# Statistical Accuracy of Approximate Filtering Methods


**J. A. Carrillo, F. Hoffmann, A. M. Stuart, and U. Vaes**

**Abstract** Estimating the statistics of the state of a dynamical system, from partial and noisy observations, is both mathematically challenging and finds wide application. Furthermore, the applications are of great societal importance, including problems such as probabilistic weather forecasting (Kalnay 2003) and prediction of epidemics (Keeling and Eames 2005). Particle filters provide a well-founded approach to the problem, leading to provably accurate approximations of the statistics (Doucet et al. 2001). However these methods perform poorly in high dimensions (Bickel et al. 2008; Snyder et al. 2008). In 1994 the idea of ensemble Kalman filtering was introduced (Evensen 1994) leading to a methodology that has been widely adopted in the geophysical sciences (van Leeuwen et al. 2019) and also finds application to quite general inverse problems (Iglesias et al. 2013). However, ensemble Kalman filters have defied rigorous analysis of their statistical accuracy, except in the linear Gaussian setting (Le Gland et al. 2011; Mandel et al. 2011). In this article we describe recent work which takes first steps to analyze the statistical accuracy of ensemble Kalman filters beyond the linear Gaussian setting (Carrillo et al. 2022). The subject is inherently technical, as it involves the evolution of probability measures according to a nonlinear and nonautonomous dynamical system; and the approximation of this evolution. It can nonetheless be presented in a fairly accessible fashion, understandable with basic knowledge of dynamical systems, numerical analysis and probability. We undertake such a presentation here.

J. A. Carrillo
Mathematical Institute, University of Oxford, Oxford, UK
e-mail: carrillo@maths.ox.ac.uk

F. Hoffmann · A. M. Stuart (✉)
Department of Computing and Mathematical Sciences, Caltech, Pasadena, CA, USA
e-mail: astuart@caltech.edu

F. Hoffmann
e-mail: franca.hoffmann@caltech.edu

U. Vaes
MATHERIALS Project-Team, Inria Paris, Paris, France
e-mail: urbain.vaes@inria.fr

CERMICS,École des Ponts, Champs-sur-Marne, France



S. Oishi et al. (eds.), *Recent Developments in Industrial and Applied Mathematics*,
ICIAM2023 Springer Series 1, https://doi.org/10.1007/978-981-95-1446-5_14

# 1 Filtering Problem

Consider a dynamical system for state $\{v_n\}_{n\in\mathbb{Z}^+}$ evolving in $\mathbb{R}^d$, partially and noisily observed through data $\{y_n\}_{n\in\mathbb{N}}$ in $\mathbb{R}^K$, with state and data determined by the following system, holding for $n \in \mathbb{Z}^+$:

$$\begin{aligned} \textbf{State}: &\quad v_{n+1} = \Psi(v_n) + \xi_n\,, \\ \textbf{Data}: &\quad y_{n+1} = h(v_{n+1}) + \eta_{n+1}\,. \end{aligned}$$

Here $\mathbb{N} := \{1, 2, \ldots\}$ and $\mathbb{Z}^+ = \{0, 1, 2, \ldots\}$. We assume that the initial state of the system is a Gaussian random variable $v_0 \sim \mathcal{N}(m_0, C_0)$. Furthermore we assume that $\xi_n \sim \mathcal{N}(0, \Sigma)$ is the mean-zero noise affecting the state evolution and $\eta_{n+1} \sim \mathcal{N}(0, \Gamma)$ is the mean-zero noise entering the data acquisition process. We assume that the state evolution and data acquisition noise sequences are i.i.d. and that the following independence assumptions hold:

$$v_0 \perp\!\!\!\perp \{\xi_n\}_{n\in\mathbb{Z}^+} \perp\!\!\!\perp \{\eta_{n+1}\}_{n\in\mathbb{Z}^+}.$$

The objective of filtering is to determine the probability distribution on $v_n$ given all the data acquired up to that point. To this end we define, for a given realization of the data, denoted by a dagger $^\dagger$,

$$Y_n^\dagger = \{y_\ell^\dagger\}_{\ell=1}^n, \qquad v_n | Y_n^\dagger \sim \mu_n.$$

With this notation we can state the objective of filtering more precisely: it is to determine the *filtering distribution* or *true filter* $\mu_n$ and update it sequentially in $n$. As we now show $\mu_n$ evolves according to a nonautonomous and nonlinear dynamical system on the space of probability measures.

To determine this dynamical system, and various approximations of it that follow, it is helpful to define

- $\mathcal{P}(\mathbb{R}^r)$: all probability measures on $\mathbb{R}^r$.
- $\mathcal{G}(\mathbb{R}^r)$: all Gaussian probability measures on $\mathbb{R}^r$.

For simplicity we will use the same symbol for measures and their densities throughout this article. Before writing the dynamical system for $\mu_n$ we first determine the evolution of the measures $\pi_n, \mathfrak{r}_n$ defined by

$$v_n \sim \pi_n, \qquad (v_n, y_n) \sim \mathfrak{r}_n.$$

These evolve according to

$$\begin{aligned} \pi_{n+1} &= \mathsf{P}\pi_n, \\ \mathfrak{r}_{n+1} &= \mathsf{Q}\pi_{n+1} \end{aligned}$$

Here $\mathsf{P}\colon \mathcal{P}(\mathbb{R}^d) \to \mathcal{P}(\mathbb{R}^d)$ is the linear operator

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

whilst $\mathsf{Q}\colon \mathcal{P}(\mathbb{R}^d) \to \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^K)$, which is also a linear operator, is determined by

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

The evolution for $\pi_n$, which describes the probability of the state $v_n$, is determined by a Markov process on $\mathbb{R}^d$ defined via the linear operator $\mathsf{P}$. The linear operator $\mathsf{Q}$ lifts $\pi_n$ to the joint space of state and data $(v_n, y_n)$. It is worth highlighting that, by moving from the evolution of $(v_n, y_n) \in \mathbb{R}^d \times \mathbb{R}^K$ (finite dimensions) to the evolution of $(\pi_n, \mathfrak{r}_n) \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^K)$ (infinite dimensions) we have converted a nonlinear stochastic problem into a linear autonomous one.

However, the dynamical system for $\mu_n$ is nonlinear and nonautonomous. To determine the dynamical system for $\mu_n$ we need to introduce a nonlinear operator on probability measures. Specifically, let $\mathsf{B}(\bullet; y^\dagger)\colon \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^K) \to \mathcal{P}(\mathbb{R}^d)$ describe conditioning of a joint random variable $(v, y)$ on observation $y = y^\dagger$:

$$\mathsf{B}(\rho; y^\dagger)(u) = \frac{\rho(u, y^\dagger)}{\int_{\mathbb{R}^d} \rho(u, y^\dagger)\, \mathrm{d}u}.$$

Armed with this we define $v_{n+1}|Y_n^\dagger \sim \widehat{\mu}_{n+1}$ and observe that

$$\begin{aligned}
\widehat{\mu}_{n+1} &= \mathsf{P}\mu_n, && v_{n+1}|Y_n^\dagger \sim \widehat{\mu}_{n+1} \\
\rho_{n+1} &= \mathsf{Q}\widehat{\mu}_{n+1}, && (v_{n+1}, y_{n+1})|Y_n^\dagger \sim \rho_{n+1} \\
\mu_{n+1} &= \mathsf{B}(\rho_{n+1}; y_{n+1}^\dagger), && \text{conditioning on } y_{n+1} = y_{n+1}^\dagger.
\end{aligned}$$

Thus we have the nonlinear and nonautonomous dynamical system

$$\mu_{n+1} = \mathsf{B}(\mathsf{QP}\mu_n; y_{n+1}^\dagger), \qquad \mu_0 = \mathcal{N}(m_0, C_0).$$

This evolution may be thought of in terms of sequential application of Bayes theorem: $\mathsf{P}\mu_n$ is prior prediction; $\mathsf{L}(\bullet; y^\dagger) := \mathsf{B}(\bullet; y^\dagger) \circ \mathsf{Q}$ maps prior to posterior according to Bayes theorem. This leads to the following equivalent formulation of the evolution:

$$\mu_{n+1} = \mathsf{L}(\mathsf{P}\mu_n; y_{n+1}^\dagger), \qquad \mu_0 = \mathcal{N}(m_0, C_0).$$

The resulting evolution is an infinite dimensional problem since $\mu_n$ is a probability density over $\mathbb{R}^d$. Approximating this evolution is thus a significant challenge.

## 2 Particle Filter

During the last quarter of the last century, the particle filter became a widely adopted methodology for solving the filtering problem in moderate dimensions $d$; the methodology is overviewed in Doucet et al. (2001).

To describe the basic form of the method we introduce a random map on the space of probability measures; this map encapsulates Monte Carlo sampling. To this end let $\Omega$ denote an abstract probability space (encapsulating the sampling underlying the Monte Carlo method) and define $\mathsf{S}^J : \mathcal{P}(\mathbb{R}^r) \times \Omega \to \mathcal{P}(\mathbb{R}^r)$ to be the empirical approximation operator:

$$\mathsf{S}^J \mu = \frac{1}{J} \sum_{j=1}^{J} \delta_{v_j}, \qquad v_j \overset{\text{i.i.d.}}{\sim} \mu .$$

Here $\Omega$ is a probability space capturing the process through which independent draws (from $\mu$) are made. We suppress explicit dependence on $\omega \in \Omega$ in the expression $\mathsf{S}^J \mu$ for notational convenience. Then, for large $J$ and fixed $\omega \in \Omega$, $\mathsf{S}^J$ is a random approximation of the identity on $\mathcal{P}(\mathbb{R}^r)$, a statement made precise in the remarks that follow Theorem 1.

Since $\mathsf{S}^J \approx I$, we may introduce the approximation $\mu_n^{\mathrm{PF}} \approx \mu_n$ evolving according to the map

$$\mu_{n+1}^{\mathrm{PF}} = \mathsf{L}(\mathsf{S}^J \mathsf{P} \mu_n^{\mathrm{PF}}; y_{n+1}^{\dagger}), \qquad \mu_0^{\mathrm{PF}} = \mu_0.$$

In the preceding expression it is implicit that each application of $\mathsf{S}^J$ is based on an independent realization of $\omega \in \Omega$. It then follows that, for $n \in \mathbb{N}$,

$$\mu_n^{\mathrm{PF}} = \sum_{j=1}^{J} w_n^{(j)} \delta_{\widehat{v}_n^{(j)}}$$

where the particles $\widehat{v}_n^{(j)}$ and weights $w_n^{(j)}$ evolve according to

$$\begin{aligned}
\widehat{v}_{n+1}^{(j)} &= \Psi\big(v_n^{(j)}\big) + \xi_n^{(j)}, \qquad v_n^{(j)} \overset{\text{i.i.d.}}{\sim} \mu_n^{\mathrm{PF}}, \\
\ell_{n+1}^{(j)} &= \exp\left( -\frac{1}{2} \big| y_{n+1}^{\dagger} - h\big(\widehat{v}_{n+1}^{(j)}\big) \big|_{\Gamma}^2 \right), \\
w_{n+1}^{(j)} &= \ell_{n+1}^{(j)} \Big/ \Big( \sum_{m=1}^{J} \ell_{n+1}^{(m)} \Big).
\end{aligned}$$

Here $\xi_n^{(j)}$ are $\mathcal{N}(0, \Sigma)$ Gaussians that are i.i.d. with respect to both $n$ and $j$.

Systemization of the analysis of this method may be found in the work of Del Moral (1997), with the significant extension to analysis over long time-intervals in Del

Moral and Guionnet (2001) and to continuous time in Crişan et al. (1999). We describe a prototypical theoretical result, based on the formulation underpinning Rebeschini and van Handel (2015); for precise statement of conditions under which the theorem holds see Law et al. (2015, Theorem 4.5). The theorem deploys a metric $\mathrm{d}(\bullet, \bullet)$ on random probability measures. This is defined as follows. First let $\mu, \nu$ be elements of the space of random probability measures defined over a probability space $\Omega$ and let $\mu_\omega$ (resp. $\nu_\omega$) denote the specific probability measure that arises when random probability measure $\mu$ (resp. $\nu$) is evaluated at $\omega \in \Omega$. Let $\mathbb{E}^\omega$ denote expectation with respect to $\omega \in \Omega$. Then

$$\mathrm{d}(\mu, \nu)^2 = \sup_{|f| \leq 1} \mathbb{E}^\omega \left| \mu_\omega[f] - \nu_\omega[f] \right|^2,$$

where $\mu_\omega[f] = \int f(u)\, \mu_\omega(\mathrm{d}u)$ and similarly for $\nu_\omega$.

The metric reduces to TV distance if applied to deterministic probability measures. In the application that follows $\Omega$ is defined to include the randomness that arises from the successive empirical approximations using $\mathsf{S}^J$ over multiple steps of the particle filter.

**Theorem 1** (Law et al. 2015, Theorem 4.5) *It holds that*

$$\sup_{0 \leq n \leq N} \mathrm{d}(\mu_n, \mu_n^{\mathrm{PF}}) \leq \frac{C(N)}{\sqrt{J}}.$$

**Remark 1** The following remarks explain and interpret the theorem:

- The theorem states that a distance between the particle filter and the true filter is bounded above by an error which decreases at the Monte Carlo rate with respect to the number of particles $J$.
- The proof follows a typical numerical analysis structure: Consistency + Stability Implies Convergence.
- Consistency: $\mathrm{d}(\mathsf{S}^J \mu, \mu) \leq \frac{1}{\sqrt{J}}$. This shows that $\mathsf{S}^J$ approximates the identity.
- Stability: $\mathsf{P}, \mathsf{L}$ are Lipschitz in $\mathrm{d}(\bullet, \bullet)$.
- The theorem assumes upper and lower bounds on the $n$-dependent likelihood arising from sequential application of Bayes theorem; the constant $C$ depends on these bounds. Furthermore $C$ grows exponentially with $N$.

In practice, especially in high dimensions $d$, the method often suffers from weight collapse (Bickel et al. 2008; Snyder et al. 2008). This refers to the phenomenon where one of the weights $\{w_n^{(j)}\}_{j=1}^J$ is close to 1 and the others are therefore necessarily close to 0. When this happens the method is of little value because the effective number of particles approximating $\mu_n$ is 1. This leads us to consider the ensemble Kalman filter which, by design, has equal weights.

## 3 Ensemble Kalman Filter

The ensemble Kalman filter, introduced by Evensen (1994), is overviewed in the texts (Evensen 2009; Evensen et al. 2022). We will state the basic particle form of the algorithm, from Evensen (1994). We will then show how it may be derived from a mean-field perspective highlighted in Calvello et al. (2025). Finally we describe analysis of this mean-field ensemble Kalman filter, and its relationship to the true filter (Carrillo et al. 2022). This analysis of the mean field method is a first step to eventually obtaining an error estimate between the ensemble Kalman filter and the true filter in the finite-ensemble setting.

To describe the particle form of the method we first introduce key notation relating to covariances. Specifically we write covariance under $\rho \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^K)$ as:

$$\mathrm{cov}(\rho) = \begin{pmatrix} C^{vv}(\rho) & C^{vy}(\rho) \\ C^{vy}(\rho)^\top & C^{yy}(\rho) \end{pmatrix}.$$

We use similar notation ($\mathrm{mean}_\rho$) for the mean under $\rho$. The ensemble Kalman filter from Evensen (1994) then has the form, for $n \in \mathbb{Z}^+$,

$$\begin{aligned}
\widehat{v}_{n+1}^{(j)} &= \Psi\big(v_n^{(j)}\big) + \xi_n^{(j)}, \\
\widehat{y}_{n+1}^{(j)} &= h(\widehat{v}_{n+1}^{(j)}) + \eta_{n+1}^{(j)}, \\
v_{n+1}^{(j)} &= \widehat{v}_{n+1}^{(j)} + C^{vy}\big(\rho_{n+1}^{\mathrm{EK},J}\big) C^{yy}\big(\rho_{n+1}^{\mathrm{EK},J}\big)^{-1} \big(y_{n+1}^{\dagger} - \widehat{y}_{n+1}^{(j)}\big), \\
\rho_{n+1}^{\mathrm{EK},J} &= \frac{1}{J} \sum_{j=1}^{J} \delta_{\left(\widehat{v}_{n+1}^{(j)}, \widehat{y}_{n+1}^{(j)}\right)}.
\end{aligned}$$

Here $\xi_n^{(j)} \sim \mathcal{N}(0, \Sigma)$ i.i.d. with respect to both $n$ and $j$ and $\eta_n^{(j)} \sim \mathcal{N}(0, \Gamma)$ i.i.d. with respect to both $n$ and $j$. Furthermore the set of $\{\xi_n^{(j)}\}$ is independent of the set of $\{\eta_n^{(j)}\}$.

**Remark 2** Note that the components of $\mathrm{cov}(\rho_{n+1}^{\mathrm{EK},J})$ involve empirical learning of the covariances $\Sigma$ and $\Gamma$ of the $\{\xi_n^{(j)}\}_{j=1}^J$ and $\{\eta_{n+1}^{(j)}\}_{j=1}^J$ respectively. It is possible, and indeed often desirable, to directly input the matrices $\Sigma$ and $\Gamma$, only using empirical estimation for the covariance based on the $\{\Psi(v_n^{(j)})\}_{j=1}^J$ and $\{h(\widehat{v}_{n+1}^{(j)})\}_{j=1}^J$. In particular to use

$$C^{vy}(\rho_{n+1}^{\mathrm{EK},J}) \approx C^{vh,J}, \tag{2a}$$

$$C^{yy}(\rho_{n+1}^{\mathrm{EK},J}) \approx C^{hh,J} + \Gamma \tag{2b}$$

where $C^{vh,J}$ is the empirical cross-covariance between the $\{\Psi(v_n^{(j)})\}_{j=1}^J$ and the $\{h(\widehat{v}_{n+1}^{(j)})\}_{j=1}^J$, whilst $C^{hh,J}$ is the empirical covariance of the $\{h(\widehat{v}_{n+1}^{(j)})\}_{j=1}^J$. Note

that this leads to a definition of $C^{yy}(\rho_{n+1}^{\mathrm{EK},J})$ which differs from the original one, which employs an empirical approximation $\Gamma_{n+1}$ of the covariance $\Gamma$, formed from $\{\eta_{n+1}^{(j)}\}_{j=1}^{J}$.

From the particles defined by this algorithm we may define an empirical measure

$$\mu_n^{\mathrm{EK},J} = \frac{1}{J}\sum_{j=1}^{J} \delta_{v_n^{(j)}},$$

noting that for $n = 0$ we choose $v_0^{(j)} \sim \mu_0$ i.i.d. Practitioners like this methodology because it assigns equal weights to the particles and cannot suffer from the weight collapse arising in the particle filter. Furthermore, in the setting where $\Psi, h$ are both linear, and $\mu_n$ is Gaussian, $\mu_n^{\mathrm{EK},J}$ converges to $\mu_n$ as $J \to \infty$, at the Monte Carlo rate (Le Gland et al. 2011; Mandel et al. 2011). In this linear setting, with Gaussian noise and initial condition, the problem can be solved explicitly by the Kalman filter (Kalman 1960); this is a nonautonomous and nonlinear dynamical system for the mean and covariance of the (in this setting) Gaussian $\mu_n$. However forming and propagating $d \times d$ covariances, when $d \gg 1$, is impractical; ensemble methods, on the other hand, operate by computing a $J \times d$ low-rank approximation of the covariance and can be used when $d \gg 1$. Thus particle methods are of some practical value even in this linear Gaussian setting, avoiding the need to work with large covariances when $d \gg 1$.

We now outline recent new work aimed at determining the statistical accuracy of the mean field ensemble Kalman filter, beyond the linear Gaussian setting. The mean-field limit of the particle system, found by letting $J \to \infty$, is determined by the map

$$\begin{aligned}
\widehat{v}_{n+1} &= \Psi(v_n) + \xi_n, \\
\widehat{y}_{n+1} &= h(\widehat{v}_{n+1}) + \eta_{n+1}, \\
v_{n+1} &= \widehat{v}_{n+1} + C^{vy}\big(\rho_{n+1}^{\mathrm{EK}}\big)C^{yy}\big(\rho_{n+1}^{\mathrm{EK}}\big)^{-1}\big(y_{n+1}^{\dagger} - \widehat{y}_{n+1}\big), \\
(\widehat{v}_{n+1}, \widehat{y}_{n+1}) &\sim \rho_{n+1}^{\mathrm{EK}}.
\end{aligned}$$

Once again $\xi_n \sim \mathcal{N}(0, \Sigma)$ i.i.d. with respect to $n$ and $\eta_n \sim \mathcal{N}(0, \Gamma)$ i.i.d. with respect to $n$; and the set of $\{\xi_n\}$ is independent of the set of $\{\eta_n\}$.

**Remark 3** The covariances $C^{vy}$ and $C^{yy}$ can be calculated exactly from the covariance of $\widehat{v}_{n+1}$ with $h(\widehat{v}_{n+1})$, from the covariance of $h(\widehat{v}_{n+1})$ with itself and from $\Gamma$. The formulae are the $J \to \infty$ limit of those given in (2).

Note that the map $v_n \mapsto v_{n+1}$ is nonlinear, stochastic and nonautonomous in the sense that it depends on the observed data. But, furthermore, the map depends on the law of $v_n$ since knowledge of this law is required to define $\rho_{n+1}^{\mathrm{EK}}$. Thus we now focus on finding an evolution equation for the law of $v_n$, which we denote by $\mu_n^{\mathrm{EK}}$.

With this goal we define $\mathfrak{T}$ as follows:

$$\mathfrak{T}(\bullet, \bullet; \rho, y^\dagger) \colon \mathbb{R}^d \times \mathbb{R}^K \to \mathbb{R}^d;$$
$$(v, y) \mapsto v + C^{vy}(\rho)C^{yy}(\rho)^{-1}\big(y^\dagger - y\big),$$

noting that this map is linear for any given pair $(\rho, y^\dagger) \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^K) \times \mathbb{R}^K$. Recalling the notation $F_\sharp \nu$ for pushforward of a probability measure $\nu$ under a map $F$, we define the operator

$$\mathsf{T}(\rho; y^\dagger) = \big(\mathfrak{T}(\bullet, \bullet; \rho, y^\dagger)\big)_\sharp \rho.$$

Notice that $\mathsf{T}$, like $\mathsf{B}$, is nonlinear as a map on probability measures. With the definition of $\mathsf{T}$ it then follows (Calvello et al. 2025) that

$$\mu_{n+1}^{\mathrm{EK}} = \mathsf{T}(\mathsf{QP}\mu_n^{\mathrm{EK}}; y_{n+1}^\dagger), \quad \mu_0^{\mathrm{EK}} = \mu_0.$$

The mean-field map $v_n \mapsto v_{n+1}$ is now well-defined by coupling it to this nonlinear evolution equation for $\mu_n^{\mathrm{EK}}$, the law of $v_n$. We refer to it as a *mean-field map* precisely because of this dependence on its own law.

The key question we wish to address is the relationship between $\mu_n^{\mathrm{EK}}$ and $\mu_n$. To focus on this question we write their evolution equations in parallel:

$$\mu_{n+1}^{\mathrm{EK}} = \mathsf{T}(\mathsf{QP}\mu_n^{\mathrm{EK}}; y_{n+1}^\dagger),$$
$$\mu_{n+1} = \mathsf{B}(\mathsf{QP}\mu_n; y_{n+1}^\dagger).$$

From this it is clear that the key is to understand when $\mathsf{T} \approx \mathsf{B}$. In fact $\mathsf{T} \equiv \mathsf{B}$ on the set of Gaussian measures $\mathcal{G}(\mathbb{R}^d \times \mathbb{R}^K)$ (Calvello et al. 2025). This is related to the fact that, in the Gaussian case, the map $\mathsf{T}$ effects an exact transport from the joint distribution of state and data to the conditional distribution of state given data. To obtain an error bound between $\mu_n^{\mathrm{EK}}$ and $\mu_n$ it is helpful to define

$$\mathsf{G} \colon \mathcal{P} \to \mathcal{G},$$
$$\mathsf{G}\pi = \arg\min_{\mathfrak{p} \in \mathcal{G}} d_{\mathrm{KL}}(\pi \| \mathfrak{p}),$$

where $d_{\mathrm{KL}}(\pi \| \mathfrak{p})$ is the Kullback–Leibler (KL) divergence of $\pi$ from $\mathfrak{p}$, defined as

$$d_{\mathrm{KL}}(\pi \| \mathfrak{p}) = \int_{\mathbb{R}^d} \frac{\mathrm{d}\pi}{\mathrm{d}\mathfrak{p}} \log\left(\frac{\mathrm{d}\pi}{\mathrm{d}\mathfrak{p}}\right) \mathrm{d}\mathfrak{p}. \tag{3}$$

It then follows that $\mathsf{G}\pi = \mathcal{N}(\mathrm{mean}_\pi, \mathrm{cov}_\pi)$ (Sanz-Alonso et al. 2023, Theorem 4.7).

**Remark 4** The map $\mathsf{T}$ is not an *optimal* transport, in the sense of transporting one measure into another at minimal cost (Villani 2009), but it is a transport that is well-adapted to numerical implementation. Links between transport and data assimilation were pioneered by Sebastian Reich; see Reich and Cotter (2015) and the references therein.

When analyzing the particle filter we used a metric on random probability measures which reduces to the TV distance in the non-random case. Here we use a different metric; we have no need to consider random probability measures, but standard TV alone does not allow us to control first and second moments. Control of these moments is useful for the analysis because of the Gaussian approximations underlying the use of $\mathsf{T}$ rather than $\mathsf{B}$. We thus use a weighted TV metric, with weight $g(v) = 1 + |v|^2$. Specifically we define

$$\mathrm{d}_g(\mu, \nu) = \sup_{|f| \leq g} \left| \mu[f] - \nu[f] \right|,$$

with $\mu[f]$ and $\nu[f]$ defined as before. In order to state our theorem about closeness of $\mu_n$ and $\mu_n^{\mathrm{EK}}$ we introduce the following measure of how close the true filter $\{\mu_n\}$ is to being Gaussian, in the lifted space of state and data:

**Definition 1** We define the closeness between the filtering distribution $\mu_n$, lifted to the joint space of state and data, and its projection onto Gaussians, over $N$ steps:

$$\varepsilon := \sup_{0 \leq n \leq N} \mathrm{d}_g(\mathsf{GQP}\mu_n, \mathsf{QP}\mu_n).$$

For precise statement and proof of the following theorem, and in particular details of the conditions under which it holds, see Carrillo et al. (2022).

**Theorem 2** *Let $\mu_0^{\mathrm{EK}} = \mu_0$ and assume that $\|\Psi\|_{L^\infty}$, $\|h\|_{L^\infty}$ and $|h|_{C^{0,1}}$ are bounded by $r$. Then there is $C := C(N, r) \in (0, \infty)$ such that*

$$\sup_{0 \leq n \leq N} \mathrm{d}_g(\mu_n, \mu_n^{\mathrm{EK}}) \leq C\varepsilon.$$

**Remark 5** The following remarks explain and interpret the theorem:

- The theorem states that a distance between the mean-field ensemble Kalman filter and the true filter is bounded above by a quantity which measures how close the true filter is to being Gaussian. This is natural because $\mathsf{T}$ and $\mathsf{B}$ are identical on Gaussians.
- As for the particle filter, the proof follows a typical numerical analysis structure: Consistency + Stability Implies Convergence.
- Consistency: in Carrillo et al. (2022) a class of problems is identified within which there are sequences of problems along which $\varepsilon \to 0$. This demonstrates that there are problem classes within which the mean-field ensemble Kalman filter accurately approximates the true filter; at $\varepsilon = 0$ a Gaussian problem is obtained and so the result concerns a class of problems in a small neighbourhood of Gaussians.
- Stability: $\mathsf{P}$, $\mathsf{Q}$ are Lipschitz in $\mathrm{d}_g(\bullet, \bullet)$. Whilst $\mathsf{B}$, $\mathsf{T}$ are not Lipschitz, stability bounds can be proved in $\mathrm{d}_g(\bullet, \bullet)$, given certain information such as moment bounds and lower bounds on covariances.
- The theorem assumes upper bounds on the vector fields $\Psi$, $h$ defining the filtering problems, and the constant $C$ depends on these bounds. Furthermore $C$ grows exponentially with $N$.

## 4 Conclusions

Ensemble Kalman filters are widely used, yet many open problems remain concerning their properties. In this article we have concentrated on understanding the sense in which they approximate the true filtering distribution. This is important for understanding the sense in which probabilistic forecasts made using ensemble Kalman filters predict accurate statistics. In the context of weather forecasting, an example is determining the probability of rain at levels that will cause flooding; in the context of forecasting epidemics, an example is determining the probability of an infection peak that will overwhelm health services, in the absence of interventions. The work described in Theorem 2 is a first step to build theory in this area, for non-Gaussian problems. Many challenges lie ahead to develop this theory so that it applies under more complex and realistic conditions.

**Acknowledgements** This article is based on an invited lecture delivered by AMS at ICIAM 2023 in Tokyo. JAC was supported by the Advanced Grant Nonlocal-CPD (Nonlocal PDEs for Complex Particle Dynamics: Phase Transitions, Patterns and Synchronization) of the European Research Council Executive Agency (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 883363). JAC was also partially supported by the Engineering and Physical Sciences Research Council (EPSRC) under grants EP/T022132/1 and EP/V051121/1. FH is supported by start-up funds at the California Institute of Technology. FH was also supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) via project 390685813-GZ 2047/1-HCM. The work of AMS is supported by a Department of Defense Vannevar Bush Faculty Fellowship, and by the SciAI Center, funded by the Office of Naval Research (ONR), under Grant Number N00014-23-1-2729. UV is partially supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 810367), and by the Agence Nationale de la Recherche under grants ANR-21-CE40-0006 (SINEQ) and ANR-23-CE40-0027 (IPSO). The authors are grateful to Eviatar Bach, Ricardo Baptista, Edo Calvello and an anonymous referee for helpful suggestions relating to the exposition.

**Competing Interests** The authors have no conflicts of interest to declare that are relevant to the content of this chapter.

## References

Bickel, P., Li, B., Bengtsson, T.: Sharp failure rates for the bootstrap particle filter in high dimensions. In: Pushing the Limits of Contemporary Statistics: Contributions in Honor of Jayanta K. Ghosh. Inst. Math. Stat. (IMS) Collect., vol 3 (2008)

Calvello, E., Reich, S., Stuart, A.M.: Ensemble Kalman methods: a mean field perspective. Acta Numer. (2025)

Carrillo, J.A., Hoffmann, F., Stuart, A.M., Vaes, U.: The Mean Field Ensemble Kalman Filter: Near-Gaussian Setting. arXiv:2212.13239 (2022)

Crişan, D., Moral, P., Lyons, T.J.: Interacting particle systems approximations of the Kushner-Stratonovitch equation. Adv. Appl. Probab. **31**(3), 819–838 (1999)

Del Moral, P.: Nonlinear filtering: interacting particle resolution. C. R. Acad. Sci. Paris Ser. I Math. **325**(6), 653–658 (1997)
Del Moral, P., Guionnet, A.: On the stability of interacting processes with applications to filtering and genetic algorithms. Ann. Inst. H. Poincaré Probab. Statist. **37**(2), 155–194 (2001)
Doucet, A., de Freitas, N., Gordon, N. (eds.): Sequential Monte Carlo methods in Practice. Statistics for Engineering and Information Science. Springer-Verlag, New York (2001)
Evensen, G., Vossepoel, F.C., van Leeuwen, P.J.: Data assimilation fundamentals: a unified formulation of the state and parameter estimation problem. In: Springer Textbooks in Earth Sciences, Geography and Environment. Springer, Cham (2022)
Evensen, G.: Data Assimilation: The Ensemble Kalman Filter, 2nd ed. Springer, Berlin (2009)
Evensen, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. J. Geophys. Res. Oceans **99**(C5), 10143–10162 (1994)
Iglesias, M.A., Law, K.J.H., Stuart, A.M.: Ensemble Kalman methods for inverse problems. Inverse Probl. **29**(4), 045001, 20 (2013)
Kalman, R.E.: A new approach to linear filtering and prediction problems. Trans. ASME Ser. D. J. Basic Eng. **82**(1), 35–45 (1960)
Kalnay, E.: Atmospheric Modeling, Data Assimilation and Predictability. Cambridge University Press (2003)
Keeling, M.J., Eames, K.T.: Networks and epidemic models. J. R. Soc. Interface **2**(4), 295–307 (2005)
Law, K., Stuart, A., Zygalakis, K.: Data Assimilation. Texts in Applied Mathematics, vol. 62. Springer, Cham (2015)
Le Gland, F., Monbet, V., Tran, V.-D.: Large sample asymptotics for the ensemble Kalman filter. In: The Oxford Handbook of Nonlinear Filtering, pp. 598–631. Oxford University Press, Oxford (2011)
Mandel, J., Cobb, L., Beezley, J.D.: On the convergence of the ensemble Kalman filter. Appl. Math. **56**(6), 533–541 (2011)
Rebeschini, P., van Handel, R.: Can local particle filters beat the curse of dimensionality? Ann. Appl. Probab. **25**(5), 2809–2866 (2015)
Reich, S., Cotter, C.: Probabilistic Forecasting and Bayesian Data Assimilation. Cambridge University Press, New York (2015)
Sanz-Alonso, D., Stuart, A., Taeb, A.: Inverse Problems and Data Assimilation. LMS Student Lecture Notes, vol. 107. Cambridge University Press (2023)
Snyder, C., Bengtsson, T., Bickel, P., Anderson, J.: Obstacles to high-dimensional particle filtering. Mon. Weather Rev. **136**(12), 4629–4640 (2008)
van Leeuwen, P.J., Künsch, H.R., Nerger, L., Potthast, R., Reich, S.: Particle filters for high-dimensional geoscience applications: a review. Q. J. R. Meteorol. Soc. **145**(723), 2335–2365 (2019)
Villani, C.: Optimal Transport. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 338. Springer, Berlin (2009). Old and new