SIAM J. NUMER. ANAL.
Vol. 62, No. 6, pp. 2549–2587

THE MEAN-FIELD ENSEMBLE KALMAN FILTER: NEAR-GAUSSIAN SETTING*

J. A. CARRILLO†, F. HOFFMANN‡, A. M. STUART‡, AND U. VAES§

Abstract. The ensemble Kalman filter is widely used in applications because, for high-dimensional filtering problems, it has a robustness that is not shared, for example, by the particle filter; in particular, it does not suffer from weight collapse. However, there is no theory which quantifies its accuracy as an approximation of the true filtering distribution, except in the Gaussian setting. To address this issue, we provide the first analysis of the accuracy of the ensemble Kalman filter beyond the Gaussian setting. We prove two types of results: The first type comprises a stability estimate controlling the error made by the ensemble Kalman filter in terms of the difference between the true filtering distribution and a nearby Gaussian, and the second type uses this stability result to show that, in a neighborhood of Gaussian problems, the ensemble Kalman filter makes a small error in comparison with the true filtering distribution. Our analysis is developed for the mean-field ensemble Kalman filter. We rewrite the update equations for this filter and for the true filtering distribution in terms of maps on probability measures. We introduce a weighted total variation metric to estimate the distance between the two filters, and we prove various stability estimates for the maps defining the evolution of the two filters in this metric. Using these stability estimates, we prove results of the first and second types in the weighted total variation metric. We also provide a generalization of these results to the Gaussian projected filter, which can be viewed as a mean-field description of the unscented Kalman filter.

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

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

DOI. 10.1137/24M1628207

1. Introduction.

1.1. Context. This paper is concerned with the study of partially and noisily observed dynamical systems. Filtering refers to the sequential updating of the probability distribution of the state of the (possibly stochastic) dynamical system, given

*Received by the editors January 2, 2024; accepted for publication (in revised form) August 26, 2024; published electronically November 15, 2024.
https://doi.org/10.1137/24M1628207
Funding: The first author was supported by the Advanced Grant Nonlocal-CPD (Nonlocal PDEs for Complex Particle Dynamics: Phase Transitions, Patterns, and Synchronization) of the European Research Council (ERC) Executive Agency under the European Union's Horizon 2020 research and innovation program (grant agreement 883363). The first author was also partially supported by the Engineering and Physical Sciences Research Council (EPSRC) under grants EP/T022132/1 and EP/V051121/1. The second author was supported by start-up funds at the California Institute of Technology, by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) via project 390685813-GZ 2047/1-HCM, and also by NSF CAREER Award 2340762. The work of the third author was supported by a Department of Defense Vannevar Bush Faculty Fellowship and by the SciAI Center, funded by the Office of Naval Research (ONR), under grant N00014-23-1-2729. The fourth author was partially supported by the ERC under the European Union's Horizon 2020 research and innovation program (grant agreement 810367) and by the Agence Nationale de la Recherche under grants ANR-21-CE40-0006 (SINEQ) and ANR-23-CE40-0027 (IPSO).
†Mathematical Institute, University of Oxford, Oxford OX2 6GG, UK (carrillo@maths.ox.ac.uk).
‡Department of Computing and Mathematical Sciences, Caltech, Pasadena, CA 91125 USA (franca.hoffmann@caltech.edu, astuart@caltech.edu).
§MATHERIALS, Inria Paris and CERMICS, École des Ponts, 77420 Champs-sur-Marne, France (urbain.vaes@inria.fr).

partial noisy observations [3, 44, 53]. The Kalman filter, introduced in 1960, provides an explicit solution to this problem in the setting of linear dynamical systems, linear observations, and additive Gaussian noise [39]; the desired probability distribution is Gaussian, and the Kalman filter provides explicit update formulae for the mean and covariance. The extended Kalman filter is a linearization-based methodology which was developed in the 1960s and 1970s to apply to situations beyond the linear-Gaussian setting [37]. It is, however, not practical in high dimensions because of the need to compute and sequentially update large covariance matrices [30]. To address this issue, the ensemble Kalman filter was introduced in 1994 [25], using ensemble-based low-rank approximations of the covariances, and is a method well adapted to high-dimensional problems. The unscented Kalman filter, introduced in 1997, provides an alternative approach to nonlinear and non-Gaussian problems [38], using quadrature to approximate covariance matrices, and is well adapted to problems of moderate dimension. The particle filter [21] is a provably convergent methodology for approximating the filtering distribution [11, 52]. However, it does not scale well to high-dimensional problems [4, 58]; this motivates the increasing adoption of ensemble Kalman methods.

Over the past two decades, ensemble Kalman filters have found widespread use in the geophysical sciences, are starting to be used in other application domains, and have been developed as a general purpose tool for solving inverse problems; for reviews of such developments, see [9, 26, 27]. Despite widespread and growing adoption, theory quantifying the accuracy of the ensemble Kalman filter, in relation to the true filtering distribution, is limited. Two exceptions are the important contributions [45, 47], which concern accuracy in the large particle limit in the setting where the underlying filtering problem is Gaussian. However, there is no proof that the ensemble Kalman methodology can accurately approximate the desired filtering distribution beyond the Gaussian setting. Indeed, in general, the methodology does not correctly reproduce the filtering distribution; for examples and analysis in the setting of filtering and Bayesian inverse problems, see [1, 45] and [23], respectively. The aim of our work is to address this issue by proving accuracy of the ensemble Kalman filter beyond the Gaussian setting; specifically, our analysis applies when the true filtering distribution is close to Gaussian after appropriate lifting to the joint space of state and data. We perform the analysis for the mean-field ensemble Kalman filter, focusing on quantifying the effect of the Gaussian approximation underlying the ensemble Kalman filter. We also study the Gaussian projected filter; this filter can be seen as a mean-field version of the unscented Kalman filter. Both the ensemble and the unscented Kalman mean-field models are defined in [9].

The three primary contributions of our paper are (i) to establish finite time stability estimates in an appropriate weighted TV metric which enables control of first- and second-order moments for various nonlinear maps required to define the nonlinear Markov processes determining filter evolution, (ii) to use these results to establish stability estimates controlling the error made by the mean-field ensemble Kalman filter in terms of the difference between the true filtering distribution and a nearby Gaussian, and (iii) to deploy these estimates to prove closeness of the mean-field ensemble Kalman filters and the true filtering distribution in the near-Gaussian setting. These results are also established for the mean-field unscented Kalman filter. As well as making the three primary contributions, the work suggests many questions for further analysis, and the numerical analysis framework we deploy ("consistency plus stability implies convergence") is a natural one in which to pursue these questions. In particular, we assume bounded vector fields and discuss only finite time error; overcoming these assumptions requires new ideas and is outside the scope of this first paper.

Although our study of the accuracy of ensemble Kalman filters beyond the Gaussian setting is new, there exists a growing body of literature analyzing ensemble Kalman methods from different perspectives. In the context of long-time behavior, the papers [41, 63, 16, 13] focus primarily on the accuracy of the filter in approximating the true trajectory over long time intervals; in contrast, the papers [31, 40] demonstrate a mechanism for filter divergence, an obstacle to obtaining accuracy. Localization is widely used in practice and important to consider for an overall understanding of ensemble Kalman filter performance; see [62, 64]. For analysis of filters in high dimensions, see [59, 46]. For continuous-time and mean-field limits, see [42, 43, 24]. In the context of inverse problems, see [36, 56, 57, 35, 34].

1.2. Overview of paper. In subsection 1.3, which follows, we define useful notation employed throughout the paper. Then, in subsection 1.4, we define the filtering problem as employed throughout the paper, building on this notation. Section 2 introduces the mean-field ensemble Kalman filter and proves our main approximation theorem in the near-Gaussian setting; the theorem is based on a stability estimate which transfers distance between the true filter and its Gaussian projection into the distance between the true filter and the ensemble Kalman filter. In section 3, we define and then state and prove related theory for the Gaussian projected filter. Although we provide theorems of a type not seen before for nonlinear Kalman filters and new methods of analysis to derive them, our work leads to many substantial open problems; we conclude in section 4 by highlighting those we have identified as being of particular value to advancing the field. The reader may wish to study the concluding section 4, in conjunction with our setup in section 2, to understand the specific problem formulation for our main theorems and to appreciate the substantial challenges to be met in order to build on and go beyond our theorems. The key auxiliary results underpinning the proof of our main theorems are given in Appendices B and C; these rely on technical results presented in Appendix A.

1.3. Notation. The Euclidean vector norm is denoted by $|\bullet|$, and the corresponding operator norm on matrices is denoted by $\|\bullet\|$. For a symmetric positive definite matrix $S \in \mathbf{R}^{n\times n}$, the notation $|\bullet|_S$ refers to the weighted norm $|S^{-1/2}\bullet|$. For a function $m\colon \mathbf{R}^n \to \mathbf{R}$ and $r \geqslant 0$, we let $B_{L^\infty}(m,r)$ denote the L^∞ ball of radius r, centered at m, and let $|\bullet|_{C^{0,1}}$ denote the $C^{0,1}$ seminorm, namely, the Lipschitz constant.

We use symbol $\perp\!\!\!\perp$ to denote independence of two random variables. For $m \in \mathbf{R}^n$ and $\Sigma \in \mathbf{R}^{n\times n}$, the notation $\mathcal{N}(m,\Sigma)$ denotes the Gaussian distribution with mean m and covariance C. The notation $\mathcal{P}(\mathbf{R}^n)$ denotes the set of probability measures over $\mathbf{R}^n$, and $\mathcal{P}^p(\mathbf{R}^n)$ is the set of probability measures over $\mathbf{R}^n$ with finite moments up to order p. The notation $\mathcal{P}_{\rm c}(\mathbf{R}^n)$ is the set of probability measures over $\mathbf{R}^n$ with continuous density with respect to the Lebesgue measure, and the notation $\mathcal{G}(\mathbf{R}^n)$ denotes the set of Gaussian probability measures over $\mathbf{R}^n$. We also introduce the Gaussian projection operator $\mathsf{G}\colon \mathcal{P}^2(\mathbf{R}^n) \to \mathcal{G}(\mathbf{R}^n)$ given by

$$\mathsf{G}\mu = \mathcal{N}\bigl(\mathcal{M}(\mu), \mathcal{C}(\mu)\bigr).$$

We observe [5] that

$$\mathsf{G}\mu = \operatorname*{arg\,min}_{\nu\in\mathcal{G}} \mathrm{KL}(\mu\|\nu),$$

where $\mathrm{KL}(\mu\|\nu)$ is the Kullback–Leibler (KL) divergence of μ from ν, defined in (B.30). Note that G defines a nonlinear mapping. We refer to G as a projection because of its characterization as finding the closest point to μ with respect to the $\mathrm{KL}(\mu\|\bullet)$ divergence. Throughout this paper, all probability measures have continuous Lebesgue density because of our assumptions concerning the noise structure in the dynamics model and the data acquisition model. Thus, we abuse notation by using the same symbols for probability measures and their densities. For a probability measure $\mu\in\mathcal{P}(\mathbf{R}^n)$, the notation $\mu(u)$ for $u\in\mathbf{R}^n$ refers to the Lebesgue density of μ evaluated at u, whereas $\mu[f]$ for a function $f\colon\mathbf{R}^n\to\mathbf{R}$ is a shorthand notation for $\int_{\mathbf{R}^n} f\,\mathrm{d}\mu$.

The notations $\mathcal{M}(\mu)$ and $\mathcal{C}(\mu)$ denote, respectively, the mean and covariance under μ:

$$\mathcal{M}(\mu)=\mu[u], \qquad \mathcal{C}(\mu)=\mu\left[(u-\mathcal{M}(\mu))\otimes(u-\mathcal{M}(\mu))\right].$$

The notation $\mathcal{P}_R(\mathbf{R}^n)$ for $R\geqslant 1$ refers to the subset of $\mathcal{P}(\mathbf{R}^n)$ of probability measures whose first and second moments satisfy the bound

$$|\mathcal{M}(\mu)|\leqslant R, \qquad \frac{1}{R^2}I_n\preccurlyeq\mathcal{C}(\mu)\preccurlyeq R^2 I_n. \tag{1.1}$$

Here I_n denotes the identity matrix in $\mathbf{R}^{n\times n}$, and $\preccurlyeq$ is the partial ordering defined by the convex cone of positive semidefinite matrices. Similarly, $\mathcal{G}_R(\mathbf{R}^n)$ is the subset of $\mathcal{G}(\mathbf{R}^n)$ of probability measures satisfying (1.1).

For a probability measure $\pi\in\mathcal{P}(\mathbf{R}^d\times\mathbf{R}^K)$ associated with random variable $(u,y)\in\mathbf{R}^d\times\mathbf{R}^K$, we use the notation $\mathcal{M}^u(\pi)\in\mathbf{R}^d$, $\mathcal{M}^y(\pi)\in\mathbf{R}^K$ for the means of the marginal distributions and the notation $\mathcal{C}^{uu}(\pi)\in\mathbf{R}^{d\times d}$, $\mathcal{C}^{uy}(\pi)\in\mathbf{R}^{d\times K}$, and $\mathcal{C}^{yy}(\pi)\in\mathbf{R}^{K\times K}$ for the blocks of the covariance matrix $\mathcal{C}(\pi)$. That is to say,

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

Throughout this work, we employ the following weighted total variation distance.

DEFINITION 1.1. *Let $g\colon\mathbf{R}^n\to[1,\infty)$ be given by $g(v)=1+|v|^2$. We define the weighted total variation metric $d_g\colon\mathcal{P}^2(\mathbf{R}^n)\times\mathcal{P}^2(\mathbf{R}^n)\to\mathbf{R}$ by*

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

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

Metrics of this type have been employed previously in the literature. See, for example, the early reference [49], where a weighted total variation metric is used for stuyding ergodicity for Markov chains, and the reference [48], where a similar metric is employed in the context of perturbation of Markov chains. See also [33], where a direct proof of Harris's ergodic theorem relying on appropriate weighted total variation norms is given, as well as [68, Theorem 6.15], which states that Wasserstein distances can be controlled by weighted total variation metrics. There are other dual metrics on probability measures commonly employed in the literature, such as the so-called maximum mean discrepancy, which is based on a dual formulation in a reproducing kernel Hilbert space; see, e.g., [61].

Remark 1.2. If μ_1,μ_2 have Lebesgue densities ρ_1, ρ_2, then the weighted metric in Definition 1.1 satisfies

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

Unlike the usual total variation distance, this weighted total variation metric enables control of the differences $|\mathcal{M}(\mu_1) - \mathcal{M}(\mu_2)|$ and $\|\mathcal{C}(\mu_1) - \mathcal{C}(\mu_2)\|$; several lemmas used to prove our main results rely on this control. More precisely, it is possible to show that for all $\mu_1, \mu_2 \in \mathcal{P}(\mathbf{R}^n)$ with finite second moments, it holds that

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

This is the content of Lemma B.4, proved in Appendix B.

1.4. Filtering distribution. We consider a general setting in subsection 1.4.1 and provide details for the Gaussian setting in particular in subsection 1.4.2.

1.4.1. General case. We consider the following stochastic dynamics and data model:

$$u_{j+1} = \Psi(u_j) + \xi_j, \qquad \xi_j \sim \mathcal{N}(0,\Sigma), \tag{1.2a}$$

$$y_{j+1} = h(u_{j+1}) + \eta_{j+1}, \qquad \eta_{j+1} \sim \mathcal{N}(0,\Gamma). \tag{1.2b}$$

Here $\{u_j\}_{j\in[\![0,J]\!]}$ is the unknown state, evolving in $\mathbf{R}^d$, and $\{y_j\}_{j\in[\![0,J]\!]}$ are the observations, evolving in $\mathbf{R}^K$. We assume throughout the paper that the covariance matrices Σ, Γ are positive definite, that the function $h\colon \mathbf{R}^d \to \mathbf{R}^K$ is continuous, that the initial state is distributed according to a Gaussian distribution $u_0 \sim \mathcal{N}(m_0, C_0)$, and that the following independence assumption is satisfied:

$$u_0 \perp\!\!\!\perp \xi_0 \perp\!\!\!\perp \cdots \perp\!\!\!\perp \xi_J \perp\!\!\!\perp \eta_1 \perp\!\!\!\perp \cdots \perp\!\!\!\perp \eta_{J+1}. \tag{1.3}$$

The objective of probabilistic filtering is to sequentially estimate the conditional distribution of the unknown state given the data as new data arrive. The true filtering distribution μ_j is the conditional distribution of the state u_j given a realization $Y_j^\dagger = \{y_1^\dagger, \dots, y_j^\dagger\}$ of the data process up to step j. Data Y_j may be thought of as arising from a realization of (1.2), but the case of model misspecification, where Y_j does not necessarily arise from (1.2), is also of interest.

It is well known [44, 53] that the true filtering distribution evolves according to

$$\mu_{j+1} = \mathsf{L}_j \mathsf{P} \mu_j, \tag{1.4}$$

where P and L_j are maps on probability measures, referred to, respectively, as the prediction and analysis steps in the data assimilation community [3]. The operator $\mathsf{P}\colon \mathcal{P}(\mathbf{R}^n) \to \mathcal{P}(\mathbf{R}^n)$ is linear and defined by the Markov kernel associated with the stochastic dynamics (1.2a). Its action on a probability measure with Lebesgue density μ reads as

$$\mathsf{P}\mu(u) = \frac{1}{\sqrt{(2\pi)^d \det \Sigma}} \int_{\mathbf{R}^d} \exp\left(-\frac{1}{2}\,|u - \Psi(v)|_\Sigma^2\right) \mu(v)\,\mathrm{d}v. \tag{1.5}$$

The operator $\mathsf{L}_j\colon \mathcal{P}(\mathbf{R}^n) \to \mathcal{P}(\mathbf{R}^n)$ is a nonlinear map which formalizes the incorporation of the new datum $y_{j+1}^\dagger$ using Bayes's theorem. Its action on a probability measure in $\mathcal{P}(\mathbf{R}^d)$ with Lebesgue density μ reads as

$$\mathsf{L}_j\mu(u) = \frac{\exp\left(-\frac{1}{2}\left|y^\dagger_{j+1} - h(u)\right|_\Gamma^2\right)\mu(u)}{\int_{\mathbf{R}^d}\exp\left(-\frac{1}{2}\left|y^\dagger_{j+1} - h(U)\right|_\Gamma^2\right)\mu(U)\,\mathrm{d}U}. \tag{1.6}$$

The operator L_j effects a reweighting of the measure to which it applies, with more weight assigned to the state values that are consistent with the observation. It is convenient in this work to decompose the analysis map L_j into the composition $\mathsf{B}_j\mathsf{Q}$. The action of the operator $\mathsf{Q}\colon \mathcal{P}(\mathbf{R}^d) \to \mathcal{P}(\mathbf{R}^d \times \mathbf{R}^K)$ on a probability measure in $\mathcal{P}(\mathbf{R}^d)$ with Lebesgue density μ is given by

$$\mathsf{Q}\mu(u,y) = \frac{1}{\sqrt{(2\pi)^K \det\Gamma}}\exp\left(-\frac{1}{2}\left|y - h(u)\right|_\Gamma^2\right)\mu(u). \tag{1.7}$$

On the other hand, the action of $\mathsf{B}_j\colon \mathcal{P}_c(\mathbf{R}^d\times\mathbf{R}^K) \to \mathcal{P}(\mathbf{R}^d)$ on a probability measure with continuous Lebesgue density μ is given by

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

The operator Q maps a probability measure with density μ into the density associated with the joint random variable $\big(U, h(U)+\eta\big)$, where $U \sim \mu$ is independent of $\eta \sim \mathcal{N}(0,\Gamma)$. The operator B_j performs conditioning on the data $y^\dagger_{j+1}$. Map Q is linear, while B_j is nonlinear. We may thus write (1.4) in the form

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

We note that for any $\mu \in \mathcal{P}(\mathbf{R}^d)$, the measure $\mathsf{Q}\mathsf{P}\mu \in \mathcal{P}_c(\mathbf{R}^d\times\mathbf{R}^K)$ has continuous Lebesgue density, and so the iteration (1.9) is well defined.

1.4.2. Gaussian case. Now consider the setting in which $\Psi(\bullet) = M\bullet$ and $h(\bullet) = H\bullet$ for matrices $M \in \mathbf{R}^{d\times d}$ and $H \in \mathbf{R}^{K\times d}$. In this case, the filtering distribution is Gaussian: $\mu_j = \mathcal{N}(m_j, C_j)$. The Kalman filter [39] gives explicit evolution equations for the pair $(m_j, C_j) \in \mathbf{R}^d \times \mathbf{R}^{d\times d}$. To write these down, it is helpful to note that $\widehat{\mu}_{j+1} := \mathsf{P}\mu_j$ is also Gaussian: $\widehat{\mu}_{j+1} = \mathcal{N}(\widehat{m}_j, \widehat{C}_j)$.

Then $(\mu_{j+1}, \widehat{\mu}_{j+1})$ are determined from μ_j by the update formulae

$$\widehat{m}_{j+1} = Mm_j, \tag{1.10a}$$

$$\widehat{C}_{j+1} = MC_jM^\mathsf{T} + \Sigma, \tag{1.10b}$$

$$m_{j+1} = \widehat{m}_{j+1} + \widehat{C}_{j+1}H^\mathsf{T}\left(H\widehat{C}_{j+1}H^\mathsf{T} + \Gamma\right)^{-1}\left(y^\dagger_{j+1} - H\widehat{m}_{j+1}\right), \tag{1.10c}$$

$$C_{j+1} = \widehat{C}_{j+1} - \widehat{C}_{j+1}H^\mathsf{T}\left(H\widehat{C}_{j+1}H^\mathsf{T} + \Gamma\right)^{-1}H\widehat{C}_{j+1}, \tag{1.10d}$$

where $\{y^\dagger_j\}$ is the observation. Recall that this observation may be thought of as arising from a realization of (1.2) but, in the case of model misspecification, may be derived from a different source.

2. The ensemble Kalman filter. Following from expository discussion of the Gaussian case in subsection 2.1, in subsection 2.2, we define the specific version of the mean-field ensemble Kalman filter that we analyze here; other versions may be found in [9] and will be amenable to similar analyses. In subsection 2.3, we prove our main stability theorem, showing that the error between the true filter and its Gaussian projection may be used to control the error between the true filter and the ensemble Kalman filter. In subsection 2.4, we prove a corollary to this theorem, establishing that the mean-field ensemble Kalman filter accurately approximates the true filter for a specific class of non-Gaussian problems.

2.1. The algorithm: Gaussian case. To motivate the mean-field ensemble Kalman filter, we first consider the Gaussian case and introduce the stochastic dynamical system

$$\widehat{v}_{j+1} = Mv_j + \xi_j, \qquad \xi_j \sim \mathcal{N}(0,\Sigma), \tag{2.1a}$$

$$\widehat{y}_{j+1} = H\widehat{v}_{j+1} + \eta_{j+1}, \qquad \eta_{j+1} \sim \mathcal{N}(0,\Gamma), \tag{2.1b}$$

$$v_{j+1} = \widehat{v}_{j+1} + \widehat{C}_{j+1}H^\top (H\widehat{C}_{j+1}H^\top + \Gamma)^{-1}(y^\dagger_{j+1} - \widehat{y}_{j+1}). \tag{2.1c}$$

Here $\widehat{C}_{j+1}$ denotes the covariance of $\widehat{v}_{j+1}$. A simple calculation reveals that if $v_0 \sim \mathcal{N}(m_0, C_0)$, then in fact $\widehat{v}_{j+1} \sim \mathcal{N}(\widehat{m}_{j+1}, \widehat{C}_{j+1})$ and $v_{j+1} \sim \mathcal{N}(m_{j+1}, C_{j+1})$, where the means and covariances are given by the Kalman filter (1.10). Note that (2.1) constitutes a mean-field stochastic dynamical system because (2.1c) requires knowledge of $\widehat{C}_{j+1}$, which depends on the law μ_j of v_j. The law of this mean-field stochastic dynamical system is thus equal to the law of the Kalman filter. The analysis in [45, 47] is concerned with studying particle approximations of this Gaussian mean-field stochastic dynamical system and convergence to the mean-field limit in the limit of infinite particles. In contrast, our work concerns the study of mean-field stochastic dynamical systems but goes beyond the Gaussian setting. To this end, the next subsection defines the mean-field ensemble Kalman filter in the general, non-Gaussian setting.

2.2. The algorithm: General case. The ensemble Kalman filter may be derived as a particle approximation of various mean-field limits [9]. The specific mean-field ensemble Kalman filter that we study in this paper reads as

$$\widehat{u}_{j+1} = \Psi(u_j) + \xi_j, \qquad \xi_j \sim \mathcal{N}(0,\Sigma), \tag{2.2a}$$

$$\widehat{y}_{j+1} = h(\widehat{u}_{j+1}) + \eta_{j+1}, \qquad \eta_{j+1} \sim \mathcal{N}(0,\Gamma), \tag{2.2b}$$

$$u_{j+1} = \widehat{u}_{j+1} + \mathcal{C}^{uy}\left(\widehat{\pi}^K_{j+1}\right)\mathcal{C}^{yy}\left(\widehat{\pi}^K_{j+1}\right)^{-1}\left(y^\dagger_{j+1} - \widehat{y}_{j+1}\right), \tag{2.2c}$$

where $\widehat{\pi}^K_{j+1} = \mathrm{Law}(\widehat{u}_{j+1}, \widehat{y}_{j+1})$ and independence of the noise terms (1.3) is still assumed to hold. The covariance matrices Γ, Σ are still assumed to be positive definite, so $\mathcal{C}^{yy}(\widehat{\pi}^K_{j+1}) \succ 0$, and the algorithm is well defined. See subsection 1.3 for the definition of the covariance matrices that appear in (2.2c).

Remark 2.1. The mean-field ensemble Kalman filter algorithm may be derived as the best linear unbiased estimator (BLUE) of the predicted state, given the data; see [9].

Let us denote by μ^K_j the law of u_j. In order to rewrite the evolution of μ^K_j in terms of maps on probability measures, let us introduce

$$\mathcal{P}^2_{\succ 0}\left(\mathbf{R}^d \times \mathbf{R}^K\right) := \left\{\pi \in \mathcal{P}^2\left(\mathbf{R}^d \times \mathbf{R}^K\right) : \mathcal{C}^{yy}(\pi) \succ 0\right\}.$$

Then, for a given $y^\dagger_{j+1}$, we denote by $\mathsf{T}_j : \mathcal{P}^2_{\succ 0}(\mathbf{R}^d \times \mathbf{R}^K) \to \mathcal{P}^2(\mathbf{R}^d)$ the map defined by

$$\mathsf{T}_j\pi = \mathscr{T}(\bullet,\bullet;\pi,y^\dagger_{j+1})_\sharp\pi. \tag{2.3}$$

Here the subscript $\sharp$ denotes the pushforward, and for any given $\pi \in \mathcal{P}(\mathbf{R}^d \times \mathbf{R}^K)$ and $z \in \mathbf{R}^K$, the map $\mathscr{T}$ is affine in its first two arguments and given by

$$\begin{aligned} \mathscr{T}(\bullet,\bullet;\pi,z)\colon \mathbf{R}^d \times \mathbf{R}^K &\to \mathbf{R}^d, \\ (u,y) &\mapsto u + \mathcal{C}^{uy}(\pi)\mathcal{C}^{yy}(\pi)^{-1}\big(z-y\big). \end{aligned} \tag{2.4}$$

As demonstrated in [9], with this notation, the evolution of the probability measure μ^K_j may be written in compact form as

$$\mu^K_{j+1} = \mathsf{T}_j\mathsf{QP}\mu^K_j. \tag{2.5}$$

We now discuss the preceding map in relation to (1.9). The specific affine map $\mathscr{T}$ used in (2.2c) is determined by the measure $\widehat{\pi}^K_{j+1}$ (here equal to $\mathsf{QP}\mu^K_j$) and the data $y^\dagger_{j+1}$. That the law of u_j in (2.2) evolves according to (2.5) follows from the following observations:

- If $u_j \sim \mu^K_j$, then $\widehat{u}_{j+1} \sim \mathsf{P}\mu^K_j$ by definition of P.
- Given the definition (1.7) of the operator Q, the random vector $(\widehat{u}_{j+1}, \widehat{y}_{j+1})$ is distributed according to $\widehat{\pi}^K_{j+1} = \mathsf{QP}\mu^K_j$.
- Equation (2.2c) then implies that $u_{j+1} \sim \mathsf{T}_j\mathsf{QP}\mu^K_j$.

As we show in Lemma B.5, the operator T_j coincides with the conditioning operator B_j over the subset $\mathcal{G}(\mathbf{R}^d \times \mathbf{R}^K) \subset \mathcal{P}(\mathbf{R}^d \times \mathbf{R}^K)$ of Gaussian probability measures. Therefore, in the particular case where μ_0 is Gaussian, which is a standing assumption in this paper, and the operators Ψ and h are linear, the mean-field ensemble Kalman filter (2.5) reproduces the exact filtering distribution (1.9), which is then the Kalman filter itself; indeed, this is what we show in subsection 2.1. In the next section, we provide error bounds between (1.9) and (2.5) when Ψ and h are not assumed to be linear.

Our theorems in subsections 2.3 and 2.4 concern relationships between the true filter (1.9) and the mean-field ensemble Kalman filter (2.5). We study the setting in which Ψ and h are not assumed to be linear, so that the true filter is not Gaussian, in subsection 2.3, and then we study small perturbations away from the Gaussian setting that arise when the vector fields Ψ and h are close to constant in subsection 2.4. We recall that $B_{L^\infty}(m,r)$ denotes the L^∞ ball of radius r, centered at m. The theorems hold under the following set of assumptions.

Assumption A. The following assumptions hold on the data $\{y^\dagger_j\}$, the vector fields (Ψ,h), and the covariances (Σ,Γ):

(H1) Fix positive integer J. The data $Y^\dagger = \{y^\dagger_j\}_{j=1}^J$ lie in set $B_y \subset \mathbf{R}^{KJ}$ defined, for some $\kappa_y > 0$, by

$$B_y := \left\{ Y^\dagger \in \mathbf{R}^{KJ} : \max_{j\in[\![0,J]\!]} |y^\dagger_j| \leqslant \kappa_y \right\}.$$

(H2) The function Ψ satisfies $\Psi(\bullet) \in B_\Psi := B_{L^\infty}(0,\kappa_\Psi)$ for some $\kappa_\Psi > 0$.

(H3) The function h is continuous and satisfies $h(\bullet) \in B_h := B_{L^\infty}(0,\kappa_h)$ for some $\kappa_h > 0$.

(H4) The covariance matrices Σ and Γ appearing in (2.2) are positive definite: $\Sigma \succcurlyeq \sigma I_d$ and $\Gamma \succcurlyeq \gamma I_K$ for positive σ and γ.

2.3. Stability theorem: Ensemble Kalman filter. Roughly speaking, our main result states that if the true filtering distributions $(\mu_j)_{j\in[\![1,J]\!]}$ are close to Gaussian after appropriate lifting to the state/data space, then the distribution μ_j^K given by the mean-field ensemble Kalman filter (2.5) is close to the true filtering distribution μ_j given by (1.9) for all $j \in [\![0,J]\!]$. Recall that $|\bullet|_{C^{0,1}}$ denotes the $C^{0,1}$ seminorm, namely, the Lipschitz constant.

THEOREM 2.2 **(stability for the mean-field ensemble Kalman filter).** *Assume that the probability measures* $(\mu_j)_{j\in[\![1,J]\!]}$ *and* $(\mu_j^K)_{j\in[\![1,J]\!]}$ *are obtained, respectively, from the dynamical systems* (1.9) *and* (2.5)*, initialized at the same Gaussian probability measure* $\mu_0 = \mu_0^K \in \mathcal{G}(\mathbf{R}^d)$*. That is,*

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

If Assumption A *holds and* $|h|_{C^{0,1}} \leqslant \ell_h < \infty$*, then there exists* $C = C(J, \kappa_y, \kappa_\Psi, \kappa_h, \ell_h, \Sigma, \Gamma)$ *such that*

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

Remark 2.3. The constant C in the stability estimate is uniform across realizations of the data from B_y. Indeed, since h is assumed bounded, the bound κ_y on the data will hold with probability exponentially close to 1[1] for $\kappa_y \gg 1$ when there is no model misspecification. Relaxing the assumptions of bounded Ψ and h, for example, to the setting of linear plus bounded functions is technically challenging and will be left for future work. Likewise, relaxing positivity assumptions on the noise leads to substantial technical obstacles, deferred for future work. And, finally, relaxing the assumption that J is finite will require further structural assumptions on the long-time behavior of the filter and is also deferred for future work.

Remark 2.4. The stability theorem shows that the error made by the ensemble Kalman filter is controlled by the difference between the true filter, lifted to the joint space of state and observations, and its Gaussian projection. To interpret the result, it is thus important to have intuition about what it means for a measure to be close to its Gaussian projection in the d_g metric. To this end, note that by Remark 1.2, it is necessary that the means and covariances of the measure and its Gaussian projection are close for the measures to be close in the d_g metric. This is automatically satisfied for the difference between a measure and its Gaussian projection by construction. Since closeness in d_g requires the expectations of all functions growing no faster than quadratic to be close, a useful rule of thumb for practitioners is that the quantity $d_g(\mathsf{Q}\mathsf{P}\mu_j, \mathsf{G}\mathsf{Q}\mathsf{P}\mu_j)$ is small when matching first and second moments allows control of all quadratically bounded functions. This will happen, for example, when Ψ and h are small nonlinear perturbations of affine functions, and hence it will happen when Ψ and h are small nonlinear perturbations of either linear functions or constant functions. For technical reasons (see previous remark), this paper excludes the case of linearly growing functions, but we do study small perturbations of constant functions in Corollary 2.5. There is also reason to expect the filtering distribution to be close to Gaussian when the data volume is high or noise is small and some version of observability applies. Then central limit theorems for inverse problems, such as the Bernstein von Mises theorem, may in the future be developed to quantify this assertion [66]. For work pointing in this direction but in the infinite-dimensional case where Bernstein von Mises–type results are harder to establish, see [50].

[1]With respect to the noise realization leading to the data.

The proof presented hereafter relies on a number of auxiliary results, which are summarized below and proved in Appendix B:

1. For any probability measure μ, the first moments of the probability measures $\mathsf{P}\mu$ and $\mathsf{QP}\mu$ are bounded from above, and their second moments are bounded both from above and from below. The constants in these bounds depend only on the parameters κ_Ψ, κ_h, Σ, and Γ. See Lemmas B.1 and B.2.
2. For any Gaussian measure $\mu \in \mathcal{G}(\mathbf{R}^d \times \mathbf{R}^K)$, it holds that $\mathsf{B}_j\mu = \mathsf{T}_j\mu$. See Lemma B.5 and also [9].
3. The map P is globally Lipschitz on $\mathcal{P}(\mathbf{R}^d)$ for the metric d_g, with a Lipschitz constant L_P depending only on the parameters κ_Ψ and Σ. See Lemma B.7.
4. The map Q is globally Lipschitz on $\mathcal{P}(\mathbf{R}^d)$ for the metric d_g, with a Lipschitz constant L_Q depending only on the parameters κ_h and Γ. See Lemma B.8.
5. The map B_j satisfies for any $\pi \in \mathrm{Im}(\mathsf{QP}) \subset \mathcal{P}(\mathbf{R}^d \times \mathbf{R}^K)$ the bound
$$\forall j \in [\![0, J]\!], \qquad d_g(\mathsf{B}_j\mathsf{G}\pi, \mathsf{B}_j\pi) \leqslant C_\mathsf{B} d_g(\mathsf{G}\pi, \pi),$$
where $C_\mathsf{B} = C_\mathsf{B}(\kappa_y, \kappa_\Psi, \kappa_h, \Sigma, \Gamma)$. This statement concerns the stability of the B_j operator between a measure and its Gaussian approximation. See Lemma B.9.
6. The map T_j satisfies the following bound: For all $R \geqslant 1$, it holds for all probability measures $\pi \in \mathcal{P}_R(\mathbf{R}^d \times \mathbf{R}^K)$ and $p \in \mathrm{Im}(\mathsf{QP}) \subset \mathcal{P}(\mathbf{R}^d \times \mathbf{R}^K)$ that
$$\forall j \in [\![0, J]\!], \qquad d_g(\mathsf{T}_j\pi, \mathsf{T}_j p) \leqslant L_\mathsf{T}\, d_g(\pi, p)$$
for a constant $L_\mathsf{T} = L_\mathsf{T}(R, \kappa_y, \kappa_\Psi, \kappa_h, \Sigma, \Gamma)$. This statement may be viewed as a local Lipschitz continuity result in the case where the second argument of d_g is restricted to the range of QP. See Lemma B.10.

Proof of Theorem 2.2. In what follows, we refer to the preceding itemized list to clarify the proof. For notational simplicity, it is helpful to define the following measure of the difference between the true filtering distribution and its Gaussian projection:

$$\varepsilon := \max_{j \in [\![0, J-1]\!]} d_g(\mathsf{QP}\mu_j, \mathsf{GQP}\mu_j). \tag{2.6}$$

Assume throughout the following that $j \in [\![0, J-1]\!]$. The main idea of the proof results from the following use of the triangle inequality:

$$d_g(\mu^K_{j+1}, \mu_{j+1}) = d_g\left(\mathsf{T}_j\mathsf{QP}\mu^K_j, \mathsf{B}_j\mathsf{QP}\mu_j\right) \tag{2.7a}$$
(2.7b)
$$\leqslant d_g\left(\mathsf{T}_j\mathsf{QP}\mu^K_j, \mathsf{T}_j\mathsf{QP}\mu_j\right) + d_g\left(\mathsf{T}_j\mathsf{QP}\mu_j, \mathsf{T}_j\mathsf{GQP}\mu_j\right) + d_g\left(\mathsf{B}_j\mathsf{GQP}\mu_j, \mathsf{B}_j\mathsf{QP}\mu_j\right).$$

We have used the fact that the second argument of the second term on the right-hand side indeed coincides with the first argument of the third term because $\mathsf{T}_j\mathsf{GQP}\mu_j = \mathsf{B}_j\mathsf{GQP}\mu_j$ by item 2 (Lemma B.5.) By item 1 (Lemma B.2), there is a constant $R \geqslant 1$, depending on the covariance matrices Σ, Γ and the bounds κ_Ψ and κ_h from Assumption A, such that $\mathrm{Im}(\mathsf{QP}) \subset \mathcal{P}_R(\mathbf{R}^d \times \mathbf{R}^K)$. By items 3, 4, and 6 (Lemmas B.7, B.8, and B.10), the first term in (2.7b) satisfies

$$d_g\left(\mathsf{T}_j\mathsf{QP}\mu^K_j, \mathsf{T}_j\mathsf{QP}\mu_j\right) \leqslant L_\mathsf{T}(R) L_\mathsf{Q} L_\mathsf{P} d_g\left(\mu^K_j, \mu_j\right), \tag{2.8}$$

where, for conciseness, we omitted the dependence of the constants on $\kappa_y, \kappa_\Psi, \kappa_h, \Sigma, \Gamma$. Equation (2.8) establishes that the composition of maps $\mathsf{T}_j\mathsf{QP}$ is globally Lipschitz over $\mathcal{P}(\mathbf{R}^d)$. Since $\mathsf{GQP}\mu_j \in \mathcal{G}_R(\mathbf{R}^d \times \mathbf{R}^K)$ by definition of R, the second term in (2.7b) may be bounded using item 6 (Lemma B.10) and the definition in (2.6) of ε:

$$d_g\left(\mathsf{T}_j\mathsf{QP}\mu_j, \mathsf{T}_j\mathsf{GQP}\mu_j\right) \leqslant L_{\mathsf{T}}(R) d_g\left(\mathsf{QP}\mu_j, \mathsf{GQP}\mu_j\right) \leqslant L_{\mathsf{T}}(R)\varepsilon.$$

Finally, the third term in (2.7b) can be bounded using item 5 (Lemma B.9) and the definition in (2.6) of ε:

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

Therefore, letting $\ell = L_{\mathsf{T}}(R) L_{\mathsf{Q}} L_{\mathsf{P}}$, we have shown that

$$d_g(\mu_{j+1}^K, \mu_{j+1}) \leqslant \ell d_g(\mu_j^K, \mu_j) + \big(L_{\mathsf{T}}(R) + C_{\mathsf{B}}\big)\varepsilon,$$

and the conclusion follows from the discrete Gronwall lemma since $\mu_0 = \mu_0^K$. ∎

2.4. Approximation theorem: Ensemble Kalman filter. Theorem 2.2 shows that the ensemble Kalman filter error can be made arbitarily small if the true filter is arbitarily close to its Gaussian projection in state-observation space. This "closeness to Gaussian" assumption can be satisfied in our setting of bounded vector fields by considering small perturbations of constant vector fields and is the content of the following corollary. The result provides a first step in the analysis of the accuracy of the ensemble Kalman filter as an approximation of the true filter; desirable generalizations of what we prove here are discussed in the conclusions in section 4.

Corollary 2.5 **(accuracy for the mean-field ensemble Kalman filter).** *Let* (Σ, Γ) *satisfy Assumption* (H4). *Suppose that* $\Psi_0 \colon \mathbf{R}^d \to \mathbf{R}^d$ *and* $h_0 \colon \mathbf{R}^d \to \mathbf{R}^K$ *are functions taking constant values, and denote by* $B_{\Psi_0,h_0}(r)$ *the set of all functions* (Ψ, h) *satisfying* $\Psi \in B_{L^\infty}(\Psi_0, r)$, $h \in B_{L^\infty}(h_0, r)$, *and Assumptions* (H2) *and* (H3). *Assume also that* $|h|_{C^{0,1}} \leqslant \ell_h < \infty$, *and denote by* $(\mu_j)_{j\in[\![1,J]\!]}$ *and* $(\mu_j^K)_{j\in[\![1,J]\!]}$ *the probability measures obtained, respectively, from the dynamical systems* (1.9) *and* (2.5), *initialized at the same Gaussian probability measure* $\mu_0 = \mu_0^K \in \mathcal{G}(\mathbf{R}^d)$. *That is,*

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

Then for any $\epsilon > 0$, *there exists* $\delta > 0$ *such that*

$$\sup_{Y^\dagger \in B_y} \sup_{(\Psi,h)\in B_{\Psi_0,h_0}(\delta)} d_g(\mu_J^K, \mu_J) \leqslant \epsilon.$$

Proof. The result follows from Theorem 2.2 and Proposition C.2 in Appendix C. ∎

3. Generalization: Gaussian projected filter. We generalize the main result to the Gaussian projected filter defined in [9]. This algorithm may be viewed as a mean-field version of the unscented Kalman filter [38]. Using a similar approach to that adopted in the previous section, we prove a similar stability theorem and accuracy corollary. The Gaussian projected filter is defined by the iteration

$$\mu_{j+1}^G = \mathsf{B}_j\mathsf{GQP}\mu_j^G. \tag{3.1}$$

This iteration is obtained from (1.9) by inserting a projection onto Gaussians before the conditioning step B_j. Because conditioning of Gaussians on linear observations preserves Gaussianity, the preceding map generates sequence of measures $\{\mu_j^G\}$ in $\mathcal{G}(\mathbf{R}^d)$.

Remark 3.1. As shown in Lemma B.6, the evolution (3.1) may also be rewritten in the form

$$\mu_{j+1}^G = \mathsf{G}\mathsf{T}_j\mathsf{Q}\mathsf{P}\mu_j^G. \tag{3.2}$$

This shows that the Gaussian projected filter may also be obtained from the mean-field ensemble Kalman filter (2.5) by adding a Gaussian projection step after the conditioning.

Like (2.5), the filter (3.1) reproduces the true filtering distributions when Ψ and h are linear, assuming that μ_0 is Gaussian. The following theorem quantifies the closeness of the Gaussian projected filter to the true filtering distribution when the linearity assumption on Ψ and h is relaxed.

THEOREM 3.2 **(stability for the Gaussian projected filter).** *Assume that the probability measures* $(\mu_j)_{j\in[\![1,J]\!]}$ *and* $(\mu_j^K)_{j\in[\![1,J]\!]}$ *are obtained, respectively, from the dynamical systems* (1.9) *and* (3.1)*, initialized at the same Gaussian probability measure* $\mu_0 = \mu_0^G$*. That is,*

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

If Assumption A *holds, then there exists* $C = C(J, \kappa_y, \kappa_\Psi, \kappa_h, \Sigma, \Gamma)$ *such that*

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

The proof of this error estimate relies on the auxiliary results Lemmas B.2 and B.7–B.9 already discussed as well as the following two additional results:

- The map G is locally Lipschitz for the metric d_g in the sense that for any $R \geqslant 1$, this map is Lipschitz continuous over the set $\mathcal{P}_R(\mathbf{R}^d)$ given in (1.1). The associated Lipschitz constant is denoted by $L_{\mathsf{G}} = L_{\mathsf{G}}(R)$ and diverges as $R \to \infty$. This result is proved in Lemma B.12, which relies on an auxiliary result shown in Lemma B.11 on the distance between Gaussians in the d_g metric.
- The map B_j is locally Lipschitz for the metric d_g over Gaussians in the sense that for any $R \geqslant 1$ and any $j \in [\![0, J-1]\!]$, this map is Lipschitz continuous over $\mathcal{G}_R(\mathbf{R}^d \times \mathbf{R}^K)$. The associated Lipschitz constant is denoted by $L_{\mathsf{B}} = L_{\mathsf{B}}(R, \kappa_y)$. See Lemma B.15.

Proof. We obtain by the triangle inequality that

$$\begin{aligned} d_g(\mu_{j+1}^G, \mu_{j+1}) &= d_g(\mathsf{B}_j\mathsf{G}\mathsf{Q}\mathsf{P}\mu_j^G, \mathsf{B}_j\mathsf{Q}\mathsf{P}\mu_j) \\ &\leqslant d_g(\mathsf{B}_j\mathsf{G}\mathsf{Q}\mathsf{P}\mu_j^G, \mathsf{B}_j\mathsf{G}\mathsf{Q}\mathsf{P}\mu_j) + d_g(\mathsf{B}_j\mathsf{G}\mathsf{Q}\mathsf{P}\mu_j, \mathsf{B}_j\mathsf{Q}\mathsf{P}\mu_j). \end{aligned}$$

It follows from Lemmas B.2, B.7, B.8, B.12, and B.15 that the composition of maps $\mathsf{B}_j\mathsf{G}\mathsf{Q}\mathsf{P}$ is globally Lipschitz continuous on $\mathcal{P}(\mathbf{R}^d)$ with a constant $\ell = \ell(R, \kappa_y, \kappa_\Psi, \kappa_h, \Sigma, \Gamma)$, where $R = R(\kappa_\Psi, \kappa_h, \Sigma, \Gamma)$ is a positive constant such that $\mathrm{Im}(\mathsf{Q}\mathsf{P}) \subset \mathcal{P}_R(\mathbf{R}^d \times \mathbf{R}^K)$. Therefore, the first term on the right-hand side may be bounded by

$$d_g(\mathsf{B}_j\mathsf{G}\mathsf{Q}\mathsf{P}\mu_j^G, \mathsf{B}_j\mathsf{Q}\mathsf{P}\mu_j) \leqslant \ell d_g(\mu_j^G, \mu_j).$$

For notational simplicity, we again let ε be as in (2.6). Using Lemma B.9 and the definition of ε, the second term may be bounded from above by

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

where $C_{\mathsf{B}} = C_{\mathsf{B}}(\kappa_y, \kappa_\Psi, \kappa_h, \Sigma, \Gamma)$ is the constant from Lemma B.9. The proof can then be concluded in the same way as that of Theorem 2.2. ∎

Corollary 3.3 **(accuracy for the Gaussian projected filter).** *Let* (Σ, Γ) *satisfy Assumption* (H4). *Suppose that* $\Psi_0 \colon \mathbf{R}^d \to \mathbf{R}^d$ *and* $h_0 \colon \mathbf{R}^d \to \mathbf{R}^K$ *are functions taking constant values, and denote by* $B_{\Psi_0,h_0}(r)$ *the set of all functions* (Ψ, h) *satisfying* $\Psi \in B_{L^\infty}(\Psi_0, r)$, $h \in B_{L^\infty}(h_0, r)$ *and items* (H2) *and* (H3). *Assume that the probability measures* $(\mu_j)_{j\in[\![1,J]\!]}$ *and* $(\mu_j^K)_{j\in[\![1,J]\!]}$ *are obtained, respectively, from the dynamical systems* (1.9) *and* (3.1) *and initialized at the same Gaussian probability measure* $\mu_0 = \mu_0^K \in \mathcal{G}(\mathbf{R}^d)$. *That is,*

$$\mu_{j+1} = \mathsf{B}_j\mathsf{QP}\mu_j, \qquad \mu_{j+1}^G = \mathsf{B}_j\mathsf{GQP}\mu_j^G.$$

Then for any $\epsilon > 0$, *there exists* $\delta > 0$ *such that*

$$\sup_{Y^\dagger \in B_y}\ \sup_{(\Psi,h)\in B_{\Psi_0,h_0}(\delta)} d_g(\mu_J^G, \mu_J) \leqslant \epsilon.$$

4. Discussion and future directions. We have provided the first analysis of the error incurred by ensemble Kalman filters as approximations of true filtering distribution beyond the linear-Gaussian setting. We have employed an appropriate weighted TV metric and obtained new stability estimates in this metric in order to establish the approximation results. Our framing of the problem is motivated by the framing of the analysis of the particle filter contained in [52, section 1], a systematization of the original proof of convergence of the particle filter appearing in [14]. Although it introduces new methodology and theoretical results for nonlinear Kalman filters, our work leaves open numerous avenues for further analysis; we now highlight those that we identify as particularly important. These remaining open problems are substantial, but the framework we map out in this paper is an appropriate one in which to address them:

(i) Our theorems concern the mean-field limit of the ensemble Kalman filter; it is of interest to study finite particle approximations of the mean-field limit along the lines of the work in [45, 7] and analogous continuous-time analyses in [18, 19, 20]. Analysis of mean-field limits of interacting particle systems is an established field; interfacing the natural metrics employed for such analyses (Wasserstein) with those employed here (weighted TV) will be required.

(ii) We have made boundedness assumptions on $\Psi(\bullet)$ and $h(\bullet)$ in this paper. Developing proofs which relax these assumptions, allowing consideration of small nonlinear perturbations of the Kalman filter setting, for example, will be very valuable.

(iii) Our error bounds are for a finite number of steps and, as is typical of finite time error estimates that employ a "consistency plus stability implies convergence" approach, lead to error constants which grow exponentially with the time horizon. The literature on analysis of numerical methods for nonautonomous dynamical systems demonstrates that going beyond finite time error estimates that grow in time is, in general, not possible [60]; in that context, generalizing to long-time estimates requires assumptions on the long-term stability or even ergodicity of the dynamical system. Such long-term stability issues are widely studied for the true filtering distribution; see [51, 15, 67, 65, 12, 54], for example. They are complicated by the fact that the nonlinear evolution equation for the filtering distribution is nonautonomous due to the observation signal. In the context of the ensemble Kalman filter, such stability estimates are used in the paper [16], which identifies linear and Gaussian filtering problems in which it is possible to generalize the large particle asymptotic analyses of [45, 47]. There is also analysis of the ensemble Kalman filter for nonlinear non-Gaussian problems, such as the data assimilation problem for the Navier–Stokes equation, but this work concerns only accuracy of state estimation, not the entire filtering distribution [41, 6].

(iv) We have assumed a model for evolution of the state which employs additive Gaussian noise. Generalizing to a general Markov chain would be valuable. Relatedly, it is of interest to relax the assumption of including noise in the dynamical system for the state to allow for deterministic dynamics.

(v) We have studied a (widely used) version of the ensemble Kalman filter which employs a specific transport map to approximate the conditioning step in the filter. Other transport maps are also used in practice, such as that leading to the ensemble square root filter; as outlined in [9, section 2]. Analyzing these other methods would be of great interest.

(vi) Filtering can be used to solve inverse problems, as outlined in [9, section 4]; in particular, some of these methods rely on filtering over the infinite time horizon, the analysis of which will require new ideas, such as in [22].

(vii) Continuous-time versions of our analysis would be of interest, as would their relationship to the Kushner–Stratonovich equation [11]; the paper [16], which studies mean-field limits of the ensemble Kalman–Bucy filter, may be important in this context. Furthermore, study of the ensemble Kalman sampler [28, 29, 10] for inverse problems would be of interest.

(viii) The papers [45] and [20] propose the reweighting of ensemble Kalman methods to make them unbiased; further analysis of this idea and the development of new methods that can carry this out in a derivative-free fashion would be of interest.

(ix) Ensemble Kalman filters in practice employ very small numbers of ensemble members and use covariance (spatial) localization when the unknown states are fields; developing analyses which account for low rank approximations and localization would be a valuable step for the field. Furthermore, it would also be of interest to study the problem of quantization of measures [32] in the context of filtering; however, this is a very hard problem even in simple low-dimensional settings [8], and studying it for the evolution of the filtering distribution will require substantial new ideas.

(x) Particle filters suffer from a curse of dimensionality on certain families of high-dimensional problems [4, 58]. Studying ensemble Kalman methods to determine whether they ameliorate this issue in the near-Gaussian setting would be of interest. There are a variety of different models that can be employed to characterize families of high-dimensional filtering problems [55], and such structural assumptions will undoubtedly affect the results that might be obtained for ensemble Kalman methods.

(xi) Exploiting small noise or large data limits to establish that the error (2.6) is small and then using this fact to analyze the error in the ensemble Kalman filter would be of great interest.

Appendix A. Auxiliary results. We begin by presenting an elementary result used throughout the article.

LEMMA A.1. *Suppose that X is a random variable with values in $\mathbf{R}^d$ and finite second moment, and let $m := \mathbf{E}[X]$. Then*

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

and

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

Proof. We have

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

Taking $a = 0$, we obtain (A.1). In addition, since the second term in the last expression is positive semidefinite, the inequality (A.2) follows. ∎

The following lemma is very similar to [2, Lemma 3.1]; the only difference is that the weight on the left-hand side is given by $1 + |u|^2$ instead of u^2. We give the proof for completeness.

LEMMA A.2 (generalized Pinsker inequality). *Let $g(u) = 1 + |u|^2$, and assume that μ_1, μ_2 are probability measures over $\mathbf{R}^d$ satisfying $\mu_1[g^2] < \infty$ and $\mu_2[g^2] < \infty$. Then, if $\mu_1 \ll \mu_2$,*

$$d_g(\mu_1,\mu_2)^2 \leqslant 2\big(\mu_1\left[g^2\right] + \mu_2\left[g^2\right]\big)\mathrm{KL}(\mu_1\|\mu_2).$$

Proof. Denote the density by $f := \frac{\mathrm{d}\mu_1}{\mathrm{d}\mu_2}$, noting that this is nonnegative. Applying Taylor's formula to the function $\ell \colon u \mapsto u \log u$ around $u = 1$, we deduce that

$$\forall u \geqslant 0, \qquad u \log u \geqslant (u-1) + \frac{1}{2}\left(\min_{v\in I} \ell''(v)\right)(u-1)^2 = (u-1) + \frac{(u-1)^2}{2\max\{1,u\}},$$

where $I = [\min\{1,u\}, \max\{1,u\}]$. Therefore,

$$\begin{aligned}\mathrm{KL}(\mu_1\|\mu_2) &= \int_{\mathbf{R}^d} (f \log f)(u)\,\mu_2(\mathrm{d}u) \\ &\geqslant \int_{\mathbf{R}^d} (f(u)-1)\,\mu_2(\mathrm{d}u) + \frac{1}{2}\int_{\mathbf{R}^d} \frac{|f(u)-1|^2}{\max\{1,f(u)\}}\,\mu_2(\mathrm{d}u).\end{aligned}$$

Let $\theta(u) = \max\{1, f(u)\}$. The first term on the right-hand side is zero, and so we have that

$$\begin{aligned}d_g(\mu_1,\mu_2)^2 &= \left(\int_{\mathbf{R}^d} g(u)\,|f(u)-1|\;\mu_2(\mathrm{d}u)\right)^2 \\ &\leqslant \int_{\mathbf{R}^d} |g(u)|^2\theta(u)\,\mu_2(\mathrm{d}u) \int_{\mathbf{R}^d} \frac{|f(u)-1|^2}{\theta(u)}\,\mu_2(\mathrm{d}u) \\ &\leqslant \int_{\mathbf{R}^d} |g(u)|^2\big(f(u)+1\big)\,\mu_2(\mathrm{d}u)\, 2\,\mathrm{KL}(\mu_1\|\mu_2) \\ &= 2\Big(\mu_1\left[g^2\right] + \mu_2\left[g^2\right]\Big)\,\mathrm{KL}(\mu_1\|\mu_2),\end{aligned}$$

which concludes the proof. ∎

LEMMA A.3. *For all $d \in \mathbf{N}^+$ and $\alpha > 0$, there is $C > 0$ such that for all $x_0, x_1, m_0, m_1 \in \mathbf{R}^d$ and all symmetric positive definite matrices $S_0, S_1 \in \mathbf{R}^{d\times d}$ satisfying*

$$S_0 \succcurlyeq \frac{1}{\alpha} I_d, \qquad S_1 \succcurlyeq \frac{1}{\alpha} I_d,$$

it holds that

$$\begin{aligned}(2\pi)^{\frac{d}{2}}\left|g_0(x_0) - g_1(x_1)\right| \leqslant \frac{\alpha^{\frac{1+d}{2}}}{\sqrt{\mathrm{e}}}|x_1 - x_0 - m_1 + m_0| + \frac{\alpha^{1+\frac{d}{2}}}{\mathrm{e}}\|S_1 - S_0\| \\ + \alpha^d \left|\sqrt{\det S_1} - \sqrt{\det S_0}\right|. \end{aligned} \tag{A.3}$$

Here g_0 and g_1 denote the densities of $\mathcal{N}(m_0, S_0)$ and $\mathcal{N}(m_1, S_1)$, respectively.

Proof. For $s \in [0,1]$, let $x_s = (1-s)x_0 + sx_1$ as well as $m_s = (1-s)m_0 + sm_1$ and $S_s = (1-s)S_0 + sS_1$. Let us also introduce the nonnormalized densities

$$\widetilde{g}_0 = \exp\left(-\frac{1}{2}\left|x - m_0\right|_{S_0}^2\right), \qquad \widetilde{g}_1 = \exp\left(-\frac{1}{2}\left|x - m_1\right|_{S_1}^2\right),$$

and let $\lambda(s) := \exp(-\frac{1}{2}\left|x_s - m_s\right|_{S_s}^2)$. It holds that

$$\begin{aligned}\widetilde{g}_1(x_1) - \widetilde{g}_0(x_0) = \int_0^1 \frac{\mathrm{d}\lambda}{\mathrm{d}s}(s)\,\mathrm{d}s = -\int_0^1 \Big[(x_1 - x_0 - m_1 + m_0)^\mathsf{T} S_s^{-1}(x_s - m_s) \\ - \frac{1}{2}(x_s - m_s)^\mathsf{T} S_s^{-1}(S_1 - S_0)S_s^{-1}(x_s - m_s)\Big]\lambda(s)\,\mathrm{d}s.\end{aligned}$$

Therefore, using that $\max_{z\in\mathbf{R}}|z\mathrm{e}^{-\frac{z^2}{2}}| = \frac{1}{\sqrt{\mathrm{e}}}$ and $\max_{z\in\mathbf{R}}|z^2\mathrm{e}^{-\frac{z^2}{2}}| = \frac{2}{\mathrm{e}}$, we deduce that

$$\begin{aligned}\left|\widetilde{g}_1(x_1) - \widetilde{g}_0(x_0)\right| &\leqslant \int_0^1 \Big[|x_1 - x_0 - m_1 + m_0|_{S_s}\,|x_s - m_s|_{S_s} \\ &\quad + \frac{1}{2}\left\|S_s^{-\frac{1}{2}}(S_1 - S_0)S_s^{-\frac{1}{2}}\right\|\,|x_s - m_s|_{S_s}^2\Big]\lambda(s)\,\mathrm{d}s \\ &\leqslant \sqrt{\frac{\alpha}{\mathrm{e}}}|x_1 - x_0 - m_1 + m_0| + \frac{\alpha}{\mathrm{e}}\|S_1 - S_0\|.\end{aligned}$$

Applying the triangle inequality

$$\left|g_0(x_0) - g_1(x_1)\right| \leqslant \frac{\left|\widetilde{g}_0(x_0) - \widetilde{g}_1(x_1)\right|}{\sqrt{(2\pi)^d \det S_0}} + \left|\frac{1}{\sqrt{(2\pi)^d \det S_0}} - \frac{1}{\sqrt{(2\pi)^d \det S_1}}\right|\left|\widetilde{g}_1(x_1)\right|,$$

we obtain (A.3), which concludes the proof. ∎

LEMMA A.4. *Denote by $g(\bullet; m, S)$ the Lebesgue density of $\mathcal{N}(m, S)$ and by $\mathcal{S}_\alpha^K$ the set of symmetric $K \times K$ matrices M satisfying*

$$\frac{1}{\alpha} I_K \preccurlyeq M \preccurlyeq \alpha I_K. \tag{A.4}$$

Then, for all $\alpha \geqslant 1$, there exists $L_\alpha > 0$ such that for all parameters $(c_1, m_1, S_1) \in \mathbf{R} \times \mathbf{R}^K \times \mathcal{S}_\alpha^K$ and $(c_2, m_2, S_2) \in \mathbf{R} \times \mathbf{R}^K \times \mathcal{S}_\alpha^K$,

$$\|\mathfrak{h}\|_\infty \leqslant L_\alpha \|\mathfrak{h}\|_1, \qquad \mathfrak{h}(y) = c_1 g(y; m_1, S_1) - c_2 g(y; m_2, S_2). \tag{A.5}$$

Remark A.5. When $c_2 = 0$, (A.5) may be viewed as an inverse inequality. It then simply states that the L^∞ norm of the density of a normal random variable is bounded from above by the L^1 norm uniformly for all densities from the set of Gaussians with a covariance matrix satisfying (A.4).

Proof. For conciseness, let $g_1(\bullet) = g(\bullet; m_1, S_1)$ and $g_2(\bullet) = g(\bullet; m_2, S_2)$.

Step 1. *Simplification.* It is sufficient to prove the statement for $c_1 = c_2 = 1$. Indeed, suppose that there is $\widetilde{L}_\alpha > 0$ such that

$$\forall (m_1, S_1) \in \mathbf{R}^K \times \mathcal{S}^K_\alpha, \qquad \forall (m_2, S_2) \in \mathbf{R}^K \times \mathcal{S}^K_\alpha, \qquad \|g_1 - g_2\|_\infty \leqslant \widetilde{L}_\alpha \|g_1 - g_2\|_1. \tag{A.6}$$

Then, by the triangle inequality, it holds that

$$\begin{aligned}\left\|c_1 g_1 - c_2 g_2\right\|_\infty &\leqslant |c_1| \left\|g_1 - g_2\right\|_\infty + |c_1 - c_2| \left\|g_2\right\|_\infty \\ &\leqslant |c_1| \widetilde{L}_\alpha \left\|g_1 - g_2\right\|_1 + |c_1 - c_2| \left\|g_2\right\|_\infty \\ &\leqslant \widetilde{L}_\alpha \left\|c_1 g_1 - c_2 g_2\right\|_1 + |c_2 - c_1| \left\|g_2\right\|_1 + |c_2 - c_1| \left\|g_2\right\|_\infty.\end{aligned}$$

By Jensen's inequality, it holds that

$$|c_1 - c_2| = \left| \int_{\mathbf{R}^d} c_1 g_1(x) - c_2 g_2(x) \, \mathrm{d}x \right| \leqslant \left\|c_1 g_1 - c_2 g_2\right\|_1,$$

and so we deduce that

$$\left\|c_1 g_1 - c_2 g_2\right\|_\infty \leqslant \left\|c_1 g_1 - c_2 g_2\right\|_1 \left(\widetilde{L}_\alpha + 1 + \sqrt{\frac{\alpha^K}{(2\pi)^K}} \right).$$

Furthermore, since both sides of the inequality (A.6) are invariant under translation, it is sufficient to consider the case where $m_2 = 0$, which we do from now on. Finally, note that

$$\begin{aligned} g_1(y) - g_2(y) = \frac{1}{\sqrt{\det S_2}} \Bigg(& g\left(\sqrt{S_2^{-1}} y; \sqrt{S_2^{-1}} m_1, \sqrt{S_2^{-1}} S_1 \sqrt{S_2^{-1}} \right) \\ & - g\left(\sqrt{S_2^{-1}} y; 0, I_K \right) \Bigg), \end{aligned}$$

and so we can also assume without loss of generality that $S_2 = I_K$. Indeed, assume that the inequality (A.6) is satisfied in this particular case with a constant $\widehat{L}_\alpha$. Since

$$\frac{1}{\alpha^2} I_K \preccurlyeq \frac{1}{\alpha \|S_2\|} I_K \preccurlyeq \sqrt{S_2^{-1}} S_1 \sqrt{S_2^{-1}} \preccurlyeq \alpha \|S_2^{-1}\| I_K \preccurlyeq \alpha^2 I_K,$$

we deduce that if $(S_1, S_2) \in \mathcal{S}^K_\alpha \times S^K_\alpha$ for some $\alpha > 0$, then

$$\left(\sqrt{S_2^{-1}} S_1 \sqrt{S_2^{-1}}, I_K \right) \in \mathcal{S}^K_{\alpha^2} \times \mathcal{S}^K_{\alpha^2}.$$

Therefore, using the change of variable $y \mapsto \sqrt{S_2^{-1}} y$ together with (A.6) in the particular case $S_2 = I_K$, we obtain that

$$
\begin{aligned}
&\|g(\bullet;m_1,S_1)-g(\bullet;0,S_2)\|_\infty\\
&\quad=\frac{1}{\sqrt{\det S_2}}\left\|g\left(\bullet;\sqrt{S_2^{-1}}m_1,\sqrt{S_2^{-1}}S_1\sqrt{S_2^{-1}}\right)-g\left(\bullet;0,I_K\right)\right\|_\infty\\
&\quad\leqslant\frac{\widehat{L}_{\alpha^2}}{\sqrt{\det S_2}}\left\|g\left(\bullet;\sqrt{S_2^{-1}}m_1,\sqrt{S_2^{-1}}S_1\sqrt{S_2^{-1}}\right)-g\left(\bullet;0,I_K\right)\right\|_1\\
&\quad=\frac{\widehat{L}_{\alpha^2}}{\sqrt{\det S_2}}\|g(\bullet;m_1,S_1)-g(\bullet;0,S_2)\|_1\leqslant\alpha^{\frac{K}{2}}\widehat{L}_{\alpha^2}\|g(\bullet;m_1,S_1)-g(\bullet;0,S_2)\|_1,
\end{aligned}
$$

and so the bound (A.6) is valid in general with $\widetilde{L}_\alpha=\alpha^{\frac{K}{2}}\widehat{L}_{\alpha^2}$.

Step 2. *Proof of the simplified statement.* It remains to show that there exists for all $\alpha>0$ a constant $\widehat{L}_\alpha$ such that the following inequality holds for all $(m,S)\in\mathbf{R}^K\times\mathcal{S}^K_\alpha$:

$$
\|\mathfrak{h}\|_\infty\leqslant\widehat{L}_\alpha\|\mathfrak{h}\|_1,\qquad\mathfrak{h}(y)=g(y;m,S)-g(y;0,I_K).\tag{A.7}
$$

To this end, fix $\alpha>0$, fix $\varepsilon\in(0,1)$, and assume first that $\|S-I_K\|\geqslant\varepsilon$. By the lower bound on the total variation distance between Gaussians in [17, Proposition 2.2], we have that

$$
\frac{1}{2}\|\mathfrak{h}\|_1\geqslant1-\frac{\det S^{1/4}\det I_K^{1/4}}{\det\left(\frac{S+I_K}{2}\right)^{1/2}}\exp\left(-\frac{1}{8}m^\mathsf{T}\left(\frac{S+I_K}{2}\right)^{-1}m\right).
$$

In particular, it holds that

$$
\frac{1}{2}\|\mathfrak{h}\|_1\geqslant1-\frac{\det S^{1/4}\det I_K^{1/4}}{\det\left(\frac{S+I_K}{2}\right)^{1/2}}=1-\prod_{i=1}^K\sqrt{\frac{\sqrt{\lambda_i}}{\frac{\lambda_i+1}{2}}}\geqslant1-\sqrt{\frac{2\sqrt{1+\varepsilon}}{2+\varepsilon}},\tag{A.8}
$$

where $(\lambda_i)_{i=1}^K$ are the eigenvalues of S. In the last inequality, we used that, since all the terms in the product are bounded from above by 1 by the arithmetic mean-geometric mean inequality and since at least one eigenvalue is not in the interval $(1-\varepsilon,1+\varepsilon)$ given that $\|S-I_K\|\geqslant\varepsilon$, it holds that

$$
\prod_{i=1}^K\sqrt{\frac{\sqrt{\lambda_i}}{\frac{\lambda_i+1}{2}}}\leqslant\max_{|\lambda-1|\geqslant\varepsilon}\sqrt{\frac{\sqrt{\lambda}}{\frac{\lambda+1}{2}}}=\sqrt{\frac{2\sqrt{1+\varepsilon}}{2+\varepsilon}}.
$$

On the other hand, since $S\succcurlyeq\frac{1}{\alpha}I_K$, it holds that $\det S\geqslant\frac{1}{\alpha^K}$, and so

$$
\|\mathfrak{h}\|_\infty\leqslant2\left(\frac{\alpha}{2\pi}\right)^{\frac{K}{2}}.\tag{A.9}
$$

Combining (A.8) and (A.9) gives that

$$
\|\mathfrak{h}\|_1\geqslant C_1\|\mathfrak{h}\|_\infty,\qquad C_1:=\left(1-\sqrt{\frac{2\sqrt{1+\varepsilon}}{2+\varepsilon}}\right)\left(\frac{2\pi}{\alpha}\right)^{\frac{K}{2}}.
$$

Consider now the case where $\|S-I_K\|\leqslant\varepsilon$. Since $S\mapsto\sqrt{\det S}$ is Lipschitz continuous over the set of symmetric positive definite matrices S such that $\|S-I_K\|\leqslant\varepsilon$, with a Lipschitz constant we denote by c_ε, it holds by Lemma A.3 that

$$\|\mathfrak{h}\|_\infty \leqslant (2\pi)^{-\frac{K}{2}} \left(\frac{\alpha^{\frac{1+K}{2}}}{\sqrt{\mathrm{e}}} |m| + \frac{\alpha^{1+\frac{K}{2}}}{\mathrm{e}} \|S - I_K\| + \alpha^K \left| \sqrt{\det S} - \sqrt{\det I_K} \right| \right)$$
$$\leqslant (2\pi)^{-\frac{K}{2}} \left(\frac{\alpha^{\frac{1+K}{2}}}{\sqrt{\mathrm{e}}} |m| + \left(\frac{\alpha^{1+\frac{K}{2}}}{\mathrm{e}} + c_\varepsilon \alpha^K \right) \|S - I_K\| \right).$$

In view of (A.9), this implies that there exists a constant C depending only on (ε, α, K) such that

$$\|\mathfrak{h}\|_\infty \leqslant C\Big(\min\{|m|, 1\} + \|S - I_K\|\Big). \tag{A.10}$$

On the other hand, since the characteristic function of $\mathcal{N}(m, S)$ is given by $u \mapsto \mathrm{e}^{\mathrm{i}m^\mathsf{T}u - \frac{1}{2}u^\mathsf{T}Su}$, it holds by definition of the characteristic function that

$$\forall u \in \mathbf{R}^K, \qquad \mathrm{e}^{\mathrm{i}m^\mathsf{T}u - \frac{1}{2}u^\mathsf{T}Su} - \mathrm{e}^{-\frac{|u|^2}{2}} = \int_{\mathbf{R}^K} \mathrm{e}^{\mathrm{i}u^\mathsf{T}x} \mathfrak{h}(x)\, \mathrm{d}x.$$

Therefore, it holds that

$$\|\mathfrak{h}\|_1 \geqslant \sup_{|u|\leqslant 1} \left| \mathrm{e}^{\mathrm{i}m^\mathsf{T}u - \frac{1}{2}u^\mathsf{T}Su} - \mathrm{e}^{-\frac{|u|^2}{2}} \right| = \sup_{|u|\leqslant 1} \left| \mathrm{e}^{-\frac{1}{2}u^\mathsf{T}Su} - \mathrm{e}^{-\mathrm{i}m^\mathsf{T}u - \frac{|u|^2}{2}} \right|.$$

It is clear from elementary geometry in the complex plane that

$$\forall u \in \mathbf{R}^K, \qquad \left| \mathrm{e}^{-\frac{1}{2}u^\mathsf{T}Su} - \mathrm{e}^{-\mathrm{i}m^\mathsf{T}u - \frac{|u|^2}{2}} \right|$$
$$\geqslant \max\left\{ \left|\sin(m^\mathsf{T}u)\right| \mathrm{e}^{-\frac{|u|^2}{2}}, \left| \mathrm{e}^{-\frac{1}{2}u^\mathsf{T}Su} - \mathrm{e}^{-\frac{|u|^2}{2}} \right| \right\},$$

and so we have

$$\|\mathfrak{h}\|_1 \geqslant \max\left\{ \sup_{|u|\leqslant 1} \left|\sin(m^\mathsf{T}u)\right| \mathrm{e}^{-\frac{|u|^2}{2}}, \sup_{|u|\leqslant 1} \left| \mathrm{e}^{-\frac{1}{2}u^\mathsf{T}Su} - \mathrm{e}^{-\frac{|u|^2}{2}} \right| \right\}$$
$$\geqslant \max\left\{ \mathrm{e}^{-\frac{1}{2}} \sup_{|u|\leqslant 1} \left|\sin(m^\mathsf{T}u)\right|, \frac{\mathrm{e}^{-1}}{2} \sup_{|u|\leqslant 1} \left|u^\mathsf{T}(S - I_K)u\right| \right\}$$
$$= \max\left\{ \mathrm{e}^{-\frac{1}{2}} \sup_{|u|\leqslant 1} \left|\sin(m^\mathsf{T}u)\right|, \frac{\mathrm{e}^{-1}}{2} \|S - I_K\| \right\}.$$

In the second inequality, we used that $u^\mathsf{T}Su \leqslant 1 + \varepsilon \leqslant 2$ for all $|u| \leqslant 1$, together with the elementary inequality $\left|\mathrm{e}^a - \mathrm{e}^b\right| \geqslant \mathrm{e}^{\min\{a,b\}}|a - b|$. To conclude, considering the particular value

$$u = \frac{m}{|m| \max\left\{1, \frac{4|m|}{\pi}\right\}},$$

we obtain that

$$\sup_{|u|\leqslant 1} \left|\sin(m^\mathsf{T}u)\right| \geqslant \sin\Big(\min\left\{|m|, \frac{\pi}{4}\right\}\Big) = \int_0^{\min\left\{|m|, \frac{\pi}{4}\right\}} \cos(t)\, \mathrm{d}t \geqslant \cos\left(\frac{\pi}{4}\right) \min\left\{|m|, \frac{\pi}{4}\right\}.$$

Thus, using that $\max\{A,B\} \geqslant \frac{A}{2}+\frac{B}{2}$, we obtain

$$\begin{aligned}\|\mathfrak{h}\|_1 &\geqslant \max\left\{\mathrm{e}^{-\frac{1}{2}}\cos\left(\frac{\pi}{4}\right)\min\left\{|m|,\frac{\pi}{4}\right\},\frac{\mathrm{e}^{-1}}{2}\|S-I_K\|\right\}\\ &\geqslant \frac{1}{2}\min\left\{\mathrm{e}^{-\frac{1}{2}}\cos\left(\frac{\pi}{4}\right)\frac{\pi}{4},\frac{\mathrm{e}^{-1}}{2}\right\}\Big(\min\{|m|,1\}+\|S-I_K\|\Big). \qquad \text{(A.11)}\end{aligned}$$

Combining (A.11) with (A.10) leads to $\|\mathfrak{h}\|_1 \geqslant C_2\|\mathfrak{h}\|_\infty$, which concludes the proof of the case $\|S-I_K\|\leqslant\varepsilon$. Consequently, the statement (A.7) holds in general with constant $\widehat{L}_\alpha = \min\{C_1,C_2\}^{-1}$. ∎

LEMMA A.6. *Let* P *and* Q *denote the operators on probability measures given, respectively, in* (1.5) *and* (1.7). *Suppose that Assumption* A *is satisfied and that* $|h|_{C^{0,1}}\leqslant \ell_h<\infty$. *Then there is* $L=L(\kappa_\Psi,\kappa_h,\ell_h,\Sigma,\Gamma)$ *such that for all* $(u_1,u_2,y)\in\mathbf{R}^d\times\mathbf{R}^d\times\mathbf{R}^K$ *and all* $\mu\in\mathcal{P}(\mathbf{R}^d)$, *the density of* $p=\mathsf{QP}\mu$ *satisfies*

$$|p(u_1,y)-p(u_2,y)|\leqslant L|u_1-u_2|\exp\left(-\frac{1}{4}\left(\min\Big\{|u_1|_\Sigma^2,|u_2|_\Sigma^2\Big\}+|y|_\Gamma^2\right)\right). \qquad \text{(A.12)}$$

Proof. Throughout this proof, C denotes a constant whose value is irrelevant in the context, depends only on $\kappa_\Psi,\kappa_h,\ell_h,\Sigma,\Gamma$, and may change from line to line. Since the function $g(x):=\mathrm{e}^{-x^2}$ has derivative $-2x\mathrm{e}^{-x^2}$ and since $|x\mathrm{e}^{-x^2}|\leqslant\mathrm{e}^{-\frac{2x^2}{3}}$ for all $x\in\mathbf{R}$, it holds for all $(a,b)\in\mathbf{R}^2$ that there is ξ between $|a|$ and $|b|$ such that

$$\left|\mathrm{e}^{-a^2}-\mathrm{e}^{-b^2}\right| = |b-a|\,|g'(\xi)|\leqslant 2|b-a|\left(\mathrm{e}^{-\frac{2a^2}{3}}+\mathrm{e}^{-\frac{2b^2}{3}}\right). \qquad \text{(A.13)}$$

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

$$\begin{aligned}&\left|\exp\left(-\frac{1}{2}|u_1-\Psi(v)|_\Sigma^2\right)-\exp\left(-\frac{1}{2}|u_2-\Psi(v)|_\Sigma^2\right)\right|\\ &\qquad\leqslant C|u_2-u_1|_\Sigma\left(\exp\left(-\frac{1}{3}|u_1-\Psi(v)|_\Sigma^2\right)+\exp\left(-\frac{1}{3}|u_2-\Psi(v)|_\Sigma^2\right)\right).\end{aligned}$$

Integrating out the v variable with respect to μ and using the equivalence of norms, we obtain that

$$\begin{aligned}|\mathsf{P}\mu(u_1)-\mathsf{P}\mu(u_2)|\leqslant{}& C|u_1-u_2|\int_{\mathbf{R}^d}\exp\left(-\frac{1}{3}|u_1-\Psi(v)|_\Sigma^2\right)\mu(\mathrm{d}v)\\ &+C|u_1-u_2|\int_{\mathbf{R}^d}\exp\left(-\frac{1}{3}|u_2-\Psi(v)|_\Sigma^2\right)\mu(\mathrm{d}v).\end{aligned}$$

By Young's inequality, it holds for all $\delta>0$ that

$$\forall(a,b)\in\mathbf{R}^d\times\mathbf{R}^d,\qquad |a-b|_\Sigma^2\geqslant\frac{1}{1+\delta}|a|_\Sigma^2-\frac{1}{\delta}|b|_\Sigma^2. \qquad \text{(A.14)}$$

Using this inequality with $\delta=\frac{1}{3}$ together with the assumption that Ψ is bounded, we deduce that

$$
\begin{aligned}
(A.15)\qquad |\mathsf{P}\mu(u_1) - \mathsf{P}\mu(u_2)| &\leqslant C|u_1-u_2| \int_{\mathbf{R}^d} \exp\left(-\frac{1}{4}|u_1|_\Sigma^2\right)\mu(\mathrm{d}v) \\
&\quad + C|u_1-u_2| \int_{\mathbf{R}^d} \exp\left(-\frac{1}{4}|u_2|_\Sigma^2\right)\mu(\mathrm{d}v) \\
&\leqslant 2C|u_1-u_2| \exp\left(-\frac{1}{4}\min\left\{|u_1|_\Sigma^2, |u_2|_\Sigma^2\right\}\right).
\end{aligned}
$$

Next, letting $h_s = (1-s)h(u_1) + sh(u_2)$ and using a reasoning similar to that in the proof of Lemma A.3, we obtain the following inequalities, which hold for all $(u_1,u_2,y) \in \mathbf{R}^d \times \mathbf{R}^d \times \mathbf{R}^K$:

$$
\begin{aligned}
(A.16)\qquad &\Big|\mathcal{N}\big(h(u_1),\Gamma\big)(y) - \mathcal{N}\big(h(u_2),\Gamma\big)(y)\Big| \\
&\quad \leqslant |h(u_2)-h(u_1)|_\Gamma \int_0^1 |h_s - y|_\Gamma \exp\left(-\frac{1}{2}|h_s-y|_\Gamma^2\right)\mathrm{d}s \\
&\quad \leqslant C|u_2-u_1| \exp\left(-\frac{1}{4}|y|_\Gamma^2\right),
\end{aligned}
$$

where we used the Lipschitz continuity of h, together with (A.14) and the boundedness of h, in the last inequality. In order to conclude the proof, using the definition of P, we calculate that

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

The first and second terms on the right-hand side can be bounded by using (A.15) and (A.16), respectively, leading to (A.12). ∎

Appendix B. Technical results for Theorems 2.2 and 3.2. We show moment bounds in Appendix B.1, and we prove that $\mathsf{T}_j\mu = \mathsf{B}_j\mu$ for any Gaussian probability measure μ in Appendix B.2. Finally, we prove the stability results used in the proofs of Theorems 2.2 and 3.2 in Appendices B.3 and B.4, respectively.

B.1. Moment bounds.

LEMMA B.1 (moment bounds). *Let μ denote a probability measure on $\mathbf{R}^d$. Under Assumption* A*, it holds that*

$$
(B.1)\qquad \big|\mathcal{M}(\mathsf{P}\mu)\big| \leqslant \kappa_\Psi, \qquad \Sigma \preccurlyeq \mathcal{C}(\mathsf{P}\mu) \preccurlyeq \kappa_\Psi^2 I_d + \Sigma.
$$

Proof. From the definition of P in (1.5), we have that

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

where the last equality is obtained by changing the order of integration using Fubini's theorem. Using the first item in Assumption A, we then deduce the first inequality in (B.1). For the second inequality in (B.1), we first note the following inequality, which holds for any $m, v \in \mathbf{R}^d$ by Lemma A.1:

$$\int_{\mathbf{R}^d} (u-m)\otimes(u-m)\exp\left(-\frac{1}{2}\,|u-\Psi(v)|_\Sigma^2\right)\mathrm{d}u \tag{B.2}$$
$$\succcurlyeq \int_{\mathbf{R}^d} \big(u-\Psi(v)\big)\otimes\big(u-\Psi(v)\big)\exp\left(-\frac{1}{2}\,|u-\Psi(v)|_\Sigma^2\right)\mathrm{d}u.$$

The result follows by using the fact that $\Psi(v)$ is the mean under Gaussian $\mathcal{N}\big(\Psi(v),\Sigma\big)$. Now choose m to be the mean under measure $\mathsf{P}\mu$, and note that by conditioning on v and using (B.2),

$$\begin{aligned}
\mathcal{C}(\mathsf{P}\mu) &= \int_{\mathbf{R}^d} (u-m)\otimes(u-m)\,\mathsf{P}\mu(u)\,\mathrm{d}u \\
&= \frac{1}{\sqrt{(2\pi)^d\det\Sigma}}\int_{\mathbf{R}^d}\left(\int_{\mathbf{R}^d}(u-m)\otimes(u-m)\exp\Big(-\frac{1}{2}\,|u-\Psi(v)|_\Sigma^2\Big)\,\mathrm{d}u\right)\mu(\mathrm{d}v) \\
&\succcurlyeq \frac{1}{\sqrt{(2\pi)^d\det\Sigma}}\int_{\mathbf{R}^d}\left(\int_{\mathbf{R}^d}\big(u-\Psi(v)\big)\otimes\big(u-\Psi(v)\big)\right. \\
&\qquad \left.\exp\Big(-\frac{1}{2}\,|u-\Psi(v)|_\Sigma^2\Big)\,\mathrm{d}u\right)\mu(\mathrm{d}v) \\
&= \int_{\mathbf{R}^d}\Sigma\,\mu(\mathrm{d}v) = \Sigma,
\end{aligned}$$

and so $\mathcal{C}(\mathsf{P}\mu)\succcurlyeq\Sigma$. On the other hand, using Lemma A.1 again together with the fact that by the Cauchy–Schwarz inequality $aa^\mathsf{T}\preccurlyeq(a^\mathsf{T}a)I_d$ for any vector $a\in\mathbf{R}^d$, we have

$$\begin{aligned}
\mathcal{C}(\mathsf{P}\mu) &\preccurlyeq \int_{\mathbf{R}^d} u\otimes u\,\mathsf{P}\mu(u)\,\mathrm{d}u \\
&= \frac{1}{\sqrt{(2\pi)^d\det\Sigma}}\int_{\mathbf{R}^d}\int_{\mathbf{R}^d} u\otimes u\exp\left(-\frac{1}{2}\,|u-\Psi(v)|_\Sigma^2\right)\mu(\mathrm{d}v)\,\mathrm{d}u \\
&= \int_{\mathbf{R}^d}\big(\Psi(v)\otimes\Psi(v)+\Sigma\big)\,\mu(\mathrm{d}v)\preccurlyeq\kappa_\Psi^2\,I_d+\Sigma, \qquad \text{(B.3)}
\end{aligned}$$

which concludes the proof. ◻

It is possible, by using a similar reasoning, to obtain bounds on the moments of $\mathsf{QP}\mu$.

Lemma B.2. *Let μ denote a probability measure on $\mathbf{R}^d$. Under Assumption* A, *it holds that*

$$\big|\mathcal{M}(\mathsf{QP}\mu)\big|\leqslant\sqrt{\kappa_\Psi^2+\kappa_h^2} \tag{B.4}$$

and

$$\min\left\{\frac{\gamma\sigma}{2\kappa_h^2+\gamma},\frac{\gamma}{2}\right\}I_{d+K}\preccurlyeq\mathcal{C}(\mathsf{QP}\mu)\preccurlyeq\begin{pmatrix}2\kappa_\Psi^2 I_d+2\Sigma & 0_{d\times K}\\ 0_{K\times d} & 2\kappa_h^2 I_K+\Gamma\end{pmatrix}. \tag{B.5}$$

Proof. The inequality (B.4) follows immediately from Assumption A and the fact that

$$\mathcal{M}(\mathsf{QP}\mu)=\begin{pmatrix}\mathcal{M}(\mathsf{P}\mu)\\ \mathsf{P}\mu[h]\end{pmatrix}. \tag{B.6}$$

For inequality (B.5), let $\phi\colon \mathbf{R}^d \to \mathbf{R}^{d+K}$ denote the map $\phi(u) = \bigl(u, h(u)\bigr)$, and let $\phi_\sharp$ denote the associated pushforward map on measures. A calculation gives

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

For any $(a,b) \in \mathbf{R}^d \times \mathbf{R}^K$, it holds that

$$2\begin{pmatrix} aa^\mathsf{T} & 0_{d\times K} \\ 0_{K\times d} & bb^\mathsf{T} \end{pmatrix} - \begin{pmatrix} a \\ b \end{pmatrix} \otimes \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} a \\ -b \end{pmatrix} \otimes \begin{pmatrix} a \\ -b \end{pmatrix} \succcurlyeq 0_{(d+K)\times(d+K)}.$$

Therefore, we obtain

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

which enables to deduce, using Lemma B.1 and Assumption A, that

$$\begin{aligned} \mathcal{C}(\phi_\sharp \mathsf{P}\mu) &\preccurlyeq \int_{\mathbf{R}^d} \begin{pmatrix} u \\ h(u) \end{pmatrix} \otimes \begin{pmatrix} u \\ h(u) \end{pmatrix} \mathsf{P}\mu(u)\,\mathrm{d}u \\ &\preccurlyeq 2\int_{\mathbf{R}^d} \begin{pmatrix} uu^\mathsf{T} & 0_{d\times K} \\ 0_{K\times d} & h(u)h(u)^\mathsf{T} \end{pmatrix} \mathsf{P}\mu(u)\,\mathrm{d}u \preccurlyeq 2\begin{pmatrix} \kappa_\Psi^2 I_d + \Sigma & 0_{d\times K} \\ 0_{K\times d} & \kappa_h^2 I_K \end{pmatrix}, \end{aligned} \tag{B.8}$$

where we used (B.3) in the last inequality. Combined with (B.7), this inequality leads to the upper bound (B.5). For the lower bound, note that by the Cauchy–Schwarz inequality, it holds for any probability measure $\pi \in \mathcal{P}(\mathbf{R}^d \times \mathbf{R}^K)$ and all $(a,b) \in \mathbf{R}^d \times \mathbf{R}^K$ that

$$\begin{aligned} \left|a^\mathsf{T} \mathcal{C}^{uy}(\pi) b\right| &= \int_{\mathbf{R}^d\times\mathbf{R}^K} \bigl(a^\mathsf{T}\bigl(u - \mathcal{M}^u(\pi)\bigr)\bigr)\bigl(b^\mathsf{T}\bigl(y - \mathcal{M}^y(\pi)\bigr)\bigr)\,\pi(\mathrm{d}u\mathrm{d}y) \\ &\leqslant \sqrt{a^\mathsf{T}\mathcal{C}^{uu}(\pi)a}\sqrt{b^\mathsf{T}\mathcal{C}^{yy}(\pi)b}. \end{aligned}$$

Therefore, by Young's inequality, it holds for all $\varepsilon \in (0,1)$ and for all $(a,b) \in \mathbf{R}^d \times \mathbf{R}^K$ that

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

where we employed (B.1) and the bound $\mathcal{C}^{yy}(\phi_\sharp \mathsf{P}\mu) \preccurlyeq \kappa_H^2 I_K$ in the last inequality. Using (B.7), we deduce that

$$\begin{aligned} \begin{pmatrix} a \\ b \end{pmatrix}^\mathsf{T} \mathcal{C}(\mathsf{QP}\mu) \begin{pmatrix} a \\ b \end{pmatrix} &\geqslant (1-\varepsilon) a^\mathsf{T}\Sigma a - \left(\frac{1}{\varepsilon} - 1\right)\kappa_h^2 |b|^2 + b^\mathsf{T}\Gamma b \\ &\geqslant (1-\varepsilon)\sigma |a|^2 + \left(\gamma - \left(\frac{1}{\varepsilon} - 1\right)\kappa_h^2\right)|b|^2. \end{aligned} \tag{B.9}$$

Letting ε be such that the coefficient of $|b|^2$ is $\gamma/2$, we finally obtain

$$\begin{pmatrix} a \\ b \end{pmatrix}^\mathsf{T} \mathcal{C}(\mathsf{QP}\mu) \begin{pmatrix} a \\ b \end{pmatrix} \geqslant \frac{\gamma\sigma}{2\kappa_h^2 + \gamma}|a|^2 + \frac{\gamma}{2}|b|^2, \tag{B.10}$$

which concludes the proof. ∎

Remark B.3. A bound sharper than (B.10) can be obtained by letting ε be such that the coefficients of $|a|^2$ and $|b|^2$ are equal in (B.9), but this is not necessary for our purposes.

LEMMA B.4. *For $\mu_1, \mu_2 \in \mathcal{P}(\mathbf{R}^n)$ with finite second moments, it holds that*

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

Proof. Let $m_i = \mathcal{M}(\mu_i)$ and $\Sigma_i = \mathcal{C}(\mu_i)$ for $i = 1, 2$. Notice that $|2a^\mathsf{T} u| \leqslant g(u)$ if $|a| = 1$, so

$$|m_1 - m_2| = \sup_{|a|=1} \left|a^\mathsf{T}\left(m_1 - m_2\right)\right| = \sup_{|a|=1} \left|\mu_1\left[a^\mathsf{T} u\right] - \mu_2\left[a^\mathsf{T} u\right]\right| \leqslant \frac{1}{2} d_g(\mu_1,\mu_2), \tag{B.11}$$

where the supremum is over the unit sphere in $\mathbf{R}^n$, centered at the origin and in the Euclidean distance. Similarly,

$$\begin{aligned} \|\Sigma_1 - \Sigma_2\| &= \sup_{|a|=1} \left|a^\mathsf{T}\Sigma_1 a - a^\mathsf{T}\Sigma_2 a\right| \\ &\leqslant \sup_{|a|=1} \left\{ \left|\mu_1\left[|a^\mathsf{T} u|^2\right] - \mu_2\left[|a^\mathsf{T} u|^2\right]\right| + \left|\mu_1\left[a^\mathsf{T} u\right]^2 - \mu_2\left[a^\mathsf{T} u\right]^2\right| \right\} \\ &\leqslant d_g(\mu_1,\mu_2) + \sup_{|a|=1} \left|\mu_1\left[a^\mathsf{T} u\right] + \mu_2\left[a^\mathsf{T} u\right]\right| \left|\mu_1\left[a^\mathsf{T} u\right] - \mu_2\left[a^\mathsf{T} u\right]\right| \\ &\leqslant \left(1 + \frac{1}{2}|m_1 + m_2|\right) d_g(\mu_1,\mu_2), \end{aligned} \tag{B.12}$$

which concludes the proof. ◻

B.2. Action of T_j on Gaussians.

LEMMA B.5 ($\mathsf{B}_j\mathsf{G} = \mathsf{T}_j\mathsf{G}$). *Fix $y^\dagger_{j+1} \in \mathbf{R}^K$. Let π be a Gaussian measure over $\mathbf{R}^d \times \mathbf{R}^K$ with mean and covariance given by*

$$m = \begin{pmatrix} m_u \\ m_y \end{pmatrix}, \qquad S = \begin{pmatrix} S_{uu} & S_{uy} \\ S_{uy}^\mathsf{T} & S_{yy} \end{pmatrix}.$$

Recall the pushforward of $\mathscr{T}$ defined in (2.3)*. Then the probability measure $\mathsf{B}_j\pi$ defined in* (1.8) *coincides with the probability measure $\mathsf{T}_j\pi = \mathscr{T}(\bullet, \bullet; \pi, y^\dagger_{j+1})_\sharp \pi$.*

Proof. For conciseness, we denote $y^\dagger = y^\dagger_{j+1}$. Using the well-known formula for the conditional distribution of a normal random variable, we have that

$$\mathsf{B}_j\pi = \mathcal{N}\left(m_u + S_{uy}S_{yy}^{-1}(y^\dagger - m_y), S_{uu} - S_{uy}S_{yy}^{-1}S_{uy}^\mathsf{T}\right). \tag{B.13}$$

Since T_j is the pushforward under an affine map, it maps Gaussian distributions in $\mathcal{P}(\mathbf{R}^d \times \mathbf{R}^K)$ to Gaussian distributions in $\mathcal{P}(\mathbf{R}^d)$ and so is sufficient to check that the first and second moments of $\mathsf{B}_j\pi$ and $\mathsf{T}_j\pi$ coincide. By definition of T_j in (2.3), we have that $U + S_{uy}S_{yy}^{-1}(y^\dagger - Y) \sim \mathsf{T}_j\pi$ if $(U, Y) \sim \pi$. It follows immediately that the mean under $\mathsf{T}_j\pi$ coincides with that under the conditional distribution (B.13). Employing the expression for the mean under $\mathsf{T}_j\pi$, we then obtain that the covariance under $\mathsf{T}_j\pi$ is given by

$$\mathcal{C}(\mathsf{T}_j\pi) = \mathbf{E}_{(U,Y)\sim\pi}\left[\left(U - m_u + S_{uy}S_{yy}^{-1}\left(m_y - Y\right)\right) \otimes \left(U - m_u + S_{uy}S_{yy}^{-1}\left(m_y - Y\right)\right)\right].$$

Developing this expression, we obtain

$$\begin{aligned}\mathcal{C}(\mathsf{T}_j\pi) &= \mathbf{E}_{(U,Y)\sim\pi}\Big[(U-m_u)(U-m_u)^\mathsf{T} + (U-m_u)\big(m_y-Y\big)^\mathsf{T} S_{yy}^{-1}S_{uy}^\mathsf{T} \\ &\qquad + S_{uy}S_{yy}^{-1}\big(m_y-Y\big)(U-m_u)^\mathsf{T} + S_{uy}S_{yy}^{-1}\big(m_y-Y\big)\big(m_y-Y\big)^\mathsf{T} S_{yy}^{-1}S_{uy}^\mathsf{T}\Big] \\ &= S_{uu} - S_{uy}S_{yy}^{-1}S_{uy}^\mathsf{T} - S_{uy}S_{yy}^{-1}S_{uy}^\mathsf{T} + S_{uy}S_{yy}^{-1}S_{uy}^\mathsf{T} = S_{uu} - S_{uy}S_{yy}^{-1}S_{uy}^\mathsf{T},\end{aligned}$$

which indeed coincides with the covariance of $\mathsf{B}_j\pi$ in (B.13). ∎

LEMMA B.6. *The maps* (3.1) *and* (3.2) *are equivalent.*

Proof. The equivalence follows from Lemma B.5 and the operator equality $\mathsf{T}_j\mathsf{G} = \mathsf{G}\mathsf{T}_j$, with both sides viewed as operators from $\mathcal{P}(\mathbf{R}^d\times\mathbf{R}^K)$ to $\mathcal{P}(\mathbf{R}^d)$. Since T_j maps Gaussians to Gaussians, the image of both operators is contained in the set $\mathcal{G}(\mathbf{R}^d)$ of Gaussian distributions. It is therefore sufficient to check that for any $\pi\in\mathcal{P}(\mathbf{R}^d\times\mathbf{R}^K)$, the probability measures $\mathsf{T}_j\mathsf{G}\pi$ and $\mathsf{G}\mathsf{T}_j\pi$ have the same first and second moments. We saw in the proof of Lemma B.5 that the first and second moments of T_jp, for any $p\in\mathcal{P}(\mathbf{R}^d\times\mathbf{R}^K)$, depend only on the first and second moments of p. Therefore, the first and second moments of $\mathsf{T}_j\pi$ and $\mathsf{T}_j\mathsf{G}\pi$ coincide since the operator G leaves the first and second moments invariant, and the conclusion follows. ∎

B.3. Stability results.

LEMMA B.7 (the map P is globally Lischitz). *Under Assumption* A*, it holds for all* $\mu\in\mathcal{P}(\mathbf{R}^d)$ *that*

$$d_g(\mathsf{P}\mu,\mathsf{P}\nu) \leqslant \Big(1+\kappa_\Psi^2+\operatorname{tr}(\Sigma)\Big)\, d_g(\mu,\nu).$$

Proof. By definition (1.5) of P, it holds that

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

Take any $f\colon\mathbf{R}^d\to\mathbf{R}$ such that $|f|\leqslant g$, where $g(u)=1+|u|^2$. Since Ψ is bounded by assumption,

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

Therefore, using Fubini's theorem, we have

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

which concludes the proof. ∎

LEMMA B.8 (the map Q is globally Lipschitz). *Under Assumption* A*, it holds for any* $\mu,\nu\in\mathcal{P}(\mathbf{R}^d)$ *that*

$$d_g(\mathsf{Q}\mu,\mathsf{Q}\nu) \leqslant \Big(1+\kappa_h^2+\operatorname{tr}(\Gamma)\Big)d_g(\mu,\nu). \tag{B.14}$$

Proof. Let us take $f\colon \mathbf{R}^d \times \mathbf{R}^K \to \mathbf{R}$ such that $|f| \leqslant g$ and introduce the operator Π given by

$$\Pi f(u) = \int_{\mathbf{R}^K} f(u,y)\mathcal{N}\big(h(u),\Gamma\big)(\mathrm{d}y).$$

Clearly, $|\Pi f(u)| \leqslant \Pi g(u)$, and it holds that

$$\begin{aligned}\Pi g(u) &= \int_{\mathbf{R}^K} \left(1+|u|^2+|y|^2\right) \mathcal{N}\big(h(u),\Gamma\big)(\mathrm{d}y) = 1+|u|^2+\int_{\mathbf{R}^K}|y|^2\mathcal{N}\big(h(u),\Gamma\big)(\mathrm{d}y)\\ &= 1+|u|^2+\big|h(u)\big|^2+\operatorname{tr}(\Gamma) \leqslant \Big(1+\kappa_h^2+\operatorname{tr}(\Gamma)\Big)\left(1+|u|^2\right).\end{aligned}$$

Therefore,

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

and we obtain (B.14). ◻

Lemma B.9. *Under Assumption* A*, there exists* $C_{\mathsf{B}} = C_{\mathsf{B}}(\kappa_y,\kappa_\Psi,\kappa_h,\Sigma,\Gamma)$ *such that for any probability measure* $\mu \in \mathcal{P}(\mathbf{R}^d)$*, it holds 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. For conciseness, we denote $y^\dagger = y^\dagger_{j+1}$. Let us introduce the y-marginal densities

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

Then

$$\begin{aligned}d_g(\mathsf{B}_j\mathsf{GQP}\mu, \mathsf{B}_j\mathsf{QP}\mu) &= \int_{\mathbf{R}^d}\left(1+|u|^2\right)\left|\frac{\mathsf{GQP}\mu(u,y^\dagger)}{\alpha_\mu(y^\dagger)} - \frac{\mathsf{QP}\mu(u,y^\dagger)}{\beta_\mu(y^\dagger)}\right|\mathrm{d}u\\ &\leqslant \frac{1}{\alpha_\mu(y^\dagger)}\int_{\mathbf{R}^d}\left(1+|u|^2\right)\left|\mathsf{GQP}\mu(u,y^\dagger)-\mathsf{QP}\mu(u,y^\dagger)\right|\mathrm{d}u\\ &\quad+\left|\frac{\alpha_\mu(y^\dagger)-\beta_\mu(y^\dagger)}{\alpha_\mu(y^\dagger)\beta_\mu(y^\dagger)}\right|\int_{\mathbf{R}^d}\left(1+|u|^2\right)\mathsf{QP}\mu(u,y^\dagger)\,\mathrm{d}u.\end{aligned} \tag{B.15}$$

Step 1: *Bounding* $\alpha_\mu(y^\dagger)$ *and* $\beta_\mu(y^\dagger)$ *from below.* The marginal distribution $\alpha_\mu(\bullet)$ is Gaussian with covariance matrix

$$\Gamma + \mathsf{P}\mu\big[h(\bullet)\otimes h(\bullet)\big] - \mathsf{P}\mu\big[h(\bullet)\big]\otimes\mathsf{P}\mu\big[h(\bullet)\big], \tag{B.16}$$

which is bounded from below by Γ and from above by $\Gamma + \kappa_h^2 I_K$ in view of Assumption A. (We again use the fact that $aa^\mathsf{T} \preccurlyeq (a^\mathsf{T}a)I_d$ for any vector $a\in\mathbf{R}^d$ by the Cauchy–Schwarz inequality.) Assumption A also implies that the mean of α_μ is bounded from above in norm by κ_h. Therefore, it holds that

$$\forall y \in \mathbf{R}^K, \qquad \alpha_\mu(y) \geqslant \frac{\exp\Big(-\frac{1}{2}(|y|+\kappa_h)^2\left\|\Gamma^{-1}\right\|\Big)}{\sqrt{(2\pi)^K\det(\Gamma+\kappa_h^2 I_K)}}. \tag{B.17}$$

The function β_μ can be bounded from below independently of μ in a similar manner. Indeed, it holds under Assumption A that for all $y \in \mathbf{R}^K$,

$$\beta_\mu(y) = \int_{\mathbf{R}^d} \mathsf{QP}\mu(u,y)\,\mathrm{d}u = \int_{\mathbf{R}^d} \frac{\exp\Big(-\frac{1}{2}\big(y-h(u)\big)^\mathsf{T}\Gamma^{-1}\big(y-h(u)\big)\Big)}{\sqrt{(2\pi)^K\det(\Gamma)}}\mathsf{P}\mu(u)\,\mathrm{d}u$$

$$\geqslant \frac{\exp\Big(-\frac{1}{2}(|y|+\kappa_h)^2\left\|\Gamma^{-1}\right\|\Big)}{\sqrt{(2\pi)^K\det(\Gamma)}}. \tag{B.18}$$

Step 2*: Bounding the first term in* (B.15). For fixed $u \in \mathbf{R}^d$, the functions $y \mapsto \mathsf{QP}\mu(u,y)$ and $y \mapsto \mathsf{GQP}\mu(u,y)$ are Gaussians up to constant factors. The covariance matrix of the former is Γ, and using the formula for the covariance of the conditional distribution of a Gaussian, we calculate that the covariance of the latter is given by

$$\mathcal{C}^{yy}(\mathsf{QP}\mu) - \mathcal{C}^{yu}(\mathsf{QP}\mu)\mathcal{C}^{uu}(\mathsf{QP}\mu)^{-1}\mathcal{C}^{uy}(\mathsf{QP}\mu). \tag{B.19}$$

Since $\mathcal{C}^{yu}(\mathsf{QP}\mu)\mathcal{C}^{uu}(\mathsf{QP}\mu)^{-1}\mathcal{C}^{uy}(\mathsf{QP}\mu)$ is positive semidefinite, it follows from Lemma B.2 that the matrix (B.19) is bounded from above by $2\kappa_h^2 I_K + \Gamma$. Then, using the same notation as in (B.7), we obtain that

$$\begin{aligned}&\mathcal{C}^{yy}(\mathsf{QP}\mu) - \mathcal{C}^{yu}(\mathsf{QP}\mu)\mathcal{C}^{uu}(\mathsf{QP}\mu)^{-1}\mathcal{C}^{uy}(\mathsf{QP}\mu)\\ &\quad= \Gamma + \Big(\mathcal{C}^{yy}(\phi_\sharp\mathsf{P}\mu) - \mathcal{C}^{yu}(\phi_\sharp\mathsf{P}\mu)\mathcal{C}^{uu}(\phi_\sharp\mathsf{P}\mu)^{-1}\mathcal{C}^{uy}(\phi_\sharp\mathsf{P}\mu)\Big) \succcurlyeq \Gamma,\end{aligned} \tag{B.20}$$

where the inequality holds because, being the Schur complement of the block $\mathcal{C}^{uu}(\phi_\sharp\mathsf{P}\mu)$ of the matrix $\mathcal{C}(\phi_\sharp\mathsf{P}\mu)$, the bracketed term is positive semidefinite. Therefore, the matrix (B.19) is bounded from below by Γ, and so the integral in the first term of (B.15) may be bounded from above by using Lemma A.4 in Appendix A with parameter $\alpha = \alpha(\kappa_h,\Gamma)$, which gives

$$\begin{aligned}&\int_{\mathbf{R}^d}\left(1+|u|^2\right)|\mathsf{GQP}\mu(u,y) - \mathsf{QP}\mu(u,y)|\,\mathrm{d}u\\ &\quad\leqslant C\int_{\mathbf{R}^K}\int_{\mathbf{R}^d}\left(1+|u|^2\right)\big|\mathsf{GQP}\mu(u,y) - \mathsf{QP}\mu(u,y)\big|\,\mathrm{d}u\,\mathrm{d}y\\ &\quad\leqslant C\int_{\mathbf{R}^K}\int_{\mathbf{R}^d}\left(1+|u|^2+|y|^2\right)\big|\mathsf{GQP}\mu(u,y) - \mathsf{QP}\mu(u,y)\big|\,\mathrm{d}u\,\mathrm{d}y\\ &\quad= Cd_g(\mathsf{GQP}\mu,\mathsf{QP}\mu)\end{aligned} \tag{B.21}$$

for an appropriate constant C depending only on Γ and κ_h.

Step 3*: Bounding the second term in* (B.15). By (B.21), we have

$$\begin{aligned}\big|\alpha_\mu(y) - \beta_\mu(y)\big| &= \left|\int_{\mathbf{R}^d}\mathsf{GQP}\mu(u,y) - \mathsf{QP}\mu(u,y)\,\mathrm{d}u\right|\\ &\leqslant \int_{\mathbf{R}^d}\left(1+|u|^2\right)\big|\mathsf{GQP}\mu(u,y) - \mathsf{QP}\mu(u,y)\big|\,\mathrm{d}u \leqslant Cd_g(\mathsf{GQP}\mu,\mathsf{QP}\mu).\end{aligned}$$

On the other hand, since $y \mapsto \mathsf{QP}\mu(u,y)/\mathsf{P}\mu(u)$ is a Gaussian density with covariance Γ, which is bounded uniformly from above by $\big((2\pi)^K\det(\Gamma)\big)^{-1/2}$, it holds that

$$
\begin{aligned}
\int_{\mathbf{R}^d} \left(1+|u|^2\right) \mathsf{QP}\mu(u,y)\,\mathrm{d}u &= \beta_\mu(y) + \int_{\mathbf{R}^d} |u|^2 \mathsf{QP}\mu(u,y)\,\mathrm{d}u \\
&\leqslant \beta_\mu(y) + \int_{\mathbf{R}^d} \frac{|u|^2 \mathsf{P}\mu(u)}{\sqrt{(2\pi)^K \det(\Gamma)}}\,\mathrm{d}u \\
&= \beta_\mu(y) + \frac{\operatorname{tr}\bigl(\mathcal{C}(\mathsf{P}\mu)\bigr) + |\mathcal{M}(\mathsf{P}\mu)|^2}{\sqrt{(2\pi)^K \det(\Gamma)}}.
\end{aligned}
$$

By Lemma B.1, the first and second moments of $\mathsf{P}\mu$ are bounded from above by a constant depending only on κ_Ψ and Σ.

Step 4: Concluding the proof. Combining the above inequalities, we deduce that

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

Using (B.17) and (B.18) together with the uniform bound κ_y on the data (Assumption A) gives the conclusion. ∎

To conclude this section, we show a stability result for T_j.

LEMMA B.10 (stability result for the mean-field map T_j). *Suppose that Assumption* A *is satisfied and that* $|h|_{C^{0,1}} \leqslant \ell_h < \infty$. *Then, for all* $R \geqslant 1$, *there is* $L_\mathsf{T} = L_\mathsf{T}(R, \kappa_y, \kappa_\Psi, \kappa_h, \ell_h, \Sigma, \Gamma)$ *such that for all* $\pi \in \mathcal{P}_R(\mathbf{R}^d \times \mathbf{R}^K)$ *and* $\mu \in \mathcal{P}(\mathbf{R}^d)$, *it holds that*

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

Proof. By Lemma B.2, the probability measure $\mathsf{QP}\mu$ belongs to $\mathcal{P}_{\widetilde{R}}(\mathbf{R}^d \times \mathbf{R}^K)$ for some $\widetilde{R} \geqslant 1$. Let us introduce

$$
r = \max\left\{R, \widetilde{R}, \kappa_y\right\}
$$

and denote by $\mathscr{T}^\pi$ and $\mathscr{T}^p$ the affine maps corresponding to evaluation of covariance information at π and $p = \mathsf{QP}\mu$. Specifically,

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

where $K = \mathcal{C}(\pi)$, $S = \mathcal{C}(p)$, and $y^\dagger = y^\dagger_{j+1}$. By the triangle inequality, we have

$$
d_g(\mathsf{T}_j \pi, \mathsf{T}_j p) \leqslant d_g(\mathscr{T}^\pi_\sharp \pi, \mathscr{T}^\pi_\sharp p) + d_g(\mathscr{T}^\pi_\sharp p, \mathscr{T}^p_\sharp p). \tag{B.22}
$$

Before separately bounding each term on the right-hand side, we obtain simple inequalities that will be helpful later in the proof. It holds that

$$
\|A_\pi\| \leqslant \|K_{uy}\| \|K_{yy}^{-1}\| \leqslant \|K\| \|K^{-1}\| \leqslant r^4 \tag{B.23}
$$

and similarly for A_p. Here we used that the matrix norm (induced by the Euclidean vector norm) of any submatrix is bounded from above by the norm of the original matrix. Using this bound again, we obtain, assuming that $r > 1$ without any loss of generality,

$$
\begin{aligned}
\|A_\pi - A_p\| = \|K_{uy}K_{yy}^{-1} - S_{uy}S_{yy}^{-1}\| &\leqslant \|(K_{uy} - S_{uy})K_{yy}^{-1}\| + \|S_{uy}\left(K_{yy}^{-1} - S_{yy}^{-1}\right)\| \\
&\leqslant r^2\Big(\|K_{uy} - S_{uy}\| + \|K_{yy}^{-1} - S_{yy}^{-1}\|\Big) \\
&= r^2\Big(\|K_{uy} - S_{uy}\| + \|K_{yy}^{-1}(S_{yy} - K_{yy})S_{yy}^{-1}\|\Big) \\
&\leqslant r^2\Big(\|K_{uy} - S_{uy}\| + \|K_{yy}^{-1}\|\|S_{yy} - K_{yy}\|\|S_{yy}^{-1}\|\Big) \\
&\leqslant 2r^6\|K - S\| \leqslant 2r^6(1+2r)\, d_g(\pi, p),
\end{aligned} \tag{B.24}
$$

where we used in the last line the inequality (B.12) from Lemma B.4.

Bounding the first term in (B.22). Let $f\colon \mathbf{R}^d \to \mathbf{R}$ denote a function satisfying $|f| \leqslant g$. It follows from the definition of the pushforward measure that

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

For all $(u, y) \in \mathbf{R}^d \times \mathbf{R}^K$, it holds that

$$
\begin{aligned}
\left|f \circ \mathscr{T}^\pi(u, y)\right| = \left|f\Big(u + A_\pi\big(y^\dagger - y\big)\Big)\right| &\leqslant g\Big(u + A_\pi\big(y^\dagger - y\big)\Big) \\
&= 1 + \left|u + A_\pi(y^\dagger - y)\right|^2 \leqslant 1 + 3|u|^2 + 3|A_\pi y^\dagger|^2 + 3|A_\pi y|^2 \\
&\leqslant 3\left(1 + |A_\pi y^\dagger|^2\right)\max\left\{1, \|A_\pi\|^2\right\} g(u, y).
\end{aligned}
$$

Therefore, using (B.23), we deduce that

$$
d_g(\mathscr{T}^\pi_\sharp \pi, \mathscr{T}^\pi_\sharp p) \leqslant 3\left(1 + r^{10}\right) r^8\, d_g(\pi, p). \tag{B.25}
$$

Bounding the second term in (B.22). Let $f\colon \mathbf{R}^d \to \mathbf{R}$ again denote a function satisfying the inequality $|f| \leqslant g$, with $g(u) = 1 + |u|^2$. It holds 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 right-hand side may be rewritten more explicitly as

$$
\begin{aligned}
&\left|p[f \circ \mathscr{T}^\pi - f \circ \mathscr{T}^p]\right| \\
&\quad = \left|\int_{\mathbf{R}^K}\int_{\mathbf{R}^d} f\left(u + A_\pi\big(y^\dagger - y\big)\right) - f\left(u + A_p\big(y^\dagger - y\big)\right)\, p(u, y)\,\mathrm{d}u\,\mathrm{d}y\right|.
\end{aligned}
$$

We apply a change of variable in order rewrite the integral in this identity as

$$
\begin{aligned}
&\int_{\mathbf{R}^K}\int_{\mathbf{R}^d}\big(f(u + A_\pi z) - f(u + A_p z)\big)\, p(u, y^\dagger - z)\,\mathrm{d}u\,\mathrm{d}z \\
&\qquad = \int_{\mathbf{R}^K}\int_{\mathbf{R}^d} f(v)\Big(p(v - A_\pi z, y^\dagger - z) - p(v - A_p z, y^\dagger - z)\Big)\mathrm{d}v\,\mathrm{d}z.
\end{aligned} \tag{B.26}
$$

It follows from the technical Lemma A.6 proved in Appendix A that

$$
\begin{aligned}
&\left|p(v - A_\pi z, y^\dagger - z) - p(v - A_p z, y^\dagger - z)\right| \\
&\quad \leqslant C|A_\pi z - A_p z| \exp\left(-\frac{1}{4}\left(\left|y^\dagger - z\right|_\Gamma^2 + \min\left\{|v - A_\pi z|_K^2, |v - A_p z|_K^2\right\}\right)\right).
\end{aligned}
$$

Using this inequality, we first bound the inner integral in (B.26) for fixed $z \in \mathbf{R}^K$. Keeping only the terms that depend on v, we obtain that

$$\begin{aligned}
&\int_{\mathbf{R}^d} |f(v)| \exp\left(-\frac{1}{4}\min\left\{|v - A_\pi z|_K^2, |v - A_p z|_K^2\right\}\right) \mathrm{d}v \\
&\quad = \int_{\mathbf{R}^d} |f(v)| \max\left\{\exp\left(-\frac{1}{4}|v - A_\pi z|_K^2\right), \exp\left(-\frac{1}{4}|v - A_p z|_K^2\right)\right\} \mathrm{d}v \\
&\quad \leqslant \int_{\mathbf{R}^d} |f(v)| \exp\left(-\frac{1}{4}|v - A_\pi z|_K^2\right) \mathrm{d}v + \int_{\mathbf{R}^d} |f(v)| \exp\left(-\frac{1}{4}|v - A_p z|_K^2\right) \mathrm{d}v.
\end{aligned}$$

Since $|f(v)| \leqslant 1 + |v|^2$, it is clear that this expression 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 we used (B.23) in the last inequality. The remaining integral in the z direction can be bounded similarly,

$$\begin{aligned}
d_g(\mathscr{T}_\sharp^\pi p, \mathscr{T}_\sharp^p p) &\leqslant Cr^8 \int_{\mathbf{R}^K} \left(1 + |z|^2\right) |A_\pi z - A_p z| \exp\left(-\frac{1}{4}\left(\left|y^\dagger - z\right|_\Gamma^2\right)\right) \mathrm{d}z \\
&\leqslant Cr^{11} \|A_\pi - A_p\| \leqslant Cr^{18} d_g(\pi, p), \qquad \text{(B.27)}
\end{aligned}$$

where we used the bound (B.24) in the last inequality. Combining (B.25) and (B.27), we deduce the statement. ∎

B.4. Additional stability results for the Gaussian projected filter. Next, we show the local Lipschitz continuity of the Gaussian projection map G. To this end, we begin by proving the following result. (The proof uses an approach similar to that employed in [17] for obtaining an upper bound on the distance between Gaussians in the usual total variation metric.)

LEMMA B.11. *Let $\mu_1 = \mathcal{N}(m_1, \Sigma_1)$ and $\mu_2 = \mathcal{N}(m_2, \Sigma_2)$ with $\Sigma_1, \Sigma_2 \in \mathbf{R}^{n\times n}$ symmetric and positive definite. It holds that*

$$d_g\big(\mu_1, \mu_2\big) \leqslant \sqrt{\mu_1\left[g^2\right] + \mu_2\left[g^2\right]}\left(3\left\|\Sigma_2^{-1}\Sigma_1 - I_n\right\|_{\mathrm{F}} + |m_1 - m_2|_{\Sigma_2}\right), \qquad \text{(B.28)}$$

where $\|\bullet\|_{\mathrm{F}}$ is the Frobenius matrix norm.

Proof. We denote by $(\lambda_i)_{1\leqslant i\leqslant n}$ the eigenvalues of $\Sigma_1^{-1}\Sigma_2$. Being the product of two symmetric positive definite matrices, $\Sigma_1^{-1}\Sigma_2$ is real diagonalizable with real and positive eigenvalues. We note that

$$d_g(\mu_1, \mu_2) \leqslant \mu_1[g] + \mu_2[g] \leqslant \sqrt{\mu_1[g^2]} + \sqrt{\mu_2[g^2]} \leqslant \sqrt{2}\sqrt{\mu_1[g^2] + \mu_2[g^2]}.$$

Now assume that $\max\{\lambda_j : 1 \leqslant j \leqslant n\} \geqslant 2$. In this case, there is integer k such that $\lambda_k \geqslant 2$, and so

$$3\|\Sigma_2^{-1}\Sigma_1 - I_n\|_{\mathrm{F}} = 3\sqrt{\sum_{i=1}^n |\lambda_i^{-1} - 1|^2} \geqslant 3\left|\lambda_k^{-1} - 1\right| \geqslant \frac{3}{2} \geqslant \sqrt{2}.$$

Combining the previous two inequalities proves (B.28) in the case where

$$\max\{\lambda_j : 1 \leqslant j \leqslant n\} \geqslant 2.$$

We assume from now on that $0 < \lambda_i < 2$ for all $i \in [\![1,n]\!]$. Employing the same reasoning as in [2, Lemma 3.1], we can prove

$$d_g(\mu_1,\mu_2)^2 \leqslant 2\Big(\mu_1\big[g^2\big] + \mu_2\big[g^2\big]\Big)\,\mathrm{KL}(\mu_2\|\mu_1), \tag{B.29}$$

where $\mathrm{KL}(\mu_2\|\mu_1)$ is the Kullback–Leibler (KL) divergence of μ_2 from μ_1, given by

$$\mathrm{KL}(\mu_1\|\mu_2) := \int_{\mathbf{R}^d} \frac{\mathrm{d}\mu_1}{\mathrm{d}\mu_2}\log\left(\frac{\mathrm{d}\mu_1}{\mathrm{d}\mu_2}\right)\mathrm{d}\mu_2. \tag{B.30}$$

The proof of the inequality (B.29) is presented in Lemma A.2 in Appendix A for completeness. The KL divergence between two Gaussians has a closed expression, which we rewrite in terms of the function $f_1(x) := x^{-1} - 1 + \log x$:

$$\begin{aligned}\mathrm{KL}(\mu_2\|\mu_1) &= \frac{1}{2}\left(\mathrm{tr}(\Sigma_2^{-1}\Sigma_1 - I_n) + (m_1-m_2)^\mathsf{T}\Sigma_2^{-1}(m_1-m_2) - \log\det(\Sigma_2^{-1}\Sigma_1)\right)\\ &= \frac{1}{2}\left(\sum_{i=1}^n(\lambda_i^{-1}-1) + (m_1-m_2)^\mathsf{T}\Sigma_2^{-1}(m_1-m_2) + \sum_{i=1}^n \log\lambda_i\right)\\ &= \frac{1}{2}\left(\sum_{i=1}^n f_1(\lambda_i) + |m_1-m_2|_{\Sigma_2}^2\right).\end{aligned}$$

The function f_1 is pointwise bounded from above by $f_2(x) := (1-x^{-1})^2$ for $x\in(0,2)$. To see this, consider the function $\varphi(x) = f_2(x) - f_1(x)$. It suffices to note that $\varphi(1) = 0$ and

$$\varphi'(x) = -\frac{1}{x}\left(1-\frac{1}{x}\right)\left(1-\frac{2}{x}\right).$$

Since $\varphi'(x) < 0$ for $x\in(0,1)$ and $\varphi'(x) > 0$ for $x\in(1,2)$, the desired upper bound on $f_1(x)$ by $f_2(x)$ for $x\in(0,2)$ follows. Therefore, since

$$\sum_{i=1}^n f_1(\lambda_i) \leqslant \sum_{i=1}^n f_2(\lambda_i) = \|\Sigma_2^{-1}\Sigma_1 - I_n\|_\mathrm{F}^2,$$

we have

$$\begin{aligned}\mathrm{KL}(\mu_2\|\mu_1) &\leqslant \frac{1}{2}\Big(\|\Sigma_2^{-1}\Sigma_1 - I_n\|_\mathrm{F}^2 + |m_1-m_2|_{\Sigma_2}^2\Big)\\ &\leqslant \frac{1}{2}\Big(\|\Sigma_2^{-1}\Sigma_1 - I_n\|_\mathrm{F} + |m_1-m_2|_{\Sigma_2}\Big)^2.\end{aligned}$$

Consequently, we deduce from (B.29) that

$$d_g(\mu_1,\mu_2) \leqslant \sqrt{\mu_1\big[g^2\big] + \mu_2\big[g^2\big]}\,\Big(\|\Sigma_2^{-1}\Sigma_1 - I_n\|_\mathrm{F} + |m_1-m_2|_{\Sigma_2}\Big),$$

which concludes the proof. ◻

Lemma B.12. *For all $R \geqslant 1$, there is $L_\mathsf{G} = L_\mathsf{G}(R,n)$ such that for all $\mu_1,\mu_2 \in \mathcal{P}_R(\mathbf{R}^n)$, the following inequality holds:*

$$d_g(\mathsf{G}\mu_1,\mathsf{G}\mu_2) \leqslant L_\mathsf{G}(R,n)\,d_g(\mu_1,\mu_2). \tag{B.31}$$

Before proving this lemma, we make two remarks on the form of the result.

Remark B.13. Lemmas B.7 and B.8 on the Lipschitz continuity of P and Q do not depend on the specific choice of the function g in Definition 1.1; they hold true also for the usual total variation distance. In contrast, in the proof of Lemma B.12, the specific choice of g is used in order to control differences between moments by means of $d_g(\bullet,\bullet)$.

Remark B.14. A simple example shows that the constant $L_{\mathsf{G}}(R,n)$ in (B.31) is divergent in the limit $R\to\infty$, indicating that the Gaussian projection operator G is not globally Lipschitz continuous. Specifically, take $\varepsilon\leqslant\frac{1}{2}$, and consider the probability distributions

$$\mu_1=\mathcal{N}(0,\varepsilon),\qquad \mu_2=\varepsilon\delta_{-1}+(1-2\varepsilon)\mu_1+\varepsilon\delta_1.$$

By definition of d_g, it follows that

$$\begin{aligned} d_g(\mu_1,\mu_2)&\leqslant\varepsilon g(-1)+2\varepsilon\mu_1[g]+\varepsilon g(1)\\ &=2\varepsilon+2\varepsilon\,(1+\varepsilon)+2\varepsilon\leqslant 8\varepsilon. \end{aligned}\tag{B.32}$$

On the other hand, $\mathsf{G}\mu_2=\mathcal{N}\big(0,2\varepsilon+(1-2\varepsilon)\varepsilon\big)$, and the variance of this Gaussian is bounded from below by $3\varepsilon-2\varepsilon^2\geqslant 2\varepsilon$. Consequently, it holds that

$$d_g(\mathsf{G}\mu_1,\mathsf{G}\mu_2)\geqslant d_1(\mathsf{G}\mu_1,\mathsf{G}\mu_2)\geqslant d_1\big(\mathcal{N}(0,\varepsilon),\mathcal{N}(0,2\varepsilon)\big)=d_1\big(\mathcal{N}(0,1),\mathcal{N}(0,2)\big),$$

where d_1 is the usual total variation metric. The right-hand side of this inequality does not depend on ε, and so

$$\frac{d_g(\mathsf{G}\mu_1,\mathsf{G}\mu_2)}{d_g(\mu_1,\mu_2)}\xrightarrow[\varepsilon\to 0]{}\infty.$$

This proves that G is not globally Lipschitz.

Proof of Lemma B.12. Let $m_i=\mathcal{M}(\mu_i)$ and $\Sigma_i=\mathcal{C}(\mu_i)$ for $i=1,2$. By Lemma B.4, it holds that

$$|m_1-m_2|_{\Sigma_2}\leqslant\frac{1}{2}\sqrt{\|\Sigma_2^{-1}\|}\,d_g(\mu_1,\mu_2).$$

In addition, also using Lemma B.4 and the facts that $\|A\|_{\mathrm{F}}\leqslant\sqrt{n}\,\|A\|$ and $\|AB\|\leqslant\|A\|\|B\|$ for any matrices $A,B\in\mathbf{R}^{n\times n}$,

$$\|\Sigma_2^{-1}\Sigma_1-I_n\|_{\mathrm{F}}\leqslant\sqrt{n}\|\Sigma_2^{-1}\|\left(1+\frac{1}{2}|m_1+m_2|\right)d_g(\mu_1,\mu_2).$$

Lemma B.11 then gives

$$\begin{aligned} d_g(\mathsf{G}\mu_1,\mathsf{G}\mu_2)&\leqslant\sqrt{\mathsf{G}\mu_1\,[g^2]+\mathsf{G}\mu_2\,[g^2]}\left(3\sqrt{n}\left(1+\frac{1}{2}|m_1+m_2|\right)\right.\\ &\qquad\left.\|\Sigma_2^{-1}\|+\frac{1}{2}\sqrt{\|\Sigma_2^{-1}\|}\right)d_g(\mu_1,\mu_2)\\ &\leqslant\sqrt{\mathsf{G}\mu_1[g^2]+\mathsf{G}\mu_2[g^2]}\left(3\sqrt{n}\left(1+\frac{1}{2}|m_1+m_2|\right)+1\right)\\ &\qquad\big(1+\|\Sigma_2^{-1}\|\big)\,d_g(\mu_1,\mu_2). \end{aligned}$$

Clearly, $\mathsf{G}\mu_i\left[g^2\right] \leqslant C\big(1+|m_i|^4+\|\Sigma_i\|^2\big)$ for our choice of g for an appropriate constant C, and so we conclude by use of Young's inequality that

$$d_g(\mathsf{G}\mu_1, \mathsf{G}\mu_2) \leqslant C\left(1 + |m_1|^3 + |m_2|^3 + \|\Sigma_1\|^{3/2} + \|\Sigma_2\|^{3/2}\right)\left(1 + \left\|\Sigma_2^{-1}\right\|\right) d_g(\mu_1, \mu_2)$$

for another constant C depending only on n; this completes the proof. ◻

We end the subsection by showing local Lipschitz continuity of the conditioning operator B_j over the set $\mathcal{G}(\mathbf{R}^d \times \mathbf{R}^K)$.

LEMMA B.15. *For all $R \geqslant 1$, there exists a constant $L_{\mathsf{B}} = L_{\mathsf{B}}(R, \kappa_y)$ such that for all Gaussian probability measures $\pi, p \in \mathcal{G}_R(\mathbf{R}^d \times \mathbf{R}^K)$,*

$$\forall j \in [\![0, J-1]\!], \qquad d_g\big(\mathsf{B}_j\pi, \mathsf{B}_j p\big) \leqslant L_{\mathsf{B}}(R, \kappa_y)\, d_g\big(\pi, p\big). \tag{B.33}$$

Proof. Let $\pi = \mathcal{N}(\tau, \Upsilon)$ and $p = \mathcal{N}(t, U)$. We use the shorthand notation $y^\dagger = y^\dagger_{j+1}$. It is well known that if $Z \sim \mathcal{N}(\tau, \Upsilon)$ for a vector $\tau \in \mathbf{R}^{d+K}$ and a matrix $\Upsilon \in \mathbf{R}^{(d+K)\times(d+K)}$, then the conditional distribution of Z_u given that $Z_y = y^\dagger$ is given by

$$Z_u | Z_y = y^\dagger \sim \mathcal{N}\big(\tau_u + \Upsilon_{uy}\Upsilon_{yy}^{-1}(y^\dagger - \tau_y), \Upsilon_{uu} - \Upsilon_{uy}\Upsilon_{yy}^{-1}\Upsilon_{yu}\big). \tag{B.34}$$

Denoting $\mathsf{B}_j\pi = \mathcal{N}(\lambda, \Delta)$ and $\mathsf{B}_j p = \mathcal{N}(\ell, D)$, we have

$$\begin{aligned}\lambda - \ell &= \tau_u - t_u + \Upsilon_{uy}\Upsilon_{yy}^{-1}(y^\dagger - \tau_y) - U_{uy}U_{yy}^{-1}(y^\dagger - t_y)\\ &= \tau_u - t_u + (\Upsilon_{uy} - U_{uy})\Upsilon_{yy}^{-1}(y^\dagger - \tau_y)\\ &\quad + U_{uy}U_{yy}^{-1}(U_{yy} - \Upsilon_{yy})\Upsilon_{yy}^{-1}(y^\dagger - \tau_y) - U_{uy}U_{yy}^{-1}(\tau_y - t_y).\end{aligned}$$

Since the 2-norm of any submatrix is bounded from above by the 2-norm of the matrix that contains it, we deduce that

$$\big|\lambda - \ell\big| \leqslant 2R^4\,|\tau - t| + 2R^6(R + \kappa_y)\|\Upsilon - U\|.$$

Similarly, $\|\Delta - D\| = \|\Upsilon_{uu} - U_{uu} - \Upsilon_{uy}\Upsilon_{yy}^{-1}\Upsilon_{yu} + U_{uy}U_{yy}^{-1}U_{yu}\| \leqslant 4R^8\|\Upsilon - U\|$. By Schur decomposition, it holds that

(B.35)
$$\begin{pmatrix} U_{uu} & U_{uy} \\ U_{uy}^\mathsf{T} & U_{yy}\end{pmatrix} = \begin{pmatrix} I_d & U_{uy}U_{yy}^{-1} \\ 0 & I_K\end{pmatrix}\begin{pmatrix} U_{uu} - U_{uy}U_{yy}^{-1}U_{uy}^\mathsf{T} & 0 \\ 0 & U_{yy}\end{pmatrix}\begin{pmatrix} I_d & 0 \\ U_{yy}^{-1}U_{uy}^\mathsf{T} & I_K\end{pmatrix}.$$

Since $D = U_{uu} - U_{uy}U_{yy}^{-1}U_{uy}^\mathsf{T}$, we have by the Courant–Fischer theorem

$$\begin{aligned}\lambda_{\min}(U) &= \min\left\{\frac{x^\mathsf{T}Ux}{x^\mathsf{T}x} : x \in \mathbf{R}^{d+K}\setminus\{0\}\right\} \leqslant \min\left\{\frac{x^\mathsf{T}Ux}{x^\mathsf{T}x} : x = \begin{pmatrix} u \\ -U_{yy}^{-1}U_{uy}^\mathsf{T}u\end{pmatrix}\right.\\ &\qquad \left.\text{and } u \in \mathbf{R}^d\setminus\{0\}\right\}\\ &= \min\left\{\frac{u^\mathsf{T}Du}{x^\mathsf{T}x} : x = \begin{pmatrix} u \\ -U_{yy}^{-1}U_{uy}^\mathsf{T}u\end{pmatrix} \text{ and } u \in \mathbf{R}^d\setminus\{0\}\right\}\\ &\leqslant \min\left\{\frac{u^\mathsf{T}Du}{u^\mathsf{T}u} : u \in \mathbf{R}^d\setminus\{0\}\right\} = \lambda_{\min}(D).\end{aligned}$$

Therefore, $\|D^{-1}\| \leqslant \|U^{-1}\| \leqslant R^2$. Using Lemma B.11, we then have

$$\begin{aligned} d_g(\mathsf{B}_j\pi, \mathsf{B}_j p) &\leqslant \sqrt{\mathsf{B}_j\pi[g^2] + \mathsf{B}_j p[g^2]}\,\big(3\|D^{-1}\Delta - I_d\|_{\mathrm{F}} + |\lambda - \ell|_D\big) \\ &\leqslant \sqrt{\mathsf{B}_j\pi[g^2] + \mathsf{B}_j p[g^2]}\,\Big(3\sqrt{d}\|\Delta - D\| + |\lambda - \ell|\Big)\,\big(1 + \|D^{-1}\|\big) \\ &\leqslant C(R, \kappa_y)\big(\|\Upsilon - U\| + |\tau - t|\big) \end{aligned}$$

for an appropriate constant C depending on R and κ_y. By Lemma B.4, it holds that

$$\|\Upsilon - U\| + \big|\tau - t\big| \leqslant \left(\frac{3}{2} + \frac{1}{2}|\tau + t|\right) d_g(\pi, p),$$

enabling us to conclude. ∎

Appendix C Technical results for Corollaries 2.5 and 3.3.

LEMMA C.1. *Suppose that* $\Psi\colon \mathbf{R}^d \to \mathbf{R}^d$ *and* $h\colon \mathbf{R}^d \to \mathbf{R}^K$ *are functions such that Assumption* A *is satisfied, and let* $(\Psi_n)_{n\in\mathbf{N}}$ *and* $(h_n)_{n\in\mathbf{N}}$ *be sequences of operators such that*

$$\Psi_n \xrightarrow[n\to\infty]{L^\infty(\mathbf{R}^d)} \Psi \qquad \text{and} \qquad h_n \xrightarrow[n\to\infty]{L^\infty(\mathbf{R}^d)} h.$$

Let P_n *and* Q_n *denote the maps* (1.5) *and* (1.7) *associated with the functions* Ψ_n *and* h_n, *and assume that they too satisfy Assumption* A. *Then*

$$\text{(C.1)} \qquad \sup_{\mu\in\mathcal{P}(\mathbf{R}^d)} d_g(\mathsf{P}\mu, \mathsf{P}_n\mu) \xrightarrow[n\to\infty]{} 0 \qquad \text{and} \qquad \sup_{\mu\in\mathcal{P}_R(\mathbf{R}^d)} d_g(\mathsf{Q}\mu, \mathsf{Q}_n\mu) \xrightarrow[n\to\infty]{} 0$$

for all $R \geqslant 1$ *for the second statement.*

Proof. For all $\mu \in \mathcal{P}(\mathbf{R}^d)$, the probability measures $\mathsf{P}\mu$ and $\mathsf{P}_n\mu$ have Lebesgue densities. By the definition in (1.5) of P and Remark 1.2, it holds that

$$d_g(\mathsf{P}\mu, \mathsf{P}_n\mu) = \frac{1}{\sqrt{(2\pi)^d \det\Sigma}} \int_{\mathbf{R}^d} g(u) \left| \int_{\mathbf{R}^d} \Delta_n(u, v)\, \mu(\mathrm{d}v) \right| \mathrm{d}u,$$

where

$$\Delta_n(u, v) := \exp\left(-\frac{1}{2}\,|u - \Psi(v)|_\Sigma^2\right) - \exp\left(-\frac{1}{2}\,|u - \Psi_n(v)|_\Sigma^2\right).$$

Letting $\psi_s(v) = (1-s)\Psi(v) + s\Psi_n(v)$ and using the same reasoning as in Lemma A.3, we obtain that

$$\begin{aligned} \big|\Delta_n(u, v)\big| &\leqslant \big|\Psi(v) - \Psi_n(v)\big|_\Sigma \int_0^1 \big|u - \psi_s\big|_\Sigma \exp\left(-\frac{1}{2}\big|u - \psi_s\big|_\Sigma^2\right) \mathrm{d}s \\ &\leqslant \big|\Psi(v) - \Psi_n(v)\big|_\Sigma \int_0^1 \Big(|u|_\Sigma + |\psi_s|_\Sigma\Big) \exp\left(-\frac{1}{3}\big|u\big|_\Sigma^2 + \big|\psi_s\big|_\Sigma^2\right) \mathrm{d}s && \text{(C.2a)} \\ &\leqslant C\big|\Psi(v) - \Psi_n(v)\big|_\Sigma \exp\left(-\frac{1}{4}\big|u\big|_\Sigma^2\right). && \text{(C.2b)} \end{aligned}$$

In (C.2a), we used that by Young's inequality, it holds for all $\delta > 0$ that

$$|u - \psi_s|_\Sigma^2 \geqslant \frac{1}{1+\delta}\,|u|_\Sigma^2 - \frac{1}{\delta}\,|\psi_s|_\Sigma^2.$$

Then, in (C.2b), we used the bound $\|\psi_s\|_{L^\infty} \leqslant \kappa_\Psi$, which holds for all $s \in [0,1]$ by Assumption (H2). The first limit in (C.1) then follows immediately.

For the second limit in (C.1), fix $\mu \in \mathcal{P}_R(\mathbf{R}^d)$, and note that by definition (1.7) of Q,

$$d_g(\mathsf{Q}\mu, \mathsf{Q}_n\mu) \leqslant \frac{1}{\sqrt{(2\pi)^K \det \Gamma}} \int_{\mathbf{R}^d \times \mathbf{R}^K} \left(1 + |u|^2 + |y|^2\right) \left|\widetilde{\Delta}_n(u,y)\right| \mu(\mathrm{d}u)\,\mathrm{d}y$$

$$\leqslant \frac{1}{\sqrt{(2\pi)^K \det \Gamma}} \int_{\mathbf{R}^d} \left(1 + |u|^2\right) \left(\int_{\mathbf{R}^K} \left(1 + |y|^2\right) \left|\widetilde{\Delta}_n(u,y)\right| \mathrm{d}y \right) \mu(\mathrm{d}u), \tag{C.3}$$

where we introduced

$$\widetilde{\Delta}_n(u,y) := \exp\left(-\frac{1}{2}\big|y - h(u)\big|_\Gamma^2\right) - \exp\left(-\frac{1}{2}\big|y - h_n(u)\big|_\Gamma^2\right).$$

Using the same strategy as above, we obtain that

$$\forall (u,y) \in \mathbf{R}^d \times \mathbf{R}^K, \qquad \left|\widetilde{\Delta}_n(u,y)\right| \leqslant C\big|h(u) - h_n(u)\big|_\Gamma \exp\left(-\frac{1}{4}|y|_\Gamma^2\right).$$

Substituting in (C.3) gives the second limit in (C.1). ∎

PROPOSITION C.2. *Suppose that $\Psi\colon \mathbf{R}^d \to \mathbf{R}^d$ and $h\colon \mathbf{R}^d \to \mathbf{R}^K$ are functions taking constant values, and let $(\Psi_n)_{n\in\mathbf{N}}$ and $(h_n)_{n\in\mathbf{N}}$ be sequences of functions such that*

$$\Psi_n \xrightarrow[n\to\infty]{L^\infty(\mathbf{R}^d)} \Psi \qquad \text{and} \qquad h_n \xrightarrow[n\to\infty]{L^\infty(\mathbf{R}^d)} h.$$

Let P_n and Q_n denote the maps (1.5) *and* (1.7) *(associated with the functions Ψ_n and h_n), and assume that they too satisfy Assumption* A. *Denote by $(\mu_j^n)_{j\in[\![1,J]\!]}$ the true filtering distribution associated with functions Ψ_n, h_n. Then, with the same notation as in Lemma* C.1, *it holds that*

$$\lim_{n\to\infty} \max_{j\in[\![0,J-1]\!]} d_g(\mathsf{G}\mathsf{Q}_n\mathsf{P}_n\mu_j^n, \mathsf{Q}_n\mathsf{P}_n\mu_j^n) = 0.$$

Proof. The conclusion follows from the stronger statement that

$$\lim_{n\to\infty} \sup_{\mu\in\mathcal{P}(\mathbf{R}^d)} d_g(\mathsf{G}\mathsf{Q}_n\mathsf{P}_n\mu, \mathsf{Q}_n\mathsf{P}_n\mu) = 0.$$

Indeed, fix $\mu \in \mathcal{P}(\mathbf{R}^d)$. By the triangle inequality, we have that

$$d_g(\mathsf{G}\mathsf{Q}_n\mathsf{P}_n\mu, \mathsf{Q}_n\mathsf{P}_n\mu) \leqslant d_g(\mathsf{G}\mathsf{Q}_n\mathsf{P}_n\mu, \mathsf{G}\mathsf{Q}\mathsf{P}\mu) + d_g(\mathsf{G}\mathsf{Q}\mathsf{P}\mu, \mathsf{Q}\mathsf{P}\mu) + d_g(\mathsf{Q}\mathsf{P}\mu, \mathsf{Q}_n\mathsf{P}_n\mu). \tag{C.4}$$

Since Ψ and h are constant, the probability measure $\mathsf{Q}\mathsf{P}\mu$ is a Gaussian distribution independent of μ, and so the second term on the right-hand side is zero. For the third term, we have using Lemma B.8 that

$$\begin{aligned} d_g(\mathsf{Q}\mathsf{P}\mu, \mathsf{Q}_n\mathsf{P}_n\mu) &\leqslant d_g(\mathsf{Q}\mathsf{P}\mu, \mathsf{Q}_n\mathsf{P}\mu) + d_g(\mathsf{Q}_n\mathsf{P}\mu, \mathsf{Q}_n\mathsf{P}_n\mu) \\ &\leqslant d_g(\mathsf{Q}\nu, \mathsf{Q}_n\nu) + \left(1 + \kappa_{h_n}^2 + \operatorname{tr}(\Gamma)\right) d_g(\mathsf{P}\mu, \mathsf{P}_n\mu), \qquad \nu := \mathsf{P}\mu. \end{aligned}$$

Noting that $\nu := \mathsf{P}\mu$ is a fixed Gaussian measure independent of μ and using Lemma C.1, we deduce that both terms on the right-hand side tend to 0 in the limit $n \to \infty$

uniformly in $\mu \in \mathcal{P}(\mathbf{R}^d)$. It remains to show that the first term in (C.4) also tends to 0 as $n \to \infty$ uniformly in $\mu \in \mathcal{P}(\mathbf{R}^d)$. In view of the moment bounds in Lemma B.2, there exist $R \geqslant 1$ and $N \in \mathbf{N}$ such that

$$\forall n \geqslant N, \qquad \forall \mu \in \mathcal{P}(\mathbf{R}^d), \qquad \mathsf{Q}_n \mathsf{P}_n \mu \in \mathcal{P}_R(\mathbf{R}^d) \qquad \text{and} \qquad \mathsf{QP}\mu \in \mathcal{P}_R(\mathbf{R}^d).$$

Therefore, by the local Lipschitz continuity of G established in Lemma B.12 and the uniform-in-μ convergence to 0 of the third term in (C.4) that we already proved, it holds that

$$\sup_{\mu \in \mathcal{P}(\mathbf{R}^d)} d_g(\mathsf{GQ}_n \mathsf{P}_n \mu, \mathsf{GQP}\mu) \xrightarrow[n \to \infty]{} 0,$$

which concludes the proof. ∎

Acknowledgments. We are grateful to the reviewers for very useful comments on a previous version of this paper. In particular, we thank the reviewers for pointing out the elementary inequality in Lemma A.3 and for showing us how this inequality could be used to significantly improve the proofs of Lemmas A.4, A.6, and C.1. The constructive proof of Lemma A.4, in particular, was proposed by a reviewer.

REFERENCES

[1] A. Apte, C. K. Jones, and A. M. Stuart, *A Bayesian approach to Lagrangian data assimilation*, Tellus A, 60 (2008), pp. 336–347.

[2] A. Arnold, J. A. Carrillo, and C. Manzini, *Refined long-time asymptotics for some polymeric fluid flow models*, Commun. Math. Sci., 8 (2010), pp. 763–782.

[3] M. Asch, M. Bocquet, and M. Nodet, *Data Assimilation*, Fundamentals of Algorithms 11, SIAM, Philadelphia, 2016, https://doi.org/10.1137/1.9781611974546.pt1.

[4] 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, Institute of Mathematical Statistics Collections, Vol. 3, B. Clarke and S. Ghosal, eds., Institute of Mathematical Statistics, Waite Hill, OH, 2008, pp. 318–329, https://doi.org/10.1214/074921708000000228.

[5] C. M. Bishop, *Pattern Recognition and Machine Learning*, Information Science and Statistics, Springer-Verlag, Berlin, 2006, https://doi.org/10.1007/978-0-387-45528-0.

[6] A. Biswas and M. Branicki, *A Unified Framework for the Analysis of Accuracy and Stability of a Class of Approximate Gaussian Filters for the Navier-Stokes Equations*, preprint, arXiv:2402.14078, 2024.

[7] F. Bolley, J. A. Cañizo, and J. A. Carrillo, *Stochastic mean-field limit: Non-Lipschitz forces and swarming*, Math. Models Methods Appl. Sci., 21 (2011), pp. 2179–2210, https://doi.org/10.1142/S0218202511005702.

[8] E. Caglioti, F. Golse, and M. Iacobelli, *Quantization of probability distributions and gradient flows in space dimension* 2, Ann. Inst. H. Poincaré Anal. Non Linéaire, 35 (2018), pp. 1531–1555, https://doi.org/10.1016/j.anihpc.2017.12.003.

[9] E. Calvello, S. Reich, and A. M. Stuart, *Ensemble Kalman Methods: A Mean Field Perspective*, preprint, arXiv:2209.11371, 2024, (to appear 2025, Acta Numerica).

[10] J. A. Carrillo and U. Vaes, *Wasserstein stability estimates for covariance-preconditioned Fokker-Planck equations*, Nonlinearity, 34 (2021), pp. 2275–2295.

[11] D. Crişan, P. Del Moral, and T. J. Lyons, *Interacting particle systems approximations of the Kushner-Stratonovitch equation*, Adv. Appl. Probab., 31 (1999), pp. 819–838, https://doi.org/10.1239/aap/1029955206.

[12] D. Crisan, A. López-Yela, and J. Miguez, *Stable approximation schemes for optimal filters*, SIAM/ASA J. Uncertain. Quantif., 8 (2020), pp. 483–509, https://doi.org/10.1137/19M1255410.

[13] 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.

[14] P. Del Moral, *Nonlinear filtering: Interacting particle resolution*, C. R. Acad. Sci. Paris Sér. I Math., 325 (1997), pp. 653–658, https://doi.org/10.1016/S0764-4442(97)84778-7.

[15] P. Del Moral and A. Guionnet, *On the stability of interacting processes with applications to filtering and genetic algorithms*, Ann. Inst. Henri Poincaré Probab. Stat., 37 (2001), pp. 155–194, https://doi.org/10.1016/S0246-0203(00)01064-5.

[16] 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.

[17] L. Devroye, A. Mehrabian, and T. Reddad, *The Total Variation Distance between High-Dimensional Gaussians*, preprint, arXiv:1810.08693, 2018.

[18] Z. Ding and Q. Li, *Ensemble Kalman inversion: Mean-field limit and convergence analysis*, Stat. Comput., 31 (2021), 9, https://doi.org/10.1007/s11222-020-09976-0.

[19] 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.

[20] Z. Ding, Q. Li, and J. Lu, *Ensemble Kalman inversion for nonlinear problems: Weights, consistency, and variance bounds*, Found. Data Sci., 3 (2021), pp. 371–411, https://doi.org/10.3934/fods.2020018.

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

[22] A. Durmus, A. Eberle, A. Guillin, and R. Zimmer, *An elementary approach to uniform in time propagation of chaos*, Proc. Amer. Math. Soc., 148 (2020), pp. 5387–5398, https://doi.org/10.1090/proc/14612.

[23] O. G. Ernst, B. Sprungk, and H. Starkloff, *Analysis of the ensemble and polynomial chaos Kalman filters in Bayesian inverse problems*, SIAM/ASA J. Uncertain. Quantif., 3 (2015), pp. 823–851, https://doi.org/10.1137/140981319.

[24] S. Ertel and W. Stannat, *Analysis of the ensemble Kalman–Bucy filter for correlated observation noise*, Ann. Appl. Probab., 34 (2024), pp. 1072–1107.

[25] 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.

[26] G. Evensen, *Data Assimilation: The Ensemble Kalman Filter*, 2nd ed., Springer-Verlag, Berlin, 2009, https://doi.org/10.1007/978-3-642-03711-5.

[27] G. Evensen, F. C. Vossepoel, and P. J. van Leeuwen, *Data Assimilation Fundamentals: A Unified Formulation of the State and Parameter Estimation Problem*, Springer Textbooks in Earth Sciences, Geography and Environment, Springer-Verlag, Berlin, 2022, https://doi.org/10.1007/978-3-030-96709-3.

[28] A. Garbuno-Inigo, F. Hoffmann, W. Li, and A. M. Stuart, *Interacting Langevin diffusions: Gradient structure and ensemble Kalman sampler*, SIAM J. Appl. Dyn. Syst., 19 (2020), pp. 412–441, https://doi.org/10.1137/19M1251655.

[29] A. Garbuno-Inigo, N. Nüsken, and S. Reich, *Affine invariant interacting Langevin dynamics for Bayesian inference*, SIAM J. Appl. Dyn. Syst., 19 (2020), pp. 1633–1658.

[30] M. Ghil, S. Cohn, J. Tavantzis, K. Bube, and E. Isaacson, *Application of estimation theory to numerical weather prediction*, in Seminar on Data Assimilation Methods, European Centre for Medium-Range Weather Forecasts, Shinfield Park, UK, 1980, pp. 139–224.

[31] G. A. Gottwald and A. J. Majda, *A mechanism for catastrophic filter divergence in data assimilation for sparse observation networks*, Nonlinear Process. Geophys., 20 (2013), pp. 705–712.

[32] S. Graf and H. Luschgy, *Foundations of Quantization for Probability Distributions*, Lecture Notes in Mathematics 1730, Springer-Verlag, Berlin, 2000.

[33] M. Hairer and J. C. Mattingly, *Yet another look at Harris' ergodic theorem for Markov chains*, in Seminar on Stochastic Analysis, Random Fields and Applications VI, Progress in Probability 63, Birkhäuser Basel, Basel, Switzerland, 2011, pp. 109–117, https://doi.org/10.1007/978-3-0348-0021-1.

[34] D. Z. Huang, J. Huang, S. Reich, and A. M. Stuart, *Efficient derivative-free Bayesian inference for large-scale inverse problems*, Inverse Problems, 38 (2022), 125006, https://doi.org/10.1088/1361-6420/ac99fa.

[35] D. Z. Huang, T. Schneider, and A. M. Stuart, *Iterated Kalman methodology for inverse problems*, J. Comput. Phys., 463 (2022), 111262,https://doi.org/10.1016/j.jcp.2022.111262.

[36] M. A. Iglesias, K. J. H. Law, and A. M. Stuart, *Ensemble Kalman methods for inverse problems*, Inverse Problems, 29 (2013), 045001, https://doi.org/10.1088/0266-5611/29/4/045001.

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

[38] S. J. Julier and J. K. Uhlmann, *New extension of the Kalman filter to nonlinear systems*, in Signal Processing, Sensor Fusion, and Target Recognition VI, Vol. 3068, SPIE, Bellingham, WA, 1997, pp. 182–193.

[39] R. E. Kalman, *A new approach to linear filtering and prediction problems*, Trans. ASME Ser. D. J. Basic Engrg., 82 (1960), pp. 35–45.

[40] D. Kelly, A. J. Majda, and X. T. Tong, *Concrete ensemble Kalman filters with rigorous catastrophic filter divergence*, Proc. Natl. Acad. Sci. USA, 112 (2015), pp. 10589–10594.

[41] 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.

[42] T. Lange and W. Stannat, *Mean field limit of ensemble square root filters—discrete and continuous time,* Found. Data Sci., 3 (2021), pp. 563–588.

[43] T. Lange and W. Stannat, *On the continuous time limit of the ensemble kalman filter*, Math. Comp., 90 (2021), pp. 233–265.

[44] K. Law, A. Stuart, and K. Zygalakis, *Data Assimilation: A Mathematical Introduction*, Texts in Applied Mathematics 62, Springer-Verlag, Berlin, 2015, https://doi.org/10.1007/978-3-319-20325-6.

[45] 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 University Press, Oxford, 2011, pp. 598–631.

[46] A. J. Majda and X. T. Tong, *Performance of ensemble Kalman filters in large dimensions*, Comm. Pure Appl. Math., 71 (2018), pp. 892–937, https://doi.org/10.1002/cpa.21722.

[47] 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.

[48] F. Medina-Aguayo, D. Rudolf, and N. Schweizer, *Perturbation bounds for Monte Carlo within Metropolis via restricted approximations*, Stochastic Process. Appl., 130 (2020), pp. 2200–2227, https://doi.org/10.1016/j.spa.2019.06.015.

[49] S. P. Meyn and R. L. Tweedie, *Stability of Markovian processes* I. *Criteria for discrete-time chains*, Adv. Appl. Probab., 24 (1992), pp. 542–574, https://doi.org/10.2307/1427479.

[50] R. Nickl and E. S. Titi, *On posterior consistency of data assimilation with Gaussian process priors: The 2D Navier-Stokes equations*, Ann. Statist., 52 (2024), pp. 1825–1844.

[51] D. Ocone and E. Pardoux, *Asymptotic stability of the optimal filter with respect to its initial condition*, SIAM J. Control Optim., 34 (1996), pp. 226–243, https://doi.org/10.1137/S0363012993256617.

[52] 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.

[53] S. Reich and C. Cotter, *Probabilistic Forecasting and Bayesian Data Assimilation*, Cambridge University Press, Cambridge, 2015, https://doi.org/10.1017/CBO9781107706804.

[54] D. Sanz-Alonso and A. M. Stuart, *Long-time asymptotics of the filtering distribution for partially observed chaotic dynamical systems*, SIAM/ASA J. Uncertain. Quantif., 3 (2015), pp. 1200–1220, https://doi.org/10.1137/140997336.

[55] D. Sanz-Alonso, A. M. Stuart, and A. Taeb, *Inverse problems and data assimilation*, London Math. Soc. Stud. Texts 107, Cambridge University Press, 2023.

[56] C. Schillings and A. M. Stuart, *Analysis of the ensemble Kalman filter for inverse problems*, SIAM J. Numer. Anal., 55 (2017), pp. 1264–1290, https://doi.org/10.1137/16M105959X.

[57] C. Schillings and A. M. Stuart, *Convergence analysis of ensemble Kalman inversion: The linear, noisy case*, Appl. Anal., 97 (2018), pp. 107–123, https://doi.org/10.1080/00036811.2017.1386784.

[58] C. Snyder, T. Bengtsson, P. Bickel, and J. Anderson, *Obstacles to high-dimensional particle filtering*, Monthly Weather Rev., 136 (2008), pp. 4629–4640.

[59] C. Snyder, T. Bengtsson, and M. Morzfeld, *Performance bounds for particle filters using the optimal proposal*, Monthly Weather Rev., 143 (2015), pp. 4750–4761.

[60] A. Stuart and A. R. Humphries, *Dynamical Systems and Numerical Analysis*, Vol. 2, Cambridge University Press, Cambridge, 1998.

[61] I. O. Tolstikhin, B. K. Sriperumbudur, and B. Schölkopf, *Minimax estimation of maximum mean discrepancy with radial kernels*, Adv. Neural Inf. Process. Syst., 29 (2016).

[62] X. T. Tong, *Performance analysis of local ensemble Kalman filter*, J. Nonlinear Sci., 28 (2018), pp. 1397–1442, https://doi.org/10.1007/s00332-018-9453-2.

[63] X. T. Tong, A. J. Majda, and D. Kelly, *Nonlinear stability of the ensemble Kalman filter with adaptive covariance inflation*, Commun. Math. Sci., 14 (2016), pp. 1283–1313, https://doi.org/10.4310/CMS.2016.v14.n5.a5.

[64] X. T. Tong and M. Morzfeld, *Localized ensemble Kalman inversion*, Inverse Problems, 39 (2023), 064002.

[65] X. T. Tong and R. van Handel, *Ergodicity and stability of the conditional distributions of nondegenerate Markov chains*, Ann. Appl. Probab., 22 (2012), pp. 1495–1540, https://doi.org/10.1214/11-AAP800.

[66] A. W. van der Vaart, *Asymptotic Statistics*, Cambridge Series in Statistical and Probabilistic Mathematics 3, Cambridge University Press, Cambridge, 1998, https://doi.org/10.1017/CBO9780511802256.

[67] R. van Handel, *Uniform observability of hidden Markov models and filter stability for unstable signals*, Ann. Appl. Probab., 19 (2009), pp. 1172–1199, https://doi.org/10.1214/08-AAP576.

[68] C. Villani, *Optimal Transport: Old and New*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences] 338, Springer-Verlag, Berlin, 2009, https://doi.org/10.1007/978-3-540-71050-9.