^2 + \mathrm{tr}(\Sigma)\\ &\leqslant 1 + 2\kappa_\Psi^2 + 2\kappa_\Psi^2|v|^2 + \mathrm{tr}(\Sigma) \leqslant \Big(1 + 2\kappa_\Psi^2 + \mathrm{tr}(\Sigma)\Big)g(v).\end{aligned}$$

By Fubini's theorem, it follows that

$$\begin{aligned}\Big|\mathsf{P}\mu[f] - \mathsf{P}\nu[f]\Big| &= \left|\int_{\mathbb{R}^{d_u}}\left(\int_{\mathbb{R}^{d_u}} f(u)\,p(v,\mathrm{d}u)\right)\big(\mu(\mathrm{d}v) - \nu(\mathrm{d}v)\big)\right|\\ &= \Big(1 + 2\kappa_\Psi^2 + \mathrm{tr}(\Sigma)\Big)\Big|\mu[g] - \nu[g]\Big| \leqslant \Big(1 + 2\kappa_\Psi^2 + \mathrm{tr}(\Sigma)\Big)\,d_g(\mu,\nu),\end{aligned}$$

thus concluding the proof. ∎

LEMMA B.9 (Map Q is Lipschitz). *Suppose that* Γ *and* h *satisfy Assumptions* (H2) *and* (H4), *respectively. Then it holds that*

$$\text{(B.18)}\qquad \forall(\mu,\nu) \in \mathcal{P}(\mathbb{R}^{d_u}) \times \mathcal{P}(\mathbb{R}^{d_u}), \qquad d_g(\mathsf{Q}\mu, \mathsf{Q}\nu) \leqslant \Big(1 + 2\kappa_h^2 + \mathrm{tr}(\Gamma)\Big)d_g(\mu,\nu).$$

Proof. Take $f\colon \mathbb{R}^{d_u} \times \mathbb{R}^{d_y} \to \mathbb{R}$ satisfying $|f| \leqslant g$, where $g(u,v) = 1 + |u|^2 + |y|^2$. By Fubini's theorem, it holds that

$$\begin{aligned}\mathsf{Q}\mu[f] - \mathsf{Q}\nu[f] &= \int_{\mathbb{R}^{d_u}} \Pi f(u)\,\Big(\mu(\mathrm{d}u) - \nu(\mathrm{d}u)\Big),\\ \text{where } \Pi f(u) &:= \int_{\mathbb{R}^{d_y}} f(u,y)\,\mathcal{N}\big(h(u),\Gamma\big)(\mathrm{d}y).\end{aligned}$$

The function $\Pi f\colon \mathbb{R}^{d_u} \to \mathbb{R}$ satisfies

$$\begin{aligned}\forall u \in \mathbb{R}^{d_u}, \qquad |\Pi f(u)| &\leqslant \int_{\mathbb{R}^{d_y}} \big|f(u,y)\big|\,\mathcal{N}\big(h(u),\Gamma\big)(\mathrm{d}y)\\ &\leqslant \int_{\mathbb{R}^{d_y}} \left(1 + |u|^2 + |y|^2\right)\,\mathcal{N}\big(h(u),\Gamma\big)(\mathrm{d}y)\\ &= 1 + |u|^2 + \big|h(u)\big|^2 + \mathrm{tr}(\Gamma) \leqslant \Big(1 + 2\kappa_h^2 + \mathrm{tr}(\Gamma)\Big)\left(1 + |u|^2\right).\end{aligned}$$

Therefore, we deduce that

$$\Big|\mathsf{Q}\mu[f] - \mathsf{Q}\nu[f]\Big| \leqslant \Big(1 + 2\kappa_h^2 + \mathrm{tr}(\Gamma)\Big)\,d_g(\mu,\nu),$$

which concludes the proof. ∎

LEMMA B.10 (Stability of Map B_j). *Suppose that Assumptions* (H1) *to* (H4) *are satisfied. Then, for any probability measure* $\mu \in \mathcal{P}(\mathbb{R}^{d_u})$ *with* $\mathcal{M}_2(\mu) < \infty$, *there exists a constant* $C_\mathsf{B} = C_\mathsf{B}\big(\mathcal{M}_2(\mu), \kappa_y, \kappa_\Psi, \kappa_h, \Sigma, \Gamma\big) > 0$ *such that*

$$\forall j \in [\![0, J]\!], \qquad d_g(\mathsf{B}_j\mathsf{GQP}\mu, \mathsf{B}_j\mathsf{QP}\mu) \leqslant C_\mathsf{B} d_g(\mathsf{GQP}\mu, \mathsf{QP}\mu).$$

Proof. The proof of this lemma follows the steps of the proof of [9, Lemma B.9]; however, each step requires different bounds reflecting the assumptions on Ψ and h in this paper. For ease of exposition, we write $y^\dagger = y^\dagger_{j+1}$. We define the y-marginal densities

$$\alpha_\mu(y) := \int_{\mathbb{R}^{d_u}} \mathsf{GQP}\mu(u,y)\,\mathrm{d}u\,, \qquad \beta_\mu(y) := \int_{\mathbb{R}^{d_u}} \mathsf{QP}\mu(u,y)\,\mathrm{d}u\,.$$

By definition of B_j, it holds that

$$\begin{aligned}(\text{B.19})\quad d_g(\mathsf{B}_j\mathsf{GQP}\mu, \mathsf{B}_j\mathsf{QP}\mu) &= \int_{\mathbb{R}^{d_u}} \left(1+|u|^2\right)\left|\frac{\mathsf{GQP}\mu(u,y^\dagger)}{\alpha_\mu(y^\dagger)} - \frac{\mathsf{QP}\mu(u,y^\dagger)}{\beta_\mu(y^\dagger)}\right|\mathrm{d}u \\ &\leqslant \frac{1}{\alpha_\mu(y^\dagger)}\int_{\mathbb{R}^{d_u}} \left(1+|u|^2\right)\left|\mathsf{GQP}\mu(u,y^\dagger) - \mathsf{QP}\mu(u,y^\dagger)\right|\mathrm{d}u \\ &\quad + \left|\frac{\alpha_\mu(y^\dagger)-\beta_\mu(y^\dagger)}{\alpha_\mu(y^\dagger)\beta_\mu(y^\dagger)}\right|\int_{\mathbb{R}^{d_u}}\left(1+|u|^2\right)\mathsf{QP}\mu(u,y^\dagger)\,\mathrm{d}u.\end{aligned}$$

Step 1: bounding $\alpha_\mu(y^\dagger)$ and $\beta_\mu(y^\dagger)$ from below. The distribution $\alpha_\mu(\bullet)$ is Gaussian with mean $\mathcal{M}^y(\mathsf{QP}\mu)$ and covariance

$$(\text{B.20})\qquad \mathcal{C}^{yy}(\mathsf{QP}\mu) = \Gamma + \mathsf{P}\mu\big[h\otimes h\big] - \mathsf{P}\mu\big[h\big]\otimes\mathsf{P}\mu\big[h\big].$$

Clearly, $\mathcal{C}^{yy}(\mathsf{QP}\mu) \succcurlyeq \Gamma$. Furthermore, it holds that $\mathcal{C}^{yy}(\mathsf{QP}\mu) \preccurlyeq \Gamma + 2\kappa_h^2\big(1+\mathrm{tr}(\Sigma) + 2\kappa_\Psi^2\big(1+\mathcal{M}_2(\mu)\big)\big)I_{d_y}$ by (B.8). Therefore, noting that (B.3b) implies that $|\mathcal{M}^y(\mathsf{QP}\mu)| \leqslant C$ for a constant C depending only on the parameters $\mathcal{M}_2(\mu), \kappa_h, \kappa_\Psi, \Sigma$, we obtain

$$\begin{aligned}(\text{B.21})\qquad \alpha_\mu(y) &= \frac{1}{\sqrt{(2\pi)^{d_u}\det\big(\mathcal{C}^{yy}(\mathsf{QP}\mu)\big)}} \\ &\quad\times\exp\left(-\frac{1}{2}\big(y-\mathcal{M}^y(\mathsf{QP}\mu)\big)^\mathsf{T}\mathcal{C}^{yy}(\mathsf{QP}\mu)^{-1}\big(y-\mathcal{M}^y(\mathsf{QP}\mu)\big)\right) \\ &\geqslant \frac{\exp\Big(-\frac{1}{2}\big(|y|+C\big)^2\left\|\Gamma^{-1}\right\|\Big)}{\sqrt{(2\pi)^{d_y}\det\big(\Gamma + 2\kappa_h^2\big(1+\mathrm{tr}(\Sigma)+2\kappa_\Psi^2\big(1+\mathcal{M}_2(\mu)\big)\big)I_{d_y}\big)}}.\end{aligned}$$

The function β_μ can be bounded from below using similar reasoning. Indeed, by Assumptions (H2) to (H4), we have that for all $y\in\mathbb{R}^{d_y}$,

$$\begin{aligned}\beta_\mu(y) &= \int_{\mathbb{R}^{d_u}}\mathsf{QP}\mu(u,y)\,\mathrm{d}u = \int_{\mathbb{R}^{d_u}}\frac{\exp\Big(-\frac{1}{2}\big(y-h(u)\big)^\mathsf{T}\Gamma^{-1}\big(y-h(u)\big)\Big)}{\sqrt{(2\pi)^{d_y}\det(\Gamma)}}\mathsf{P}\mu(u)\,\mathrm{d}u \\ (\text{B.22a})\quad &\geqslant \frac{1}{\sqrt{(2\pi)^{d_y}\det(\Gamma)}}\exp\left(-\left\|\Gamma^{-1}\right\||y|^2 - \left\|\Gamma^{-1}\right\|\kappa_h^2\int_{\mathbb{R}^{d_u}}\big(2+|u|^2\big)\mathsf{P}\mu(u)\,\mathrm{d}u\right) \\ (\text{B.22b})\quad &\geqslant C\exp\Big(-\left\|\Gamma^{-1}\right\||y|^2\Big),\end{aligned}$$

where we applied Jensen's inequality in (B.22a), and the constant in (B.22b) depends on $\mathcal{M}_2(\mu), \kappa_\Psi, \kappa_h, \Sigma, \Gamma$.

Step 2: bounding the first term in (B.19). First note that, for fixed $u\in\mathbb{R}^{d_u}$, the functions $\mathsf{QP}\mu(u,\bullet)$ and $\mathsf{GQP}\mu(u,\bullet)$ are Gaussians up to constant factors, with covariance matrices given respectively by Γ and

$$(\text{B.23})\qquad \mathcal{C}^{yy}(\mathsf{QP}\mu) - \mathcal{C}^{yu}(\mathsf{QP}\mu)\mathcal{C}^{uu}(\mathsf{QP}\mu)^{-1}\mathcal{C}^{uy}(\mathsf{QP}\mu),$$

where we have used the formula for the covariance of the conditional distribution of a Gaussian. We note that $\mathcal{C}^{yu}(\mathsf{QP}\mu)\mathcal{C}^{uu}(\mathsf{QP}\mu)^{-1}\mathcal{C}^{uy}(\mathsf{QP}\mu)$ is positive semidefinite; hence, by (B.20) and its upper bound, it follows that the matrix (B.23) is bounded from above by $2\kappa_h^2\big(1+\operatorname{tr}(\Sigma)+2\kappa_\Psi^2\big(1+\mathcal{M}_2(\mu)\big)\big)I_{d_y}+\Gamma$. As shown in the proof of [9, Lemma B.9], the matrix (B.23) is bounded from below by Γ (see [9, (B.20)]). It follows from using Lemma A.2 in Appendix A, with parameter $\tau=\tau(\mathcal{M}_2(\mu),\kappa_h,\kappa_\Psi,\Sigma,\Gamma)$, that

$$\int_{\mathbb{R}^{d_u}}\big(1+|u|^2\big)\,|\mathsf{GQP}\mu(u,y)-\mathsf{QP}\mu(u,y)|\,\mathrm{d}u\leqslant C d_g(\mathsf{GQP}\mu,\mathsf{QP}\mu), \tag{B.24}$$

where C is a constant depending on $\mathcal{M}_2(\mu),\kappa_h,\kappa_\Psi,\Sigma,\Gamma$. We refer to [9, (B.21)] for the detailed steps used to establish this bound.

Step 3: bounding the second term in (B.19). In view of (B.24), it holds that

$$\big|\alpha_\mu(y)-\beta_\mu(y)\big|\leqslant\int_{\mathbb{R}^{d_u}}\big(1+|u|^2\big)\,\big|\mathsf{GQP}\mu(u,y)-\mathsf{QP}\mu(u,y)\big|\,\mathrm{d}u\leqslant C d_g(\mathsf{GQP}\mu,\mathsf{QP}\mu).$$

Now, since $\mathsf{QP}\mu(u,\bullet)/\mathsf{P}\mu(u)$ defines a Gaussian density which has covariance Γ and is bounded uniformly from above by $\big((2\pi)^{d_y}\det(\Gamma)\big)^{-1/2}$, we also have that

$$\int_{\mathbb{R}^{d_u}}\big(1+|u|^2\big)\,\mathsf{QP}\mu(u,y)\,\mathrm{d}u\leqslant\int_{\mathbb{R}^{d_u}}\frac{\big(1+|u|^2\big)\,\mathsf{P}\mu(u)}{\sqrt{(2\pi)^{d_y}\det(\Gamma)}}\,\mathrm{d}u=\frac{1+\operatorname{tr}\big(\mathcal{C}(\mathsf{P}\mu)\big)+|\mathcal{M}(\mathsf{P}\mu)|^2}{\sqrt{(2\pi)^{d_y}\det(\Gamma)}}.$$

By the moment bounds in Lemma B.1, the right-hand side is bounded from above by a constant which depends on $\mathcal{M}_2(\mu),\kappa_\Psi,\Sigma$.

Step 4: concluding the proof. Putting together the above bounds, we conclude that

$$d_g(\mathsf{B}_j\mathsf{GQP}\mu,\mathsf{B}_j\mathsf{QP}\mu)\leqslant\frac{C\big(\mathcal{M}_2(\mu),\kappa_\Psi,\kappa_h,\Sigma,\Gamma\big)}{\alpha_\mu(y^\dagger_{j+1})}\left(1+\frac{1}{\beta_\mu(y^\dagger_{j+1})}\right)d_g(\mathsf{GQP}\mu,\mathsf{QP}\mu).$$

Applying the inequalities (B.21) and (B.22b) yields the desired result. ∎

LEMMA B.11 (Stability of Map T_j). *Suppose that Assumption* H *is satisfied. Then, for all $R\geqslant 1$, it holds that for any $\pi\in\mathcal{P}_R(\mathbb{R}^{d_u}\times\mathbb{R}^{d_y})$ and $p\in\{\mathsf{QP}\mu:\mu\in\mathcal{P}(\mathbb{R}^{d_u})$ and $\mathcal{M}_{2\cdot\max\{3+d_u,4+d_y\}}(\mu)<\infty\}\subset\mathcal{P}(\mathbb{R}^{d_u}\times\mathbb{R}^{d_y})$, there is $L_{\mathsf{T}}:=L_{\mathsf{T}}(R,\mathcal{M}_{2\cdot\max\{3+d_u,4+d_y\}}(\mu),\kappa_y,\kappa_\Psi,\kappa_h,\ell_h,\Sigma,\Gamma)$, such that*

$$\forall j\in[\![1,J]\!],\qquad d_g(\mathsf{T}_j\pi,\mathsf{T}_jp)\leqslant L_{\mathsf{T}}\,d_g(\pi,p).$$

Proof. By the results of Lemma B.2, it holds that $\mathsf{QP}\mu\in\mathcal{P}_{\widetilde{R}}(\mathbb{R}^{d_u}\times\mathbb{R}^{d_y})$ for some $\widetilde{R}(\mathcal{M}_2(\mu),\kappa_\Psi,\kappa_h,\Sigma,\Gamma)\geqslant 1$. Using analogous notation to the one found in the proof of [9, Lemma B.10] we define

$$r=\max\Big\{R,\widetilde{R},\kappa_y\Big\}.$$

Letting $K=\mathcal{C}(\pi)$, $S=\mathcal{C}(p)$ and $y^\dagger=y^\dagger_{j+1}$, we also define the affine maps $\mathscr{T}^\pi$ and $\mathscr{T}^p$ corresponding to the use of covariance information at the probability measures π and $p=\mathsf{QP}\mu$,

$$\mathscr{T}^\pi(u,y)=u+A_\pi(y^\dagger-y),\qquad A_\pi:=K_{uy}K_{yy}^{-1},$$
$$\mathscr{T}^p(u,y)=u+A_p(y^\dagger-y),\qquad A_p:=S_{uy}S_{yy}^{-1}.$$

By a straightforward application of the triangle inequality, it holds that

$$d_g(\mathsf{T}_j\pi, \mathsf{T}_j p) \leqslant d_g(\mathscr{T}^\pi_\sharp \pi, \mathscr{T}^\pi_\sharp p) + d_g(\mathscr{T}^\pi_\sharp p, \mathscr{T}^p_\sharp p). \tag{B.25}$$

In the following steps, we will separately bound the two terms on the right-hand side of (B.25). Before proceeding, we outline two auxiliary bounds that will be used in the rest of the proof. These bounds are identical to the ones found in the proof of [9, Lemma B.10]; we include statements here for expository purposes. Noting that the operator 2-norm of any submatrix is bounded from above by the operator 2-norm of the full matrix, and we observe (see [9, (B.23)]) that

$$\|A_\pi\| \leqslant \|K_{uy}\|\|K_{yy}^{-1}\| \leqslant \|K\|\|K^{-1}\| \leqslant r^4. \tag{B.26}$$

Note that the above bound similarly holds for A_p. Using (B.26) and assuming without loss of generality that $r > 1$, we deduce (see [9, (B.24)]) that

$$\begin{aligned}\|A_\pi - A_p\| &\leqslant \|(K_{uy} - S_{uy})K_{yy}^{-1}\| + \|S_{uy}\left(K_{yy}^{-1} - S_{yy}^{-1}\right)\| \\ &\leqslant 2r^6\|K - S\| \leqslant 2r^6(1+2r)\, d_g(\pi, p),\end{aligned} \tag{B.27}$$

where the second inequality follows again from fact that the 2-norm of any submatrix is bounded from above by the 2-norm of the full matrix, while the last inequality follows from the result in Lemma B.6.

Bounding the first term in (B.25). With an identical argument to the one used in the proof of [9, Lemma B.10], we have that

$$d_g(\mathscr{T}^\pi_\sharp \pi, \mathscr{T}^\pi_\sharp p) \leqslant 3\left(1 + r^{10}\right) r^8\, d_g(\pi, p). \tag{B.28}$$

Bounding the second term in (B.25). Let f again satisfy $|f| \leqslant g$. We note that

$$\left|\mathscr{T}^\pi_\sharp p[f] - \mathscr{T}^p_\sharp p[f]\right| = \left|p[f\circ \mathscr{T}^\pi] - p[f \circ \mathscr{T}^p]\right| = \left|p[f\circ\mathscr{T}^\pi - f\circ\mathscr{T}^p]\right|.$$

The last term may be expressed as

$$\begin{aligned}&\left|p[f\circ\mathscr{T}^\pi - f\circ\mathscr{T}^p]\right| \\ &\quad= \left|\int_{\mathbb{R}^{d_y}}\int_{\mathbb{R}^{d_u}} \Big(f\left(u + A_\pi(y^\dagger - y)\right) - f\left(u + A_p(y^\dagger - y)\right)\Big)\, p(u,y)\,\mathrm{d}u\,\mathrm{d}y\right| \\ &\quad= \int_{\mathbb{R}^{d_y}}\int_{\mathbb{R}^{d_u}} \left(f(u + A_\pi z) - f(u + A_p z)\right) p(u, y^\dagger - z)\,\mathrm{d}u\,\mathrm{d}z \\ &\quad= \int_{\mathbb{R}^{d_y}}\int_{\mathbb{R}^{d_u}} f(v)\Big(p(v - A_\pi z, y^\dagger - z) - p(v - A_p z, y^\dagger - z)\Big)\,\mathrm{d}v\,\mathrm{d}z,\end{aligned} \tag{B.29}$$

where we used a change of variables in the second equality. Since $\mathcal{M}_{2\cdot\max\{3+d_u, 4+d_y\}}(\mu) < \infty$ by assumption, it follows from Lemma A.3 and from the inequality $\min\{a,b\} \leqslant \sqrt{a}\sqrt{b}$ that there is a constant C so that

$$\begin{aligned}&\left|p(v - A_\pi z, y^\dagger - z) - p(v - A_p z, y^\dagger - z)\right| \\ &\quad\leqslant C|A_\pi z - A_p z| \cdot \max\left\{\frac{1}{1 + |v - A_\pi z|^{\max\{3+d_u, 4+d_y\}}}, \frac{1}{1 + |v - A_p z|^{\max\{3+d_u, 4+d_y\}}}\right\} \\ &\qquad\times \frac{1}{1 + |y^\dagger - z|^{\max\{3+d_u, 4+d_y\}}}.\end{aligned}$$

We apply this inequality to bound for fixed $z \in \mathbb{R}^{d_y}$ the inner integral in (B.29). Considering only the terms that depend on v gives

(B.30)
$$\begin{aligned}
&\int_{\mathbb{R}^{d_u}} |f(v)| \max\left\{\frac{1}{1+|v-A_\pi z|^{\max\{3+d_u,4+d_y\}}}, \frac{1}{1+|v-A_p z|^{\max\{3+d_u,4+d_y\}}}\right\} \mathrm{d}v \\
&\quad\leqslant \int_{\mathbb{R}^{d_u}} \frac{|f(v)|}{1+|v-A_\pi z|^{\max\{3+d_u,4+d_y\}}} \mathrm{d}v + \int_{\mathbb{R}^{d_u}} \frac{|f(v)|}{1+|v-A_p z|^{\max\{3+d_u,4+d_y\}}} \mathrm{d}v \\
&\quad\leqslant \int_{\mathbb{R}^{d_u}} \frac{1+|v|^2}{1+|v-A_\pi z|^{\max\{3+d_u,4+d_y\}}} \mathrm{d}v + \int_{\mathbb{R}^{d_u}} \frac{1+|v|^2}{1+|v-A_p z|^{\max\{3+d_u,4+d_y\}}} \mathrm{d}v \\
&\quad\leqslant \int_{\mathbb{R}^{d_u}} \frac{1+|w+A_\pi z|^2}{1+|w|^{\max\{3+d_u,4+d_y\}}} \mathrm{d}w + \int_{\mathbb{R}^{d_u}} \frac{1+|w+A_p z|^2}{1+|w|^{\max\{3+d_u,4+d_y\}}} \mathrm{d}w,
\end{aligned}$$

where we used that $|f(v)| \leqslant 1+|v|^2$, as well as a change of variable in the last line. It follows that the integral in (B.30) is bounded from above by

$$C\Big(1+|A_\pi z|^2+|A_p z|^2\Big) \leqslant Cr^8\left(1+|z|^2\right),$$

where the inequality follows from (B.26). Finally, the resulting integral in the z variable can be bounded analogously, which gives

$$\begin{aligned}
d_g(\mathscr{T}^\pi_\sharp p, \mathscr{T}^p_\sharp p) &\leqslant Cr^8 \int_{\mathbb{R}^{d_y}} \frac{1+|z|^2}{1+|y^\dagger - z|^{\max\{3+d_u,4+d_y\}}} |A_\pi z - A_p z| \,\mathrm{d}z \\
&\leqslant Cr^{11}\|A_\pi - A_p\| \leqslant Cr^{18} d_g(\pi,p),
\end{aligned}$$

where the last inequality follows from (B.27). Combining (B.28) and (B.31) yields the desired result. ∎

Appendix C. Technical Results for Approximation Result in Proposition 2.2. In Lemma C.1, we recall the local Lipschitz continuity result for the operator G established in [9]. Lemma C.3 establishes that the filtering distribution is a locally Lipschitz function of (Ψ, h), viewed as a mapping from Banach space equipped with the $\|\cdot\|_\infty$ norm into the space of probability measures meterized using the d_g distance. This is preceded by Lemma C.2, which establishes bounds used to prove this Lipschitz property. The two lemmas do not require Ψ_0 and h_0 to be affine but simply require that they both satisfy Assumptions (H3) and (H4). This is in contrast with the more specific setting of Proposition 2.2, which imposes an affine assumption on (Ψ_0, h_0).

LEMMA C.1. *For all $R \geqslant 1$, there exists $L_\mathsf{G} = L_\mathsf{G}(R,n)$ so that for any $\mu_1,\mu_2 \in \mathcal{P}_R(\mathbb{R}^n)$, it holds that*

$$d_g(\mathsf{G}\mu_1, \mathsf{G}\mu_2) \leqslant L_\mathsf{G}(R,n) \cdot d_g(\mu_1,\mu_2).$$

Proof. The lemma as stated may be found in [9, Lemma B.12], where a complete proof is given. ∎

LEMMA C.2. *Suppose that the matrices* (Σ,Γ) *satisfy Assumption* (H2). *Fix $\kappa_\Psi, \kappa_h > 0$ and assume that $\Psi_0\colon \mathbb{R}^{d_u} \to \mathbb{R}^{d_u}$ and $h_0\colon \mathbb{R}^{d_u} \to \mathbb{R}^{d_y}$ are functions satisfying Assumptions* (H3) *and* (H4). *Then, the following statements hold.*

- *There is a constant $C_p = C_p(\kappa_\Psi, \Sigma)$ such that for all $\varepsilon \in [0,1]$ and all $(\Psi,h) \in B_{L^\infty}((\Psi_0,h_0),\varepsilon)$,*

$$\forall \mu \in \mathcal{P}(\mathbb{R}^{d_u}), \qquad d_g(\mathsf{P}_0\mu, \mathsf{P}\mu) \leqslant C_p\varepsilon \cdot \big(1+\mathcal{M}_2(\mu)\big). \tag{C.1}$$

- *There is a constant* $C_q = C_q(\kappa_h, \Gamma)$ *such that for all* $\varepsilon \in [0,1]$ *and all* $(\Psi, h) \in B_{L^\infty}((\Psi_0, h_0), \varepsilon)$,

$$\forall \mu \in \mathcal{P}(\mathbb{R}^{d_u}), \qquad d_g(\mathsf{Q}_0\mu, \mathsf{Q}\mu) \leqslant C_q \varepsilon \cdot \bigl(1 + \mathcal{M}_2(\mu)\bigr). \tag{C.2}$$

- *There is* $C_b = C_b(\kappa_h, \Gamma)$ *such that for all* $y^\dagger \in \mathbb{R}^{d_y}$, *all* $\varepsilon \in [0,1]$, *all* $(\Psi, h) \in B_{L^\infty}((\Psi_0, h_0), \varepsilon)$, *and all probability measures* $(\mu_0, \mu) \in \mathcal{P}(\mathbb{R}^{d_u}) \times \mathcal{P}(\mathbb{R}^{d_u})$,

$$d_g\Bigl(\mathsf{B}(\mathsf{Q}_0\mu_0; y^\dagger), \mathsf{B}(\mathsf{Q}\mu, y^\dagger)\Bigr) \leqslant \exp(C_b \mathcal{R})\Bigl(\varepsilon + d_g(\mu_0, \mu)\Bigr), \tag{C.3}$$

where $\mathcal{R} \in [1, \infty]$ *is given by*

$$\mathcal{R} = \max\Bigl\{\bigl|y^\dagger\bigr|^2, 1 + \mathcal{M}_2(\mu_0), 1 + \mathcal{M}_2(\mu)\Bigr\}.$$

Here, P_0 *and* Q_0 *denote the maps associated with* (Ψ_0, h_0), *and* P *and* Q *are the maps associated with* (Ψ, h).

Proof. Throughout this proof, C_p denotes a constant depending only on (κ_Ψ, Σ), and C_q is a constant that depends only on (κ_h, Γ). Both may change from line to line.

Proof of (C.1). It holds that

$$\mathsf{P}_0\mu(u) - \mathsf{P}\mu(u) = C_p \int_{\mathbb{R}^{d_u}} \exp\left(-\frac{1}{2}\bigl|u - \Psi_0(v)\bigr|_\Sigma^2\right) - \exp\left(-\frac{1}{2}\bigl|u - \Psi(v)\bigr|_\Sigma^2\right) \mu(\mathrm{d}v).$$

By the elementary inequality (A.7), the integrand on the right-hand side is bounded in absolute value by

$$2\bigl|\Psi_0(v) - \Psi(v)\bigr| \left(\exp\left(-\frac{1}{4}\bigl|u - \Psi_0(v)\bigr|_\Sigma^2\right) + \exp\left(-\frac{1}{4}\bigl|u - \Psi(v)\bigr|_\Sigma^2\right)\right).$$

By Young's inequality, it holds that $|a+b|^2 \geqslant \frac{1}{2}|a|^2 - |b|^2$, and so this is bounded by

$$\begin{aligned} &4\bigl|\Psi_0(v) - \Psi(v)\bigr| \left(\exp\left(-\frac{1}{8}\bigl|u - \Psi_0(v)\bigr|_\Sigma^2 + \frac{1}{4}\bigl|\Psi_0(v) - \Psi(v)\bigr|_\Sigma^2\right)\right) \\ &\qquad \leqslant C_p \varepsilon \exp\left(-\frac{1}{8}\bigl|u - \Psi_0(v)\bigr|_\Sigma^2\right). \end{aligned}$$

It follows, by Fubini's theorem, that

$$\begin{aligned} d_g(\mathsf{P}_0\mu, \mathsf{P}\mu) &\leqslant C_p \varepsilon \int_{\mathbb{R}^{d_u}} \int_{\mathbb{R}^{d_u}} \left(1 + |u|^2\right) \exp\left(-\frac{1}{8}\bigl|u - \Psi_0(v)\bigr|_\Sigma^2\right) \mathrm{d}u\, \mu(\mathrm{d}v) \\ &\leqslant C_p \varepsilon \int_{\mathbb{R}^{d_u}} \left(1 + \bigl|\Psi_0(v)\bigr|^2\right) \mu(\mathrm{d}v) \leqslant C_p \varepsilon \left(1 + \kappa_\Psi^2\right) \int_{\mathbb{R}^{d_u}} \left(1 + |v|^2\right) \mu(\mathrm{d}v). \end{aligned}$$

This concludes the proof of (C.1).

Proof of (C.2). Recall that

$$\begin{aligned} d_g(\mathsf{Q}_0\mu, \mathsf{Q}\mu) = C_q \int_{\mathbb{R}^{d_u}} \int_{\mathbb{R}^{d_y}} g(u,y) \Biggl(&\exp\left(-\frac{1}{2}\bigl|y - h_0(u)\bigr|_\Gamma^2\right) \\ &- \exp\left(-\frac{1}{2}\bigl|y - h(u)\bigr|_\Gamma^2\right)\Biggr) \mathrm{d}y\, \mu(\mathrm{d}u), \end{aligned}$$

where $g(u,y) = 1 + |u|^2 + |v|^2$. Using the same reasoning as above, we obtain that

$$\left|\exp\left(-\frac{1}{2}\left|y - h_0(u)\right|_\Gamma^2\right) - \exp\left(-\frac{1}{2}\left|y - h(u)\right|_\Gamma^2\right)\right| \leqslant C_q\varepsilon \exp\left(-\frac{1}{8}\left|y - h_0(u)\right|_\Gamma^2\right). \tag{C.4}$$

Therefore, we deduce that

$$d_g(\mathsf{Q}_0\mu, \mathsf{Q}\mu) \leqslant C_q\varepsilon \int_{\mathbb{R}^{d_u}} \left(1 + \left|h_0(u)\right|^2\right) \mu(\mathrm{d}u) \leqslant C_q\varepsilon(1+\kappa_h^2) \int_{\mathbb{R}^{d_u}} \left(1 + |u|^2\right) \mu(\mathrm{d}u),$$

which proves (C.2).

Proof of (C.3). We assume for simplicity that μ_0 and μ have densities, but note that this is not required. Let $\nu_0 = \mathsf{B}(\mathsf{Q}_0\mu_0; y^\dagger)$ and $\nu = \mathsf{B}(\mathsf{Q}\mu, y^\dagger)$ and recall that

$$\nu_0(u) = \frac{\exp\left(-\frac{1}{2}\left|y^\dagger - h_0(u)\right|_\Gamma^2\right)\mu_0(u)}{\int_{\mathbb{R}^{d_u}} \exp\left(-\frac{1}{2}\left|y^\dagger - h_0(U)\right|_\Gamma^2\right)\mu_0(U)\,\mathrm{d}U} =: \frac{f_0(u)}{\int_{\mathbb{R}^{d_u}} f_0(U)\,\mathrm{d}U} =: \frac{f_0(u)}{Z_0}$$

and similarly

$$\nu(u) = \frac{\exp\left(-\frac{1}{2}\left|y^\dagger - h(u)\right|_\Gamma^2\right)\mu(u)}{\int_{\mathbb{R}^{d_u}} \exp\left(-\frac{1}{2}\left|y^\dagger - h(U)\right|_\Gamma^2\right)\mu(U)\,\mathrm{d}U} =: \frac{f(u)}{\int_{\mathbb{R}^{d_u}} f(U)\,\mathrm{d}U} =: \frac{f_0(u)}{Z}.$$

Note that

$$\begin{aligned} d_g(\nu_0, \nu) &= \int_{\mathbb{R}^{d_u}} \left(1 + |u|^2\right) \left|\frac{f_0(u)}{Z_0} - \frac{f(u)}{Z}\right| \mathrm{d}u \\ &= \frac{1}{Z}\int_{\mathbb{R}^{d_u}} \left(1 + |u|^2\right)\left|f_0(u) - f(u)\right| \mathrm{d}u + \left|\frac{1}{Z_0} - \frac{1}{Z}\right| \int_{\mathbb{R}^{d_u}} \left(1 + |u|^2\right) f_0(u)\,\mathrm{d}u. \end{aligned}$$

In order to bound the first term, we write

$$\begin{aligned} f_0(u) - f(u) = &\left(\exp\left(-\frac{1}{2}\left|y^\dagger - h_0(u)\right|_\Gamma^2\right) - \exp\left(-\frac{1}{2}\left|y^\dagger - h(u)\right|_\Gamma^2\right)\right)\mu_0(u) \\ &+ \exp\left(-\frac{1}{2}\left|y^\dagger - h(u)\right|_\Gamma^2\right)\Big(\mu_0(u) - \mu(u)\Big). \end{aligned}$$

Using (C.4), we obtain that

$$\begin{aligned} \left|f_0(u) - f(u)\right| \leqslant\ &C_q\varepsilon \exp\left(-\frac{1}{8}\left|y^\dagger - h_0(u)\right|_\Gamma^2\right)\mu_0(u) \\ &+ \exp\left(-\frac{1}{2}\left|y^\dagger - h(u)\right|_\Gamma^2\right)\Big|\mu_0(u) - \mu(u)\Big|, \end{aligned} \tag{C.6}$$

and so

$$\int_{\mathbb{R}^{d_u}} \left(1 + |u|^2\right)\left|f_0(u) - f(u)\right| \mathrm{d}u \leqslant C_q\varepsilon\mathcal{R} + d_g(\mu_0, \mu).$$

Therefore it holds that

$$d_g(\nu_0, \nu) \leqslant \mathcal{R}\left(\frac{C_q\varepsilon}{Z} + \frac{|Z_0 - Z|}{Z_0 Z}\right) + \frac{1}{Z} d_g(\mu_0, \mu).$$

By (C.6), it holds that

$$|Z_0 - Z| \leqslant \int_{\mathbb{R}^{d_u}} |f_0(U) - f(U)| \,\mathrm{d}U \leqslant C_q\varepsilon + d_g(\mu_0,\mu),$$

and so we obtain finally

$$d_g(\nu_0,\nu) \leqslant \mathcal{R}\left(\frac{1}{Z} + \frac{1}{Z_0 Z}\right)\Big(C_q\varepsilon + d_g(\mu_0,\mu)\Big).$$

Furthermore, Z_0 is bounded from below because, by Jensen's inequality,

$$\begin{aligned} Z_0 &= \int_{\mathbb{R}^{d_u}} \exp\left(-\frac{1}{2}\left|y^\dagger - h_0(U)\right|_\Gamma^2\right)\mu_0(U)\,\mathrm{d}U \\ &\geqslant \exp\left(-\frac{1}{2}\int_{\mathbb{R}^{d_u}} \left|y^\dagger - h_0(U)\right|_\Gamma^2 \mu_0(U)\,\mathrm{d}U\right) \\ &\geqslant \exp\left(-\left|y^\dagger\right|_\Gamma^2 - C_q\kappa_h^2 \int_{\mathbb{R}^{d_u}} \left(1+|u|^2\right)\mu_0(U)\,\mathrm{d}U\right) = \exp\left(-C_q\mathcal{R}\right). \end{aligned}$$

The same bound holds for Z, and so we obtain the result. ◻

LEMMA C.3. *Suppose that the data* $Y^\dagger = \{y_j^\dagger\}_{j=1}^J$ *and the matrices* (Σ,Γ) *satisfy Assumptions* (H1) *and* (H2). *Fix* $\kappa_\Psi,\kappa_h > 0$ *and assume that* $\Psi_0\colon \mathbb{R}^{d_u} \to \mathbb{R}^{d_u}$ *and* $h_0\colon \mathbb{R}^{d_u} \to \mathbb{R}^{d_y}$ *are functions satisfying Assumptions* (H3) *and* (H4), *respectively. Let* $(\mu_j^0)_{j\in[\![1,J]\!]}$ *and* $(\mu_j)_{j\in[\![1,J]\!]}$ *denote the true filtering distributions associated with functions* (Ψ_0,h_0) *and* (Ψ,h), *respectively, initialized at the same Gaussian probability measure* $\mu_0 = \mathcal{N}(m_0,C_0) \in \mathcal{G}(\mathbb{R}^{d_u})$. *For all* $J\in\mathbb{Z}^+$, *there is* $C = C(m_0,C_0,\kappa_y,\kappa_\Psi,\kappa_h,\Sigma,\Gamma,J) > 0$ *such that for all* $\varepsilon\in[0,1]$ *and all* $(\Psi,h)\in B_{L^\infty}((\Psi_0, h_0),\varepsilon)$, *it holds that*

$$\max_{j\in[\![0,J]\!]} d_g(\mu_j^0,\mu_j) \leqslant C\varepsilon. \tag{C.7}$$

Proof. By Lemma B.3, the filtering distributions have bounded second moments. Let

$$\mathcal{R} = \max_{j\in[\![0,J-1]\!]}\left(\left|y_{j+1}^\dagger\right|^2, 1+\mathcal{M}_2(\mu_j^0), 1+\mathcal{M}_2(\mu_j)\right).$$

Throughout this proof, C denotes a constant whose value is irrelevant in the context, depends only on the constants $m_0, C_0, \kappa_y, \kappa_\Psi, \kappa_h, \Sigma, \Gamma, k$ (but neither on ε, nor on Ψ and h), and may change from line to line.

The statement is obviously true for $J=0$. Reasoning by induction, we assume that the statement is true up to $J=k$ and show that there is $C>0$ such that

$$\forall\varepsilon\in[0,1],\qquad \forall(\Psi,h)\in B_{L^\infty}\big((\Psi_0,h_0),\varepsilon\big),\qquad d_g(\mu_{k+1}^0,\mu_{k+1}) \leqslant C\varepsilon.$$

To this end, let P_0 and Q_0 denote the maps associated with (Ψ_0,h_0), fix $\varepsilon\in[0,1]$, and fix $(\Psi,h)\in B_{L^\infty}((\Psi_0,h_0),\varepsilon)$. Using (C.3), then the triangle inequality, and finally (C.1) and Lemma B.8, we have that

$$\begin{aligned} d_g(\mu_{k+1}^0,\mu_{k+1}) &= d_g\big(\mathsf{B}_k\mathsf{Q}_0\mathsf{P}_0\mu_k^0, \mathsf{B}_k\mathsf{Q}\mathsf{P}\mu_k\big) \\ &\leqslant \mathrm{e}^{C\mathcal{R}}\Big(\varepsilon + d_g\big(\mathsf{P}_0\mu_k^0,\mathsf{P}\mu_k\big)\Big) \\ &\leqslant \mathrm{e}^{C\mathcal{R}}\Big(\varepsilon + d_g\big(\mathsf{P}_0\mu_k^0,\mathsf{P}\mu_k^0\big) + d_g\big(\mathsf{P}\mu_k^0,\mathsf{P}\mu_k\big)\Big) \\ &\leqslant \mathrm{e}^{C\mathcal{R}}\Big(\varepsilon + C\varepsilon\mathcal{R} + C d_g(\mu_k^0,\mu_k)\Big). \end{aligned}$$

Using the induction hypothesis gives us the desired bound. ◻

Appendix D. Technical Result for Theorem 2.4. In [24], machinery is established to prove Monte Carlo error estimates between the finite particle ensemble Kalman filter and its mean field limit. We use such results as a component in proving Theorem 2.4 and, in so doing, explicit dependence on moments must be tracked. In this section, we give a self-contained proof of [24, Theorem 5.2], following the analysis closely[2] and, in addition, tracking dependence on moments; this moment dependence may be useful in the context of future work generalizing what we do in this paper. This leads to the following error estimate stating the desired Monte Carlo error estimate.

LEMMA D.1. *Assume that the probability measures* $(\mu_j^{\mathrm{EK}})_{j\in[\![0,J]\!]}$ *and* $(\mu_j^{\mathrm{EK},N})_{j\in[\![0,J]\!]}$ *are obtained, respectively, from the dynamical systems* (2.4) *and* (2.6), *initialized at the Gaussian probability measure* $\mu_0^{\mathrm{EK}} \in \mathcal{G}(\mathbb{R}^{d_u})$ *and at the empirical measure* $\mu_0^{\mathrm{EK},N} = \frac{1}{N}\sum_{i=1}^N \delta_{u_0^{(i)}}$ *for* $u_0^{(i)} \sim \mu_0^{\mathrm{EK}}$ *i.i.d. samples. That is,*

$$\mu_{j+1}^{\mathrm{EK}} = \mathsf{T}_j\mathsf{QP}\mu_j^{\mathrm{EK}}, \qquad \mu_{j+1}^{\mathrm{EK},N} = \frac{1}{N}\sum_{i=1}^N \delta_{u_{j+1}^{(i)}},$$

where $u_{j+1}^{(i)}$ *evolve according to the iteration in* (2.5). *Suppose that the data* $Y^\dagger = \{y_j^\dagger\}_{j=1}^J$ *and the matrices* (Σ,Γ) *satisfy Assumptions* (H1) *and* (H2). *We assume that the vector field* Ψ *satisfies Assumption* (H3) *and that* h *is a linear transformation, i.e., that Assumption* (V2) *is satisfied and hence also Assumption* (H4). *Furthermore, if the vector field* Ψ *additionally satisfies* $|\Psi|_{C^{0,1}} \leqslant \ell_\Psi < \infty$, *then for all* ϕ *satisfying Assumption* P1, *there exists a constant* $C = C(\mathcal{M}_q(\mu_0^{\mathrm{EK}}), J, R_\phi, L_\phi, \varsigma, \kappa_y, \kappa_\Psi, \kappa_h, \ell_\Psi, \Sigma, \Gamma)$, *where* $q := \max\{4^{J+1}, 4\varsigma, 2(\varsigma+1)\}$ *such that*

$$\left(\mathbb{E}\left|\mu_J^{\mathrm{EK},N}[\phi] - \mu_J^{\mathrm{EK}}[\phi]\right|^2\right)^{\frac{1}{2}} \leqslant \frac{C}{\sqrt{N}}.$$

Proof. To prove the proposition, we apply the coupling argument used in [24]. Using similar notation to the one in [24], to the interacting N-particle system $\{u_j^{(i)}\}_{n=1}^N$ evolving according to the ensemble Kalman dynamics (2.5), we couple N copies of the mean field dynamics $\{\overline{u}_j^{(j)}\}_{n=1}^N$ evolving according to the mean field ensemble Kalman dynamics (2.1). The mean field replicas are synchronously coupled to the interacting particle system, in the sense that they are initialized at the same initial condition and driven by the same noises; namely, the two particle systems are initialized at i.i.d. samples $u_0^{(i)} = \overline{u}_0^{(i)} \sim \mu_0^{\mathrm{EK}}$ for $i = 1,\dots,N$. For simplicity of notation, the forecast particles are denoted by the letter v, and we drop the hat notation from the forecast and simulated observations. Furthermore, we add a bar $\overline{\bullet}$ to all the variables related to the synchronously coupled mean field particles, including the probability measures $\overline{\pi}_j^{\mathrm{EK}}$. We also define for $\pi \in \mathcal{P}(\mathbb{R}^{d_u\times d_y})$ the Kalman gain

$$\mathcal{K}(\pi) := \mathcal{C}^{uh}(\pi)\left(\mathcal{C}^{hh}(\pi) + \Gamma\right)^{-1}.$$

With this notation, the interacting particle system and synchronously coupled system read as follows:

[2]The result of [24] holds more generally for functions Ψ that are locally Lipschitz and that grow at most polynomially at infinity [24, Assumption B]. However, our proof uses that Ψ is globally Lipschitz, for instance in (D.9).

Interacting particle system	*Synchronous coupling*
Initialization: $u_0^{(i)} = \overline{u}_0^{(i)}$	**Initialization:** $\overline{u}_0^{(i)} = u_0^{(i)}$
$v_{j+1}^{(i)} = \Psi\left(u_j^{(i)}\right) + \xi_j^{(i)},$	$\overline{v}_{j+1}^{(i)} = \Psi\left(\overline{u}_j^{(i)}\right) + \xi_j^{(i)},$
$y_{j+1}^{(i)} = h\left(v_{j+1}^{(i)}\right) + \eta_{j+1}^{(i)},$	$\overline{y}_{j+1}^{(i)} = h\left(\overline{v}_{j+1}^{(i)}\right) + \eta_{j+1}^{(i)},$
$u_{j+1}^{(i)} = v_{j+1}^{(i)} + \mathcal{K}\left(\pi_{j+1}^{\mathrm{EK},N}\right)\left(y_{j+1}^{\dagger} - y_{j+1}^{(i)}\right).$	$\overline{u}_{j+1}^{(i)} = \overline{v}_{j+1}^{(i)} + \mathcal{K}\left(\overline{\pi}_{j+1}^{\mathrm{EK}}\right)\left(y_{j+1}^{\dagger} - \overline{y}_{j+1}^{(i)}\right).$

Note that the synchronously coupled particles are independent and identically distributed. With this set-up, we proceed with the proof. Let ϕ satisfy Assumption P1. By applying the triangle inequality, we deduce that

$$\text{(D.1)}\qquad \left(\mathbb{E}\left|\mu_J^{\mathrm{EK},N}[\phi] - \mu_J^{\mathrm{EK}}[\phi]\right|^2\right)^{\frac{1}{2}} \leqslant \left(\mathbb{E}\left|\mu_J^{\mathrm{EK},N}[\phi] - \frac{1}{N}\sum_{i=1}^N \phi\big(\overline{u}_J^{(i)}\big)\right|^2\right)^{\frac{1}{2}} + \left(\mathbb{E}\left|\frac{1}{N}\sum_{i=1}^N \phi\big(\overline{u}_J^{(i)}\big) - \mu_J^{\mathrm{EK}}[\phi]\right|^2\right)^{\frac{1}{2}}.$$

Step 1*: bounding the second term in* (D.1). Noting that the random variables $\phi(\overline{u}_J^{(i)}) - \mu_J^{\mathrm{EK}}[\phi]$ are i.i.d. and expanding the square, we obtain

$$\left(\mathbb{E}\left|\sum_{i=1}^N \frac{1}{N}\left(\phi\big(\overline{u}_J^{(i)}\big) - \mu_J^{\mathrm{EK}}[\phi]\right)\right|^2\right)^{\frac{1}{2}} = \frac{1}{\sqrt{N}}\left(\mathbb{E}\left|\phi\big(\overline{u}_J^{(1)}\big) - \mu_J^{\mathrm{EK}}[\phi]\right|^2\right)^{\frac{1}{2}},$$

Since by Assumption P1, it holds that $|\phi(u)| \leqslant R_\phi(1 + |u|^{\varsigma+1})$ for any $u \in \mathbb{R}^{d_u}$, it follows that

$$\text{(D.2)}\qquad \left(\mathbb{E}\left|\frac{1}{N}\sum_{i=1}^N \phi\big(\overline{u}_J^{(i)}\big) - \mu_J^{\mathrm{EK}}[\phi]\right|^2\right)^{\frac{1}{2}} \leqslant \frac{C\left(\mathcal{M}_{2(\varsigma+1)}\big(\mu_0^{\mathrm{EK}}\big), R_\phi, J, \kappa_y, \kappa_\Psi, \kappa_h, \Sigma, \Gamma\right)}{\sqrt{N}},$$

where we used Lemma B.5 to bound $\mathcal{M}_{2(\varsigma+1)}(\mu_J^{\mathrm{EK}})$ in terms of $\mathcal{M}_{2(\varsigma+1)}(\mu_0^{\mathrm{EK}})$.

Step 2*: bounding the first term in* (D.1). For the first term, by Jensen's inequality, exchangeability, and finally by Assumption P1 on ϕ together with the Cauchy–Schwarz inequality, it holds without loss of generality that

$$\text{(D.3)}\qquad \begin{aligned}&\left(\mathbb{E}\left|\frac{1}{N}\sum_{i=1}^N \phi\left(u_J^{(i)}\right) - \phi\left(\overline{u}_J^{(i)}\right)\right|^2\right)^{\frac{1}{2}}\\ &\leqslant \left(\frac{1}{N}\sum_{i=1}^N \mathbb{E}\left|\phi\left(u_J^{(i)}\right) - \phi\left(\overline{u}_J^{(i)}\right)\right|^2\right)^{\frac{1}{2}}\\ &= \left(\mathbb{E}\left|\phi\left(u_J^{(1)}\right) - \phi\left(\overline{u}_J^{(1)}\right)\right|^2\right)^{\frac{1}{2}}\\ &\leqslant 3L_\phi\left(\mathbb{E}\left|u_J^{(1)} - \overline{u}_J^{(1)}\right|^4\right)^{\frac{1}{4}}\left(\mathbb{E}\left[1 + \left|u_J^{(1)}\right|^{4\varsigma} + \left|\overline{u}_J^{(1)}\right|^{4\varsigma}\right]\right)^{\frac{1}{4}}.\end{aligned}$$

By Lemma B.5, there exists $C = C\left(\mathcal{M}_{4\varsigma}(\mu_0^{\mathrm{EK}}), J, \kappa_y, \kappa_\Psi, \kappa_h, \Sigma, \Gamma\right)$ so that

$$\mathbb{E}\left|\overline{u}_J^{(1)}\right|^{4\varsigma} \leqslant C \quad \text{and} \quad \mathbb{E}\left|u_J^{(1)}\right|^{4\varsigma} \leqslant 2^{4\varsigma-1}\left(\mathbb{E}\left|\overline{u}_J^{(1)}\right|^{4\varsigma} + \mathbb{E}\left|u_J^{(1)} - \overline{u}_J^{(1)}\right|^{4\varsigma}\right) \\ \leqslant 2^{4\varsigma-1}\left(C + \mathbb{E}\left|u_J^{(1)} - \overline{u}_J^{(1)}\right|^{4\varsigma}\right).$$

Hence, it remains to bound terms of the form $\mathbb{E}\left|u_J^{(1)} - \overline{u}_J^{(1)}\right|^p$ with $p = 4$ and $p = 4\varsigma$. Such a bound will follow from a *propagation of chaos* result, which will be the object of Step 4. In Step 3, we show auxiliary moment bounds for the mean field dynamics.

Step 3*: moment bounds for the mean field dynamics.* By Lemma B.5, for any $j \in [\![0, J-1]\!]$, there exists a constant $C = C(\mathcal{M}_p(\mu_0^{\mathrm{EK}}), J, \kappa_y, \kappa_\Psi, \kappa_h, \Sigma, \Gamma)$ so that

$$\left(\mathbb{E}\left|\overline{u}_j^{(1)}\right|^p\right)^{\frac{1}{p}} \leqslant C. \tag{D.4}$$

From this, it immediately follows, by Assumptions (H3) and (H4), that there exist constants C_v, C_y depending on $\mathcal{M}_p(\mu_0^{\mathrm{EK}}), J, \kappa_y, \kappa_\Psi, \kappa_h, \Sigma, \Gamma$ so that

$$\left(\mathbb{E}\left|\overline{v}_{j+1}^{(1)}\right|^p\right)^{\frac{1}{p}} \leqslant C_v, \qquad \left(\mathbb{E}\left|\overline{y}_{j+1}^{(1)}\right|^p\right)^{\frac{1}{p}} \leqslant C_y. \tag{D.5}$$

Furthermore, it holds that for the Kalman gain that

$$\left\|\mathcal{K}(\overline{\pi}_{j+1}^{\mathrm{EK}})\right\|^p \leqslant \left\|\mathcal{C}^{uh}(\overline{\pi}_{j+1}^{\mathrm{EK}})\right\|^p \left\|\Gamma^{-1}\right\|^p = \left\|\Gamma^{-1}\right\|^p \left\|\mathcal{C}^{uu}(\overline{\pi}_{j+1}^{\mathrm{EK}}) H^\top\right\|^p \leqslant C, \tag{D.6}$$

where C depends on $\mathcal{M}_2(\mu_0^{\mathrm{EK}}), J, \kappa_y, \kappa_\Psi, \kappa_h, \Sigma, \Gamma$.

Step 4*: propagation of chaos.* In this step, we prove a propagation of chaos result stating for all p, there exists a constant $C_p = C_p(\mathcal{M}_{4^J \cdot p}(\mu_0^{\mathrm{EK}}), J, \kappa_y, \kappa_\Psi, \kappa_h, \ell_\Psi, \ell_h, \Sigma, \Gamma)$ such that

$$\sup_{N \in \mathbb{N}} \sqrt{N}\left(\mathbb{E}\left|u_J^{(1)} - \overline{u}_J^{(1)}\right|^p\right)^{\frac{1}{p}} \leqslant C_p. \tag{D.7}$$

To prove this result, we follow the strategy in [24] and reason by induction. Fix $j \in [\![0, J-1]\!]$ and assume that for all $p \in \mathbb{N}$, there is $C_{j,p} = C_{j,p}(\mathcal{M}_{4^j \cdot p}(\mu_0^{\mathrm{EK}}), J, \kappa_y, \kappa_\Psi, \kappa_h, \ell_\Psi, \ell_h, \Sigma, \Gamma)$ such that

$$\sup_{N \in \mathbb{N}} \sqrt{N}\left(\mathbb{E}\left|u_j^{(1)} - \overline{u}_j^{(1)}\right|^p\right)^{\frac{1}{p}} \leqslant C_{j,p}. \tag{D.8}$$

By Lipschitz continuity of Ψ and h, we deduce immediately that

(D.9)

$$\sup_{N \in \mathbb{N}} \sqrt{N}\left(\mathbb{E}\left|v_{j+1}^{(1)} - \overline{v}_{j+1}^{(1)}\right|^p\right)^{\frac{1}{p}} \leqslant \ell_\Psi C_{j,p}, \qquad \sup_{N \in \mathbb{N}} \sqrt{N}\left(\mathbb{E}\left|y_{j+1}^{(1)} - \overline{y}_{j+1}^{(1)}\right|^p\right)^{\frac{1}{p}} \leqslant \ell_h \ell_\Psi C_{j,p}.$$

Now, we write

$$u_{j+1}^{(1)} - \overline{u}_{j+1}^{(1)} = \left(v_{j+1}^{(1)} - \overline{v}_{j+1}^{(1)}\right) + \mathcal{K}(\overline{\pi}_{j+1}^{\mathrm{EK}})\left(\overline{y}_{j+1}^{(1)} - y_{j+1}^{(1)}\right) \\ + \left(\mathcal{K}(\pi_{j+1}^{\mathrm{EK},N}) - \mathcal{K}(\overline{\pi}_{j+1}^{\mathrm{EK}})\right)\left(y_{j+1}^{\dagger} - y_{j+1}^{(1)}\right).$$

Thus, by the triangle and Hölder inequalities,

$$\begin{aligned}\sqrt{N}\left(\mathbb{E}\left|u_{j+1}^{(1)}-\overline{u}_{j+1}^{(1)}\right|^{p}\right)^{\frac{1}{p}} &\\ \leqslant \ell_\Psi C_{j,p} + \left\|\mathcal{K}(\overline{\pi}_{j+1}^{\mathrm{EK}})\right\| \ell_h \ell_\Psi C_{j,p} &\\ + \sqrt{N}\left(\mathbb{E}\left\|\mathcal{K}(\pi_{j+1}^{\mathrm{EK},N})-\mathcal{K}(\overline{\pi}_{j+1}^{\mathrm{EK}})\right\|^{2p}\right)^{\frac{1}{2p}}\left(\mathbb{E}\left|y_{j+1}^{\dagger}-y_{j+1}^{(1)}\right|^{2p}\right)^{\frac{1}{2p}}.\end{aligned}$$

The first two terms on the right-hand side are bounded uniformly in N in view of (D.6). In order to bound the last term, we note that the term $\mathbb{E}|y_{j+1}^{\dagger}-y_{j+1}^{(1)}|^{2p}$ is bounded uniformly in N by a constant which depends on $\mathcal{M}_{4^j\cdot 2p}(\mu_0^{\mathrm{EK}})$; this follows by (D.5) and (D.9). To complete the inductive step, it remains to show that

$$\sup_{N\in\mathbb{N}}\sqrt{N}\left(\mathbb{E}\left\|\mathcal{K}\left(\pi_{j+1}^{\mathrm{EK},N}\right)-\mathcal{K}\left(\overline{\pi}_{j+1}^{\mathrm{EK}}\right)\right\|^{2p}\right)^{\frac{1}{2p}}<\infty.$$

To this end, in line with the classical propagation of chaos approach, we decompose

$$\begin{aligned}\sqrt{N}\left(\mathbb{E}\left\|\mathcal{K}\left(\pi_{j+1}^{\mathrm{EK},N}\right)-\mathcal{K}\left(\overline{\pi}_{j+1}^{\mathrm{EK}}\right)\right\|^{2p}\right)^{\frac{1}{2p}} &\leqslant \sqrt{N}\left(\mathbb{E}\left\|\mathcal{K}\left(\pi_{j+1}^{\mathrm{EK},N}\right)-\mathcal{K}\left(\overline{\pi}_{j+1}^{\mathrm{EK},N}\right)\right\|^{2p}\right)^{\frac{1}{2p}}\\ &\quad+\sqrt{N}\left(\mathbb{E}\left\|\mathcal{K}\left(\overline{\pi}_{j+1}^{\mathrm{EK},N}\right)-\mathcal{K}\left(\overline{\pi}_{j+1}^{\mathrm{EK}}\right)\right\|^{2p}\right)^{\frac{1}{2p}},\end{aligned}\tag{D.10}$$

where we introduced the empirical measure associated with the mean field particles, as follows:

$$\overline{\pi}_{j+1}^{\mathrm{EK},N}:=\frac{1}{N}\sum_{i=1}^{N}\delta_{\left(\overline{v}_{j+1}^{(i)},\overline{y}_{j+1}^{(i)}\right)}.$$

To conclude the proof of propagation of chaos, we must prove that

- the first term on the right-hand side of (D.10) is bounded uniformly in N, which is achieved by employing the induction hypothesis (D.8), as well as (D.9); and
- The second term on the right-hand side of (D.10) is also bounded uniformly in N. This follows from classical arguments based on the law of large number in L^p spaces.

Let us first bound the second term. To this end, note that

$$\begin{aligned}\mathcal{K}\left(\overline{\pi}_{j+1}^{\mathrm{EK},N}\right)-\mathcal{K}\left(\overline{\pi}_{j+1}^{\mathrm{EK}}\right)&=\mathcal{C}^{uh}\left(\overline{\pi}_{j+1}^{\mathrm{EK}}\right)\left(\left(\Gamma+\mathcal{C}^{hh}\left(\overline{\pi}_{j+1}^{\mathrm{EK},N}\right)\right)^{-1}-\left(\Gamma+\mathcal{C}^{hh}\left(\overline{\pi}_{j+1}^{\mathrm{EK}}\right)\right)^{-1}\right)\\ &\quad+\left(\mathcal{C}^{uh}\left(\overline{\pi}_{j+1}^{\mathrm{EK},N}\right)-\mathcal{C}^{uh}\left(\overline{\pi}_{j+1}^{\mathrm{EK}}\right)\right)\left(\Gamma+\mathcal{C}^{hh}\left(\overline{\pi}_{j+1}^{\mathrm{EK},N}\right)\right)^{-1}.\end{aligned}$$

Therefore, using $A^{-1}-B^{-1}=B^{-1}(B-A)A^{-1}$, we have that

$$\begin{aligned}\left\|\mathcal{K}\left(\overline{\pi}_{j+1}^{\mathrm{EK},N}\right)-\mathcal{K}\left(\overline{\pi}_{j+1}^{\mathrm{EK}}\right)\right\|&\leqslant\left\|\mathcal{C}^{uh}\left(\overline{\pi}_{j+1}^{\mathrm{EK}}\right)\right\|\|\Gamma^{-1}\|^2\left\|\mathcal{C}^{hh}\left(\overline{\pi}_{j+1}^{\mathrm{EK},N}\right)-\mathcal{C}^{hh}\left(\overline{\pi}_{j+1}^{\mathrm{EK}}\right)\right\|\\ &\quad+\|\Gamma^{-1}\|\left\|\mathcal{C}^{uh}\left(\overline{\pi}_{j+1}^{\mathrm{EK},N}\right)-\mathcal{C}^{uh}\left(\overline{\pi}_{j+1}^{\mathrm{EK}}\right)\right\|\\ &\leqslant C(\Gamma,H)\left(1+\left\|\mathcal{C}^{uu}\left(\overline{\pi}_{j+1}^{\mathrm{EK}}\right)\right\|\right)\left\|\mathcal{C}^{uu}\left(\pi_{j+1}^{\mathrm{EK},N}\right)-\mathcal{C}^{uu}\left(\overline{\pi}_{j+1}^{\mathrm{EK}}\right)\right\|.\end{aligned}$$

The term $\|\mathcal{C}^{uu}(\overline{\pi}_{j+1}^{\mathrm{EK}})\|$ is bounded by (D.5), and the L^p norm of the second term tends to zero with rate $N^{-\frac{p}{2}}$ by classical law of large number arguments; see, for example, [14, Lemma 3] and [29, Lemma 3]; indeed it holds that

$$\mathbb{E}\left\|\mathcal{C}^{uu}\left(\pi_{j+1}^{\mathrm{EK},N}\right)-\mathcal{C}^{uu}\left(\overline{\pi}_{j+1}^{\mathrm{EK}}\right)\right\|^{2p}\leqslant\frac{C\left(\mathcal{M}_{4p}\left(\mu_0^{\mathrm{EK}}\right)\right)}{N^p}.$$

It remains to bound the other term. Reasoning similarly to above, we have that

$$\begin{aligned}&\left\|\mathcal{K}\left(\pi_{j+1}^{\mathrm{EK},N}\right)-\mathcal{K}\left(\overline{\pi}_{j+1}^{\mathrm{EK},N}\right)\right\|\\&\quad\leqslant C(\Gamma,H)\left(1+\left\|\mathcal{C}^{uu}\left(\overline{\pi}_{j+1}^{\mathrm{EK},N}\right)\right\|\right)\left\|\mathcal{C}^{uu}\left(\pi_{j+1}^{\mathrm{EK},N}\right)-\mathcal{C}^{uu}\left(\overline{\pi}_{j+1}^{\mathrm{EK},N}\right)\right\|.\end{aligned}\tag{D.11}$$

By [29, Lemma 2], it holds that

$$\begin{aligned}&\left\|\mathcal{C}^{uu}\left(\pi_{j+1}^{\mathrm{EK},N}\right)-\mathcal{C}^{uu}\left(\overline{\pi}_{j+1}^{\mathrm{EK},N}\right)\right\|^2\\&\quad\leqslant 2\left(\frac{1}{N}\sum_{i=1}^N\left|v_{j+1}^{(i)}-\overline{v}_{j+1}^{(i)}\right|^2\right)\left(\frac{1}{N}\sum_{i=1}^N\left|v_{j+1}^{(i)}\right|^2+\frac{1}{N}\sum_{i=1}^N\left|\overline{v}_{j+1}^{(i)}\right|^2\right).\end{aligned}$$

Using the Cauchy–Schwarz and Jensen's inequalities, moment bounds and finally the induction hypothesis (D.7) enables the conclusion that

$$\left(\mathbb{E}\left\|\mathcal{C}^{uu}\left(\pi_{j+1}^{\mathrm{EK},N}\right)-\mathcal{C}^{uu}\left(\overline{\pi}_{j+1}^{\mathrm{EK},N}\right)\right\|^{2p}\right)^{\frac{1}{2p}}\leqslant\frac{C\left(\mathcal{M}_{4^{j+1}p}\left(\mu_0^{\mathrm{EK}}\right),J,\kappa_y,\kappa_\Psi,\kappa_h,\ell_\Psi,\ell_h,\Sigma,\Gamma\right)}{\sqrt{N}}.$$

Step 5*: concluding the proof.* Using the propagation of chaos result from (D.7) in (D.3) gives that

$$\left(\mathbb{E}\left|\frac{1}{N}\sum_{i=1}^N\phi\left(u_J^{(i)}\right)-\phi\left(\overline{u}_J^{(i)}\right)\right|^2\right)^{\frac{1}{2}}\leqslant\frac{C}{\sqrt{N}},$$

for some constant C depending on $\mathcal{M}_q\left(\mu_0^{\mathrm{EK}}\right),J,L_\phi,\varsigma,\kappa_y,\kappa_\Psi,\kappa_h,\ell_\Psi,\ell_h,\Sigma,\Gamma$ with $q=\max\{4^{J+1},4\varsigma\}$. Combining this result with (D.2) yields the desired result. ∎

Acknowledgments. The work was initiated during The Jacques-Louis Lions Lectures given by AMS at Sorbonne Université in December 2023; the authors are grateful to Albert Cohen for hosting this event.

REFERENCES

[1] S. Agapiou, O. Papaspiliopoulos, D. Sanz-Alonso, and A. M. Stuart, *Importance sampling: Intrinsic dimension and computational cost*, Statist. Sci., 32 (2017), pp. 405–431, https://doi.org/10.1214/17-STS611.

[2] B. D. Anderson and J. B. Moore, *Optimal Filtering*, Courier Corporation, 2012.

[3] M. Asch, M. Bocquet, and M. Nodet, *Data Assimilation*, Fundamentals of Algorithms 11, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2016, pp. 1–1, https://doi.org/10.1137/1.9781611974546.pt1.

[4] E. Bach and M. Ghil, *A multi-model ensemble Kalman filter for data assimilation and forecasting*, J. Adv. Model. Earth Syst., 15 (2023), e2022MS003123, https://doi.org/10.1029/2022MS003123.

[5] P. Bickel, B. Li, and T. Bengtsson, *Sharp failure rates for the bootstrap particle filter in high dimensions, in Pushing the limits of contemporary statistics: Contributions in honor of Jayanta K*, Ghosh, Inst. Math. Stat. (IMS) Collect., 3 (2008), pp. 318–329, https://doi.org/10.1214/074921708000000228.

[6] A. N. Bishop and P. Del Moral, *On the mathematical theory of ensemble (linear-Gaussian) Kalman-Bucy filtering*, Math. Control Signals Systems, 35 (2023), pp. 835–903, https://doi.org/10.1007/s00498-023-00357-2.

[7] C. M. Bishop, *Pattern Recognition and Machine Learning*, Information Science and Statistics, Springer, New York, 2006, https://doi.org/10.1007/978-0-387-45528-0, https://doi-org.extranet.enpc.fr/10.1007/978-0-387-45528-0.

[8] E. Calvello, S. Reich, and A. M. Stuart, *Ensemble Kalman methods: A mean field perspective*, Acta Numer., 34 (2025), pp. 123–291, https://doi.org/10.1017/S0962492924000060.

[9] J. A. Carrillo, F. Hoffmann, A. M. Stuart, and U. Vaes, *The mean-field ensemble Kalman filter: Near-Gaussian setting*, SIAM J. Numer. Anal., 62 (2024), pp. 2549–2587, https://doi.org/10.1137/24M1628207.

[10] N. Chopin and O. Papaspiliopoulos, *An Introduction to Sequential Monte Carlo*, Springer Nature Switzerland AG, Cham, Switzerland, 2020.

[11] J. de Wiljes, S. Reich, and W. Stannat, *Long-time stability and accuracy of the ensemble Kalman–Bucy filter for fully observed processes and small measurement noise*, SIAM J. Appl. Dyn. Syst., 17 (2018), pp. 1152–1181, https://doi.org/10.1137/17M1119056.

[12] P. Del Moral and E. Horton, *A theoretical analysis of one-dimensional discrete generation ensemble Kalman particle filters*, Ann. Appl. Probab., 33 (2023), pp. 1127–1172, https://doi.org/10.1214/22-AAP1843.

[13] P. Del Moral and J. Tugaut, *On the stability and the uniform propagation of chaos properties of ensemble Kalman–Bucy filters*, Ann. Appl. Probab., 28 (2018), pp. 790–850, https://doi.org/10.1214/17-AAP1317, https://doi-org.extranet.enpc.fr/10.1214/17-AAP1317.

[14] Z. Ding and Q. Li, *Ensemble Kalman sampler: Mean-field limit and convergence analysis*, SIAM J. Math. Anal., 53 (2021), pp. 1546–1578, https://doi.org/10.1137/20M1339507.

[15] A. Doucet, N. de Freitas, and N. Gordon, *Sequential Monte Carlo Methods in Practice*, Statistics for Engineering and Information Science, Springer-Verlag, New York, 2001, https://doi.org/10.1007/978-1-4757-3437-9, https://doi-org.extranet.enpc.fr/10.1007/978-1-4757-3437-9.

[16] G. Evensen, *Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics*, J. Geophys. Res. Oceans, 99 (1994), pp. 10143–10162, https://doi.org/10.1029/94JC00572.

[17] N. J. Gerber, F. Hoffmann, and U. Vaes, *Mean-field Limits for Consensus-Based Optimization and Sampling*, preprint, 2312.07373, 2023.

[18] M. Ghil, S. Cohn, J. Tavantzis, K. Bube, and E. Isaacson, *Applications of Estimation Theory to Numerical Weather Prediction*, Springer, New York, New York, NY, 1981, pp. 139–224, https://doi.org/10.1007/978-1-4612-5970-1_5.

[19] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*, Dover Publications, New York, 2007.

[20] R. E. Kalman, *A new approach to linear filtering and prediction problems*, Trans. ASME Ser. D. J. Basic Engrg., 82 (1960), pp. 35–45, https://doi.org/10.1115/1.3662552.

[21] D. T. B. Kelly, K. J. H. Law, and A. M. Stuart, *Well-posedness and accuracy of the ensemble Kalman filter in discrete and continuous time*, Nonlinearity, 27 (2014), pp. 2579–2604, https://doi.org/10.1088/0951-7715/27/10/2579.

[22] T. Lange and W. Stannat, *On the continuous time limit of the ensemble Kalman filter*, Math. Comp., 90 (2021), pp. 233–265, https://doi.org/10.1090/mcom/3588.

[23] K. Law, A. Stuart, and K. Zygalakis, *Data Assimilation*, Texts in Applied Mathematics 62, Springer, Cham, 2015, https://doi.org/10.1007/978-3-319-20325-6, https://doi-org.extranet.enpc.fr/10.1007/978-3-319-20325-6.

[24] F. Le Gland, V. Monbet, and V.-D. Tran, *Large sample asymptotics for the ensemble Kalman filter*, in The Oxford handbook of nonlinear filtering, Oxford Univ. Press, Oxford, 2011, pp. 598–631.

[25] J. Mandel, L. Cobb, and J. D. Beezley, *On the convergence of the ensemble Kalman filter*, Appl. Math., 56 (2011), pp. 533–541, https://doi.org/10.1007/s10492-011-0031-2.

[26] P. Rebeschini and R. van Handel, *Can local particle filters beat the curse of dimensionality?*, Ann. Appl. Probab., 25 (2015), pp. 2809–2866, https://doi.org/10.1214/14-AAP1061, https://doi-org.extranet.enpc.fr/10.1214/14-AAP1061.

[27] S. Reich and C. Cotter, *Probabilistic Forecasting and Bayesian Data Assimilation*, Cambridge University Press, New York, 2015, https://doi.org/10.1017/CBO9781107706804, https://doi-org.extranet.enpc.fr/10.1017/CBO9781107706804.
[28] C. Snyder, T. Bengtsson, P. Bickel, and J. Anderson, *Obstacles to high-dimensional particle filtering*, Mon. Weather Rev., 136 (2008), pp. 4629–4640, https://doi.org/10.1175/2008MWR2529.1.
[29] U. Vaes, *Sharp propagation of chaos for the ensemble Langevin sampler*, J. London Math. Soc., 110 (2024), e13008, https://doi.org/10.1112/jlms.13008.