# Propagation of Chaos, Large Time Behavior and Accuracy for Filtering Algorithms

**Franca Hoffmann**

Computing and Mathematical Sciences
California Institute of Technology

**Lectures 4 & 5**

**Course on Gradient Flows in Data Science**
**ICTS, Bengaluru, India**

Caltech

# Overview

- **Part 1:** The Ensemble Kalman Filter (EnKF) in the near-Gaussian setting
- **Part 2:** Propagation of Chaos for Transport Ensemble Filters
- **Part 3:** Large-Time Behavior of the Mean-Field EnKF

| | **Setting** | **Result** |
|---|---|---|
| **Part 1** | EnKF, near-Gaussian, $J = \infty$, discrete time | finite-time, local accuracy |
| **Part 2** | TEFs, $J \to \infty$, discrete time | quantitative PoC |
| **Part 3** | EnKF, $J = \infty$, discrete and continuous time | all-time, global accuracy |

# PART 1

# Ensemble Kalman filtering in the near-Gaussian setting

collaboration with:

José A. Carrillo, Andrew M. Stuart and Urbain Vaes

SIAM Journal on Numerical Analysis (2024) 62:6, 2549-2587.

https://doi.org/10.1137/24M1628207

# Kalman Filter (Navigation)

## State Space Model

Dynamics Model: $v^{n+1} = Mv^n + \xi^n, \quad n \in \mathbb{Z}^+$

Data Model: $y^{n+1} = Hv^{n+1} + \eta^{n+1}, \quad n \in \mathbb{Z}^+$

Probabilistic Structure: $v^0 \sim \mathsf{N}(m^0, C^0), \quad \xi^n \sim \mathsf{N}(0, \Sigma), \quad \eta^n \sim \mathsf{N}(0, \Gamma)$

Probabilistic Structure: $v^0 \perp\!\!\!\perp \{\xi^n\} \perp\!\!\!\perp \{\eta^n\}$ independent


- Rudolph Kalman (1960): Kalman Filter.
- $\approx 51,600$ citations (Google Scholar 08/26).
- Apollo 11.
- The Algorithm:
- $Y^n = \{y^\ell\}_{\ell=1}^n$.
- $v^n | Y^n \sim \mathsf{N}(m^n, C^n)$.
- $(m^n, C^n) \mapsto (m^{n+1}, C^{n+1})$.

# Ensemble Kalman Filter

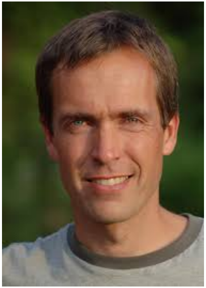
## State Space Model

$$\text{Dynamics Model:} \quad v^{n+1} = \Psi(v^n) + \xi^n, \quad n \in \mathbb{Z}^+$$

$$\text{Data Model:} \quad y^{n+1} = h(v^{n+1}) + \eta^{n+1}, \quad n \in \mathbb{Z}^+$$

$$\text{Probabilistic Structure:} \quad v^0 \sim \mathsf{N}(m^0, C^0), \quad \xi^n \sim \mathsf{N}(0, \Sigma), \quad \eta^n \sim \mathsf{N}(0, \Gamma)$$

$$\text{Probabilistic Structure:} \quad v^0 \perp\!\!\!\perp \{\xi^n\} \perp\!\!\!\perp \{\eta^n\} \text{ independent}$$



- Geir Evensen (1994): Ensemble Kalman Filter (EnKF).
- $\approx 7,200$ citations (Google Scholar 08/26).
- Originally ocean dynamics; now weather.
- $\mu^n := \mathrm{Law}(v^n|Y^n)$. (Here $Y^n = \{y^\ell\}_{\ell=1}^n$.)
- Mean-Field Algorithm:
- $(v^n, \mu^{EK,n}) \mapsto (v^{n+1}, \mu^{EK,n+1})$. $\quad \mu^{EK,n} := \mathrm{Law}(v^n)$.
- When is this approximation valid: $\mu^{EK,n} \approx \mu^n$?

# The discrete-time filtering problem

We consider the following stochastic dynamics and data model:

## Dynamics and observations

$$
\begin{aligned}
\textbf{Stochastic dynamics:} \quad & u^{n+1} = \Psi(u^n) + \xi^n, && \xi^n \sim \mathsf{N}(0, \Sigma), \\
\textbf{Data model:} \quad & y^{n+1} = h(u^{n+1}) + \eta^{n+1}, && \eta^{n+1} \sim \mathsf{N}(0, \Gamma).
\end{aligned}
$$

Independence assumption:

$$u^0 \perp\!\!\!\perp \xi^n \perp\!\!\!\perp \eta^n$$

Initial state: $u^0 \sim \mathsf{N}(m^0, C^0)$.

**Notations**:

- $\{u^n\}_{n \in [\![0, N]\!]}$ is the unknown state in $\mathbf{R}^d$.
- $\{y^n\}_{n \in [\![1, N]\!]}$ are the observations in $\mathbf{R}^K$.
- $\Psi \colon \mathbf{R}^d \to \mathbf{R}^d$ and $h \colon \mathbf{R}^d \to \mathbf{R}^K$ are nonlinear operators.
- $Y^n = \{y^{\dagger,1}, \dots, y^{\dagger,n}\}$ is a given realization of the data up to time $n$.

**Goal**: Approximate sequentially the probability measure $\mathbf{P}(u^n | Y^n)$ for $n \in [\![1, N]\!]$.

**Remark**: can be used to solve inverse problems too.

# Evolution of the true filtering distribution

## Filtering problem

| | | |
|---|---|---|
| Stochastic dynamics: | $u^{n+1} = \Psi(u^n) + \xi^n,$ | $\xi^n \sim \mathsf{N}(0, \Sigma),$ |
| Data model: | $y^{n+1} = h(u^{n+1}) + \eta^{n+1},$ | $\eta^{n+1} \sim \mathsf{N}(0, \Gamma).$ |

The true filtering distributions $\rho^n = \mathbf{P}(u^n | Y^n)$ evolves according to

$$\rho^{n+1} = B^n Q P \rho^n.$$

- $P\colon \mathcal{P}(\mathbf{R}^d) \to \mathcal{P}(\mathbf{R}^d)$

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

- $Q\colon \mathcal{P}(\mathbf{R}^d) \to \mathcal{P}(\mathbf{R}^d \times \mathbf{R}^K)$

$$Q\rho(u, y) = \exp\left(-\frac{1}{2}\left|y - h(u)\right|_\Gamma^2\right) \rho(u).$$

- $B^n\colon \mathcal{P}(\mathbf{R}^d \times \mathbf{R}^K) \to \mathcal{P}(\mathbf{R}^d)$ is the conditioning on the observation $y^{n+1} = y^{\dagger,n+1}$.

$$B^n\rho(u) = \frac{\rho(u, y^{\dagger,n+1})}{\int \rho(u, y^{\dagger,n+1})\,\mathrm{d}u}.$$

Schematically,

$$\mathbf{P}(u^n|Y^n) \xrightarrow{P} \mathbf{P}(u^{n+1}|Y^n) \xrightarrow{Q} \mathbf{P}\big((u^{n+1}, y^{n+1}) \mid Y^n\big) \xrightarrow{B^n} \mathbf{P}(u^{n+1}|Y^{n+1})$$

# The ensemble Kalman filter from a mean field perspective[1]

## One iteration of ensemble Kalman at the mean field level

$$
\begin{aligned}
\widehat{u}^{n+1} &= \psi(u^n) + \xi^n, & \xi^n &\sim \mathsf{N}(0,\Sigma),\\
\widehat{y}^{n+1} &= h(\widehat{u}^{n+1}) + \eta^{n+1}, & \eta^{n+1} &\sim \mathsf{N}(0,\Gamma),\\
u^{n+1} &= \widehat{u}^{n+1} + \mathcal{C}^{uy}(\widehat{\pi}^{n+1})\mathcal{C}^{yy}(\widehat{\pi}^{n+1})^{-1}\bigl(y^{\dagger,n+1} - \widehat{y}^{n+1}\bigr).
\end{aligned}
$$

Here $\widehat{\pi}^{n+1} = \mathrm{Law}(\widehat{u}^{n+1}, \widehat{y}^{n+1})$ and, for $\pi \in \mathcal{P}(\mathbf{R}^d \times \mathbf{R}^K)$,

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

The third equation may be rewritten

$$
u^{n+1} = \mathcal{T}(\widehat{u}^{n+1}, \widehat{y}^{n+1}; \widehat{\pi}^{n+1}, y^{\dagger,n+1}),
$$

with $\mathcal{T}$ the following mean-field transport map: for $\pi \in \mathcal{P}(\mathbf{R}^d \times \mathbf{R}^K)$ and $z \in \mathbf{R}^K$,

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

In the following, we use the shorthand notation $T^n\pi = \bigl(\mathcal{T}(\bullet, \bullet; \pi, y^{\dagger,n+1})\bigr)_\sharp \pi$ to define the operator $T^n : \mathcal{P}(\mathbf{R}^d \times \mathbf{R}^K) \to \mathcal{P}(\mathbf{R}^d)$.

[1] E. Calvello, S. Reich, and A. M. Stuart. Acta Numerica, arXiv 2209.11371, 2025.

# The ensemble Kalman filter from a mean field perspective[2]

## One iteration of ensemble Kalman at the mean field level

$$
\begin{aligned}
\widehat{u}^{n+1} &= \psi(u^n) + \xi^n, & \xi^n &\sim \mathsf{N}(0,\Sigma),\\
\widehat{y}^{n+1} &= h(\widehat{u}^{n+1}) + \eta^{n+1}, & \eta^{n+1} &\sim \mathsf{N}(0,\Gamma),\\
u^{n+1} &= \widehat{u}^{n+1} + \mathcal{C}^{uy}(\widehat{\pi}^{n+1})\mathcal{C}^{yy}(\widehat{\pi}^{n+1})^{-1}\big(y^{\dagger,n+1} - \widehat{y}^{n+1}\big).
\end{aligned}
$$

Let $\rho^{EK,n} = \mathrm{Law}(u^n)$. Then

$$
\begin{aligned}
\rho^{n+1} &= B^n Q P \rho^n && \text{(True filtering distribution)}\\
\rho^{EK,n+1} &= T^n Q P \rho^{EK,n} && \text{(Ensemble Kalman filtering distribution)}
\end{aligned}
$$

Schematically, for mean field ensemble Kalman:

$$
\mathrm{Law}(u^n) \xrightarrow{P} \widehat{\rho}^{n+1} \xrightarrow{Q} \mathrm{Law}\big((\widehat{u}^{n+1}, \widehat{y}^{n+1})\big) \xrightarrow{T^n} \mathrm{Law}(u^{n+1})
$$

[2] E. Calvello, S. Reich, and A. M. Stuart. Acta Numerica, arXiv 2209.11371, 2025.

# Gaussian Projections

## Filtering distributions

$$\rho_{n+1} = B^n Q P \rho_n \qquad \text{(True filtering distribution)}$$

$$\rho_{n+1}^{EK} = T^n Q P \rho_{n+1}^{EK} \qquad \text{(Ensemble Kalman filtering distribution)}$$

Define the Gaussian projection

$$G\rho := \mathsf{N}\big(\mathcal{M}(\rho), \mathcal{C}(\rho)\big).$$

where

$$\mathcal{M}(\rho) = \int \theta \rho(\mathrm{d}\theta)\,, \qquad \mathcal{C}(\rho) = \int \big(\theta - \mathcal{M}(\rho)\big) \otimes \big(\theta - \mathcal{M}(\rho)\big)\rho(\mathrm{d}\theta)\,.$$

**Key result:** For any Gaussian $\rho$

$$T^n \rho = B^n \rho.$$

$\rightsquigarrow$ as expected, mean field ensemble Kalman is exact in the Gaussian setting.

**Key question:** If $\rho_n$ is close to Gaussian (in a suitable sense), how close is $\rho_n^{EK}$ to $\rho_n$?

# Distance between distributions

## Weighted Total Variation Metric

Let $g \colon \mathbf{R}^n \to [1, \infty)$ be given by $g(v) = 1 + |v|^2$. Define $d_g \colon \mathcal{P}(\mathbf{R}^n) \times \mathcal{P}(\mathbf{R}^n) \to \mathbf{R}$ by

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

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.

- If $\mu_1, \mu_2$ have Lebesgue densities $\rho_1$, $\rho_2$, then

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

- Unlike the usual total variation distance, $d_g$ enables to control the differences $|\mathcal{M}(\mu_1) - \mathcal{M}(\mu_2)|$ and $\|\mathcal{C}(\mu_1) - \mathcal{C}(\mu_2)\|$.

# Estimate in the near-Gaussian setting

## What we want to show

In the near-Gaussian setting, the probability measures $\{\rho_n^{EK}\}_{n\in[\![1,N]\!]}$ are close to the true filtering distributions $\{\rho_n\}_{n\in[\![1,N]\!]}$.

**Assumptions**:

- (Near-Gaussian setting) There is $\varepsilon > 0$ such that
$$d_g(QP\rho_n, GQP\rho_n) \le \varepsilon \qquad \forall n \in [\![0,N]\!].$$
- (Boundedness) There is $\kappa < \infty$ such that
$$\|\Psi\|_{L^\infty(\mathbf{R}^d)} \vee \|h\|_{L^\infty(\mathbf{R}^d)} \le \kappa.$$
- (Lipschitz continuity) The map $h$ is globally Lipschitz continuous.
- (Non-degenerate noise) $\Sigma \ge \sigma I_d$ and $\Gamma \ge \gamma I_K$ for positive $\sigma$ and $\gamma$.

**Main theorem**: Initialize at $\rho_0 = \rho_0^{EK}$ Gaussian. Then there exists $C_N$ independent of $\varepsilon$ such that
$$\forall n \in [\![0,N]\!], \qquad d_g(\rho_n^{EK}, \rho_n) \le C_N \varepsilon.$$

# Strategy of proof

For $R > 0$, define the set of probability measures with bounded first and second moments,

$$\mathcal{P}_R(\mathbf{R}^d) = \left\{ \mu \in \mathcal{P}(\mathbf{R}^d) : \max\left\{ |\mathcal{M}(\mu)|, \|\mathcal{C}(\mu)\|, \|\mathcal{C}(\mu)\|^{-1} \right\} \leq R \right\}.$$

- The maps $P$, $Q$ are globally Lipschitz on $\mathcal{P}(\mathbf{R}^d)$.
- The maps $B^n$ satisfy: $\forall \mu \in \mathcal{P}(\mathbf{R}^d)$,
$$d_g(B^n GQP\mu, B^n QP\mu) \leq \ell_B \, d_g(GQP\mu, QP\mu).$$
- The maps $T^n$ satisfy: $\forall (\mu, \nu) \in \mathcal{P}(\mathbf{R}^d) \times \mathcal{P}_R(\mathbf{R}^d \times \mathbf{R}^K)$,
$$d_g(T^n QP\mu, T^n \nu) \leq \ell_T(R) \, d_g(QP\mu, \nu),$$
- Moment bounds: $\exists R_* > 0$ such that for any $\mu \in \mathcal{P}(\mathbf{R}^d)$, we have
$$P\mu \in \mathcal{P}_{R_*}(\mathbf{R}^d), \qquad QP\mu \in \mathcal{P}_{R_*}(\mathbf{R}^d \times \mathbf{R}^K).$$

The main idea of the proof is to use the triangle inequality. Since

$$d_g(\rho_{n+1}^{EK}, \rho_{n+1}) = d_g\left(T^n QP\rho_n^{EK}, B^n QP\rho_n\right)$$

and "$T^n G = B_n G$" we have

$$\begin{aligned} d_g\left(T^n QP\rho_n^{EK}, B^n QP\rho_n\right) &\leq d_g\left(T^n QP\rho_n^{EK}, T^n QP\rho_n\right) + d_g\left(T^n QP\rho_n, T^n GQP\rho_n\right) \\ &\quad + d_g\left(B^n GQP\rho_n, B^n QP\rho_n\right) \\ &\leq c_1 d_g(\rho_n^{EK}, \rho_n) + c_2\varepsilon. \end{aligned}$$

# Removing boundedness & finite particle approximations[3]

Extension to unbounded (linearly growing) $\Psi, h$, and to the finite-particle EnKF $\rho_n^{EK,J}$:

1. **Stability (mean field)**: the same type of bound holds for Lipschitz $\Psi, h$ with linear growth:
$$d_g(\rho_N^{EK}, \rho_N) \leq C_N \max_{n \in [\![0, N-1]\!]} d_g(QP\rho_n, GQP\rho_n);$$
2. **Error estimate (mean field)**: if $(\Psi, h)$ lie within $\varepsilon$ (sup norm) of some affine $(\Psi_0, h_0)$, then
$$d_g(\rho_N^{EK}, \rho_N) \leq C\varepsilon;$$
3. **Error estimate (finite particle)**: for linear $h$ and globally Lipschitz $\Psi$,
$$\Big(\mathbf{E}\big|\rho_N^{EK,J}[\phi] - \rho_N[\phi]\big|^2\Big)^{1/2} \leq C\Big(\frac{1}{\sqrt{J}} + \varepsilon\Big).$$

**Summary**: the first quantitative error bound between the finite-particle EnKF and the true filter, allowing unbounded, linearly-growing dynamics and observations.

[3] E. Calvello, P. Monmarché, A. M. Stuart, and U. Vaes. SIAM Journal on Numerical Analysis, 2026.

**Remaining open questions at that point**:

- Other 'close to Gaussian' assumptions and their interpretation in practice;
- Obtain error estimates with a better scaling with respect to $N$;
- Obtain error estimates for continuous-time Gaussian filtering methods;
- Propagation of chaos;
- Generalization to other filtering methods.

# PART 2
# Propagation of Chaos

collaboration with:
Frederic J. N. Jorgensen, Ricardo Baptista and Youssef Marzouk
"Quantitative Wasserstein Propagation of Chaos for Transport Ensemble Filters"
https://arxiv.org/abs/2606.25346

# The Filtering Problem

## Hidden Markov Model

$$\begin{aligned} \text{Dynamics:} \quad & X^{n+1} = \Psi(X^n) + \xi^n, \\ \text{Observation:} \quad & Y^{n+1} = h(X^{n+1}) + \eta^{n+1}, \end{aligned}$$

with $X^0 \sim \mu^0$, $\xi^n \sim \nu_\xi^n$, $\eta^n \sim \nu_\eta^n$ jointly independent.

**Goal**: characterize the true filtering (analysis) distribution $\mu_X^{a,n} = \rho_n$

$$\mu_X^{a,n} := \mathrm{Law}\big(X^n \mid Y^{1:n} = y^{1:n}\big).$$

## Recursion for true filter

$$\mu_X^{f,n} = Q^n P^{n-1} \mu_X^{a,n-1} \qquad \text{(Forecast Step: Dynamics \& Observation)}$$

$$\mu_X^{a,n} = B_{y^n} \mu_X^{f,n} \qquad \text{(Analysis Step: Conditioning)}$$

**Notation:** $B_{y^n} = B^n$, and $P^n, Q^n$ with an explicit time index $n$ as noise laws are allowed to vary with $n$.

# From Exact Filtering to Ensemble Methods

Computing $\mu_X^{a,n} = B_{y^n} Q^n P^{n-1} \mu_X^{a,n-1}$ exactly is generally intractable:

- Measures $\mu_X^{a,n}$ may be of black-box nature;
- Evaluating the conditioning operator $B_{y^n}$ requires high-dimensional integration.

**Ensemble filters** resolve this using $J$ particles per step (superscript $f$ = forecast, $a$ = analysis):

- represent $\mu_X^{a,n-1}$ by an empirical measure of particles $\{x_\ell^{a,n-1}\}_{\ell=1}^J$;
- propagate each particle through the dynamics/observation model to get a forecast ensemble $\{(x_\ell^{f,n}, y_\ell^{f,n})\}_{\ell=1}^J$, with empirical forecast measure

$$\widehat{\mu}_{XY,J}^{f,n} := \frac{1}{J} \sum_{\ell=1}^{J} \delta_{(x_\ell^{f,n}, y_\ell^{f,n})};$$

- approximate $B_{y^n}$ by a transport map $T_{y^n}^{\pi}$ acting particle-wise, using auxiliary randomness $\omega \sim \kappa$ (a fixed, known reference distribution, e.g. allowing stochastic maps):

$$\tilde{B}_y \pi := \left(T_y^{\pi}\right)_\sharp (\pi \otimes \kappa) \approx B_y \pi.$$

Particles interact through their shared dependence on the empirical forecast measure $\widehat{\mu}_{XY,J}^{f,n}$.

# Transport Ensemble Filters (TEFs)

## Algorithm: Transport Ensemble Filter

**Forecast:**

$$x_\ell^{f,n} = \Psi(x_\ell^{a,n-1}) + \xi_\ell^{n-1}, \qquad y_\ell^{f,n} = h(x_\ell^{f,n}) + \eta_\ell^n.$$

**Analysis:**

$$\widehat{\mu}_{XY,J}^{f,n} = \frac{1}{J}\sum_{\ell=1}^{J} \delta_{(x_\ell^{f,n}, y_\ell^{f,n})}, \qquad x_\ell^{a,n} = T_{y^n}^{\widehat{\mu}_{XY,J}^{f,n}}(x_\ell^{f,n}, y_\ell^{f,n}, \omega_\ell^n).$$

TEFs cover many algorithms, depending on the choice of map $T$, including

- **Ensemble Kalman Filter (EnKF)**: $T$ is affine in $(x, y)$ **[Evensen 1994]**

$$T_{y^n}^{\pi}(x, y) = x + \mathrm{Cov}(\pi)_{XY}\ \mathrm{Cov}(\pi_Y)^{\dagger}\,(y^n - y).$$

- **Ensemble Stochastic Map Filter (EnSMF)**: $T$ is a nonlinear triangular transport map **[Spantini, Baptista, Marzouk 2022]**;
- **Conditional Mean Filter (CMF)** **[Hoang et al. 2021]**, and other learning-enhanced ensemble filters **[Bach et al. 2025]**.

# Propagation of Chaos

**Propagation of chaos** [Kac 1956], [Sznitman 1991]: particles should become asymptotically independent.

## Questions:

1. What is the **mean-field law** $\tilde{\mu}_X^{a,n}$ that $\widehat{\mu}_{X,J}^{a,n}$ converges to as $J \to \infty$?
2. Convergence in which sense?
3. At what **rate** does $\widehat{\mu}_{X,J}^{a,n}$ converge to $\tilde{\mu}_X^{a,n}$?

**Why $\mathcal{P}_2$ and $W_2$?** Most TEFs (EnKF, EnSMF, CMF, ...) are built from first/second moments; $\mathcal{P}_2(\mathbb{R}^d \times \mathbb{R}^K)$ with the Wasserstein distance gives a natural, statistically meaningful geometry for both algorithmic and empirical-measure error.

# Deriving the Mean-Field Dynamics for TEFs

Replace the random empirical forecast $\widehat{\mu}^{f,n}_{XY,J}$ by its deterministic limit $\tilde{\mu}^{f,n}_{XY}$:

- a representative particle $U^{n-1} \sim \tilde{\mu}^{a,n-1}_X$ evolves as

$$U^{f,n} = \Psi(U^{n-1}) + \xi^{n-1}, \qquad Y^{f,n} = h(U^{f,n}) + \eta^n,$$

- it is updated using the map evaluated at the mean-field forecast law,

$$U^n = T^{\tilde{\mu}^{f,n}_{XY}}_{y^n}\left(U^{f,n}, Y^{f,n}, \omega^n\right).$$

## Mean-field recursion

$$\mu^{a,n}_X = B_{y^n} Q^n P^{n-1} \mu^{a,n-1}_X \qquad \text{(True filtering distribution)}$$
$$\tilde{\mu}^{a,n}_X = \tilde{B}_{y^n} Q^n P^{n-1} \tilde{\mu}^{a,n-1}_X. \qquad \text{(TEF distribution)}$$

Recall:

$$\tilde{B}_y \pi := \left(T^{\pi}_y\right)_\sharp (\pi \otimes \kappa) \approx B_y \pi.$$

- **Ensemble Kalman Filter (EnKF)**: The map $T$ is given by

$$T^{\pi}_{y^n}(x, y) = x + \mathrm{Cov}(\pi)_{XY}\ \mathrm{Cov}(\pi_Y)^{\dagger}\,(y^n - y).$$

and we denoted $\tilde{\mu}^{a,n}_X = \rho^{EK}_n$ and $\tilde{B}_{y^n} = T^n$.

# Propagation of Chaos by Synchronous Coupling

Construct an i.i.d. mean-field ensemble $\{v_\ell^{a,n}\}_{\ell=1}^J$, driven by the same noise as the interacting particles at every step $n$ (synchronous coupling [McKean 1967], [Sznitman 1991]):

**interacting particle system**

$$
\begin{aligned}
x_\ell^{f,n} &= \Psi(x_\ell^{a,n-1}) + \xi_\ell^{n-1} \\
y_\ell^{f,n} &= h(x_\ell^{f,n}) + \eta_\ell^n \\
x_\ell^{a,n} &= T_{y^n}^{\widehat{\mu}_{XY,J}^{f,n}}(x_\ell^{f,n}, y_\ell^{f,n}, \omega_\ell^n)
\end{aligned}
$$

**i.i.d. mean-field**

$$
\begin{aligned}
v_\ell^{f,n} &= \Psi(v_\ell^{a,n-1}) + \xi_\ell^{n-1} \\
o_\ell^{f,n} &= h(v_\ell^{f,n}) + \eta_\ell^n \\
v_\ell^{a,n} &= T_{y^n}^{\tilde{\mu}_{XY}^{f,n}}(v_\ell^{f,n}, o_\ell^{f,n}, \omega_\ell^n)
\end{aligned}
$$

Same randomness $\Rightarrow$ any discrepancy is due only to the interaction through

$$\widehat{\mu}_{XY,J}^{f,n} \text{ vs. } \tilde{\mu}_{XY}^{f,n}.$$

**Decomposing the error:**

$$W_2(\widehat{\mu}_{X,J}^{a,n}, \tilde{\mu}_X^{a,n}) \le \underbrace{W_2(\widehat{\mu}_{X,J}^{a,n}, \tilde{\mu}_{X,J}^{a,n})}_{\text{interaction error}} + \underbrace{W_2(\tilde{\mu}_{X,J}^{a,n}, \tilde{\mu}_X^{a,n})}_{\text{i.i.d. sampling error}}$$

- **Interaction error**: controlled by our main theorem, at the optimal Monte Carlo rate $J^{-1/2}$;
- **i.i.d. sampling error**: unavoidable empirical-measure error, no algorithm can improve it without more structure.

# Assumptions (Informally)

## Definition: sub-Gaussian random variables [Vershynin 2018]

$X$ is sub-Gaussian if its tails decay at least as fast as a Gaussian's:

$$\|X\|_{\psi_2} := \inf\{t > 0 : \mathbf{E}\exp(X^2/t^2) \le 2\} < \infty, \quad \text{e.g. } \mathcal{P}(|X| > t) \lesssim \exp(-ct^2/\|X\|_{\psi_2}^2).$$

## Dynamics

- $\Psi$ and $h$ are Lipschitz; process/observation noise and initialization are sub-Gaussian.

## Transport map $T$

- **Lipschitz after estimation**: $T_y^\mu$ is Lipschitz in $(x, y, \omega)$, with constant allowed to grow polynomially in $\mathrm{Tr}\ \mathrm{Cov}(\mu)$;
- **Lipschitz estimation stability**: $T_y^\mu$ depends continuously on the measure argument $\mu$ (in a suitable covariance-based sense);
- **Transport sensitivity**: the update $T_{y^\star}^\mu(x, y, \omega) - x$ is controlled by $\|y - y^\star\|$, $\|\omega\|$ and covariance of $\mu$;
- Auxiliary randomness $\omega \sim \kappa$ is sub-Gaussian.

Remark: Both the EnKF and EnSMF satisfy these assumptions.

# Main Result: Convergence to i.i.d. Mean-Field Particles

## Theorem Interaction Error (informal)

Under Lipschitz/stability conditions on $T$ and sub-Gaussian noise, for fixed horizon $N$, there is a coupling of $\widehat{\mu}^{a,n}_{X,J}$ with i.i.d. mean-field particles $\tilde{\mu}^{a,n}_{X,J}$ such that, with probability at least $1-\frac{1}{k}$,

$$\sup_{1\le n\le N} W_2\big(\widehat{\mu}^{a,n}_{X,J},\tilde{\mu}^{a,n}_{X,J}\big)\le C_N\,\frac{(\log k)^{1+N/2}}{\sqrt{J}}.$$

- Holds **with high probability** jointly over particle draws and the random observation path $y^{1:N}$;
- Taking expectations gives

  $$\mathbf{E}\Big(\sup_{1\le n\le N} W_2\big(\widehat{\mu}^{a,n}_{X,J},\tilde{\mu}^{a,n}_{X,J}\big)\Big)\lesssim J^{-1/2}$$

  For this, no sub-Gaussian tails are needed, moment control of the driving noises and the initialization is sufficient;
- Finite-time result: $C_N$ diverges as $N\to\infty$;
- **Proof idea**: synchronous coupling + inductive control of moments/tails of the forecast laws under conditioning on $y^{1:n}$.

# Mean-Field Convergence: Full Statement

Standard i.i.d. Wasserstein concentration [Fournier, Guillin 2015]: with probability $1 - \frac{1}{k}$,

$$\sup_{1 \le n \le N} W_2\big(\tilde{\mu}^{a,n}_{X,J}, \tilde{\mu}^{a,n}_{X}\big) \;\lesssim\; \gamma^d_{k,J} \;=\; \begin{cases} \left(\frac{\log k}{J}\right)^{1/d} & \text{if } \; d > 4 \\ \left(\frac{\log k}{J}\right)^{1/4} \big(\log(2+J)\big)^{1/2} & \text{if } \; d = 4 \\ \left(\frac{\log k}{J}\right)^{1/4} & \text{if } \; d < 4\,. \end{cases}$$

$$\boxed{\begin{array}{c}\hat{\mu}^{a,n}_{X,J} \\ \textit{interacting} \\ \textit{finite particles}\end{array}} \xrightarrow{J^{-1/2}} \boxed{\begin{array}{c}\tilde{\mu}^{a,n}_{X,J} \\ \textit{i.i.d.} \\ \textit{finite particles}\end{array}} \xrightarrow{J^{-1/d}} \boxed{\begin{array}{c}\tilde{\mu}^{a,n}_{X} \\ \textit{mean-field} \\ \textit{limit}\end{array}}$$

optimal Monte Carlo rate — curse of dimensionality

**Full statement:** with probability at least $1 - \frac{1}{k}$,

$$\sup_{1 \le n \le N} W_2\big(\widehat{\mu}^{a,n}_{X,J}, \tilde{\mu}^{a,n}_{X}\big) \;\lesssim\; \gamma^d_{k,J} + \frac{(\log k)^{1+N/2}}{\sqrt{J}}.$$

- For fixed $k, N$: the interaction term ($J^{-1/2}$) is dominated by the empirical-measure term ($J^{-1/d}$) once $d > 4$;
- This $J^{-1/d}$ curse of dimensionality is sharp in general: it cannot be avoided without further structure;
- We isolate the avoidable algorithmic error from the unavoidable sampling error.

# Proof Sketch (1/2): Splitting the Coupling Error

To control $W_2(\widehat{\mu}^{a,n}_{X,J}, \tilde{\mu}^{a,n}_{X,J})$, define $\varepsilon^n := \sqrt{\frac{1}{J}\sum_{\ell=1}^{J} \|x^{a,n}_\ell - v^{a,n}_\ell\|_2^2}$ (with $\varepsilon^0 = 0$), where

$$x^{a,n}_\ell = T^{\widehat{\mu}^{f,n}_{XY,J}}_{y^n}(x^{f,n}_\ell, y^{f,n}_\ell, \omega^n_\ell), \qquad v^{a,n}_\ell = T^{\tilde{\mu}^{f,n}_{XY}}_{y^n}(v^{f,n}_\ell, o^{f,n}_\ell, \omega^n_\ell).$$

1. **Triangle inequality**: Insert a mixed term $T^{\widehat{\mu}^{f,n}_{XY,J}}_{y^n}(v^{f,n}_\ell, o^{f,n}_\ell, \omega^n_\ell)$ (writing $\|\cdot\|_{2,J} := \sqrt{\frac{1}{J}\sum_\ell \|\cdot\|_2^2}$):

$$\varepsilon^n \le \underbrace{\left\| T^{\widehat{\mu}^{f,n}_{XY,J}}_{y^n}(x^{f,n}_\ell, y^{f,n}_\ell, \omega^n_\ell) - T^{\widehat{\mu}^{f,n}_{XY,J}}_{y^n}(v^{f,n}_\ell, o^{f,n}_\ell, \omega^n_\ell) \right\|_{2,J}}_{\text{Term A}}$$

$$+ \underbrace{\left\| T^{\widehat{\mu}^{f,n}_{XY,J}}_{y^n}(v^{f,n}_\ell, o^{f,n}_\ell, \omega^n_\ell) - T^{\tilde{\mu}^{f,n}_{XY}}_{y^n}(v^{f,n}_\ell, o^{f,n}_\ell, \omega^n_\ell) \right\|_{2,J}}_{\text{Term B}}$$

- **Term A**: same map $T^{\widehat{\mu}^{f,n}_{XY,J}}_{y^n}$, applied to the interacting vs. mean-field forecast inputs;
- **Term B**: same (mean-field) inputs, plugged into the map under two different measure arguments, $\widehat{\mu}^{f,n}_{XY,J}$ (empirical, random) vs. $\tilde{\mu}^{f,n}_{XY}$ (mean-field, deterministic)

# Proof Sketch (2/2): Bounding Each Term and Unrolling

2. **Term A (contraction)**: Lipschitzness of $T$ in $(x, y, \omega)$ for fixed measure argument (Assumption 1a) + coupled dynamics give Term A $\lesssim L_T^{\widehat{\mu}^{f,n}_{XY,J}} \varepsilon^{n-1}$;
3. **Term B (interaction)**: "estimation stability" of $T$ in its measure argument (Assumption 1b) gives Term B $\lesssim \|\operatorname{Cov}(\widehat{\mu}^{f,n}_{XY,J}) - \operatorname{Cov}(\tilde{\mu}^{f,n}_{XY})\|_2 \lesssim \sqrt{\log k/J}$, via empirical covariance concentration;
4. **Moment control**: both bounds need high-probability sub-Gaussian tails for $\widehat{\mu}^{f,n}_{XY,J}$ and $\tilde{\mu}^{f,n}_{XY}$, even conditionally on $Y^{1:n}$ — proved by induction on $n$;
5. **Induction**: combining Steps 1–4 gives the self-recursion $\varepsilon^n \lesssim (1 + \delta_{k,J})\varepsilon^{n-1} + \delta_{k,J}^{1/2}$ with $\delta_{k,J} := \log k/\sqrt{J}$, which (union bound, probability $1 - \frac{1}{k}$) inductively over $n = 1, \ldots, N$ gives:

$$\sup_{1 \leq n \leq N} \varepsilon^n \;\lesssim\; \frac{(\log k)^{1+N/2}}{\sqrt{J}}.$$

# Application 1: The Ensemble Kalman Filter (EnKF)

The EnKF map is **affine** in $(x, y)$:

$$T^{\pi}_{y^n}(x, y) = x + \mathrm{Cov}(\pi)_{XY} \left(\mathrm{Cov}(h_\sharp \pi_X) + \Lambda\right)^{-1} (y^n - y).$$

## Theorem (EnKF, informal)

Under Lipschitz dynamics/observations and sub-Gaussian noise with $\mathrm{Cov}(\eta^n) \succeq \lambda_{\min} I_K$, for all $k \geq 2$ and $J \geq (\log k)^2$, with probability $1 - \frac{1}{k}$,

$$W_2\left(\widehat{\mu}^{a,n}_{X,J}, \tilde{\mu}^{a,n}_{X,J}\right) \leq C_N \, \frac{(\log k)^{1+n/2}}{\sqrt{J}}, \qquad 1 \leq n \leq N.$$

**Contribution:** first non-asymptotic, high-probability Wasserstein convergence guarantee for the discrete-time EnKF with sub-Gaussian noise, general Lipschitz dynamics, and random observation paths.

# Application 2: Ensemble Stochastic Map Filter (EnSMF)

**Motivation**: the EnKF's affine update is biased whenever the true posterior depends nonlinearly on the observations (e.g., non-Gaussian targets) **[Spantini et al. 2022], [Jorgensen, Marzouk 2025]**.

- EnSMF learns a **triangular transport map** $S(x, y)$ pushing the joint forecast to $\gamma_{\mathrm{G}} \otimes \mu_Y^{f,n}$ (Gaussian $\times$ observation marginal);
- Each component $S_k(x, y) = \alpha_k x_k + c_k + \sum_i \theta_k^i f_k^i(x_{1:k-1}, y)$ is a monomial/RBF expansion;
- Analysis step: $\mu_X^{a,n} = \big(S(\cdot, y^n)^{-1} \circ S\big)_\sharp \mu_{XY}^{f,n}$ (the "stochastic map").

**Contribution:** verifying Lipschitz/stability of the (nonlinear!) triangular map yields the first non-asymptotic mean-field convergence guarantee for a nonlinear TEF.

# Improving on Prior EnKF Mean-Field Theory

Prior results **[Le Gland et al. 2011], [Mandel et al. 2011], [Kwiatkowski, Mandel 2015]** established $O(J^{-1/2})$ convergence in expectation, for fixed observations, in the linear-Gaussian setting.

**Our contributions:**

- random (not fixed) observation path $Y^{1:N}$;
- high-probability, not just expectation, guarantees;
- bounds in Wasserstein distance $W_2$ (distributional, not single test functions);
- sub-Gaussian noise (not just Gaussian);
- Lipschitz (not just linear) observation map $h$;
- fully non-asymptotic, explicit finite-$J$ rates.

# Summary & Open Questions: Propagation of Chaos

In Part 2,

- we introduced **transport ensemble filters (TEFs)**, a broad class covering the EnKF and nonlinear filters like the EnSMF;
- we derived the **mean-field limit** of the interacting particle system;
- we proved **non-asymptotic, high-probability** Wasserstein propagation of chaos at the optimal Monte Carlo rate $J^{-1/2}$;
- applying this to the EnKF and EnSMF gave the **first** finite-sample convergence guarantees for both.

**Perspectives for future work**:

- Uniform-in-time / long-horizon bounds (current constants blow up as $N \to \infty$);
- Structural assumptions (dissipation, observability, inflation/localization **[Kelly et al. 2014], [Tong et al. 2015]**) to control long-time behavior;
- Sharper rates in structured/low-effective-dimension settings;
- Extending the framework to other nonlinear ensemble filters (CMF, learning-enhanced filters).

# PART 3
# Large-Time Behavior of the Mean-Field EnKF

collaboration with:
Sangmin Park and Andrew M. Stuart
"Large-time behavior and accuracy of the Mean-Field Ensemble Kalman Filter in the Linear Detectable Setting"
https://arxiv.org/abs/2606.07729

# Setting: Linear Signal, Linear Observations, Gaussian Noise

## Signal and observation

$$X^{n+1} = AX^n + \xi^n, \qquad Y^{n+1} = HX^{n+1} + \eta^{n+1}, \qquad \xi^n \sim \mathsf{N}(0,\Sigma),\ \eta^n \sim \mathsf{N}(0,\Gamma).$$

## Filtering distributions (initialized at $X_0 \sim \mu$)

$$\rho^{\mu}_{n+1} = B\big(QP\rho^{\mu}_n\,; y^{n+1}\big) \qquad \text{(True filter)}$$

$$\rho^{EK,\mu}_{n+1} = T\big(QP\rho^{EK,\mu}_n\,; y^{n+1}\big) \qquad \text{(Ensemble Kalman filter)}$$

- $P$, $Q$, $B^n = B(\cdot; y^{n+1})$, $T^n = T(\cdot; y^{n+1})$ as in Part 1 (noise distributions independent of time);
- **Key fact**: $\mu$ Gaussian $\Rightarrow$ $\rho^{EK,\mu}_n = \rho^{\mu}_n$ for **all** $n$;

**Question**: what if $\mu$ (hence $\rho^{\mu}_0$) is **not** Gaussian?

- Started far from Gaussian, does the mean-field EnKF (MFEnKF) eventually agree with the optimal filter? (not obvious: each EnKF step can poorly approximate Bayes when far from Gaussian — errors could accumulate)
- **Main result**: under detectability, they don't — MFEnKF is exponentially attracted to the Gaussian manifold, and becomes asymptotically exact.

# Key contributions

(C1) A variational approximation of the (linear) Bayesian inverse problem using a covariance-weighted optimal transport metric, recovering the MFEnKF;

(C2) Exponential contraction of the MFEnKF distributions towards the Gaussian manifold with respect to a covariance-weighted optimal transport metric (with the contraction rates explicitly characterized in terms of the covariance);

(C3) Large-time stability and accuracy of the continuous-time MFEnKF in relation to the true filtering distribution.

## True/Optimal Filter (initialized at $X_0 \sim \mu$)

$$\widehat{\rho}^{\mu}_{n+1} = P\rho^{\mu}_{n} \qquad \text{(Dynamics)}$$

$$\rho^{\mu}_{n+1} = B\big(Q\widehat{\rho}^{\mu}_{n+1}\,;y^{n+1}\big) \qquad \text{(Observations)}$$

## MFEnKF (initialized at $X_0 \sim \mu$)

$$\widehat{\rho}^{EK,\mu}_{n+1} = P\rho^{EK,\mu}_{n} \qquad \text{(Dyamics)}$$

$$\rho^{EK,\mu}_{n+1} = T^{n}\big(Q\widehat{\rho}^{EK,\mu}_{n+1}\big) \qquad \text{(Observations)}$$

# Overview of Equations

## Discrete time

| | Particle Evolution | Evolution of Distribution |
|---|---|---|
| **True Filter** | $X^{n+1} = AX^n + \xi^n$<br>$Y^{n+1} = HX^{n+1} + \eta^{n+1}$ | $\widehat{\rho}^{\mu}_{n+1} = P\rho^{\mu}_n$<br>$\rho^{\mu}_{n+1} = B\big(Q\widehat{\rho}^{\mu}_{n+1}\,; y^{n+1}\big)$ |
| **MFEnKF** | $\widehat{X}^{n+1} = AX^n + \xi^n$<br>$X^{n+1} = T(\widehat{X}^{n+1}; \rho^{EK,\mu}_n, y^{n+1} - \eta^{n+1})$ | $\widehat{\rho}^{EK,\mu}_{n+1} = P\rho^{EK,\mu}_n$<br>$\rho^{EK,\mu}_{n+1} = T^n\big(Q\widehat{\rho}^{EK,\mu}_{n+1}\big)$ |

where

$$T(x;\mu,v) := x - K(\mu)(Hx - v), \qquad K(\mu) := \mathrm{Cov}(\mu)H^T\big(H\,\mathrm{Cov}(\mu)H^T + \Gamma\big)^{-1}$$

## Discrete-to-continuous limit:

$$A = I + \tau\mathsf{F}, \quad H = \tau\mathsf{H}, \quad \Gamma = \tau\mathsf{G} \qquad (\tau \to 0)$$

## Continuous time

| | SDE |
|---|---|
| **True Filter** | $dX_t = \mathsf{F}X_t\,dt + \sqrt{\Sigma}\,dB_t;$<br>$dZ_t = \mathsf{H}X_t\,dt + \sqrt{\mathsf{G}}\,dW_t$ |
| **MFEnKF** | $dX_t = \mathsf{F}X_t\,dt + \sqrt{\Sigma}\,dB_t - \mathrm{Cov}(\rho^{EK,\mu}_t)\mathsf{H}^T\mathsf{G}^{-1}(\mathsf{H}X_t\,dt + \sqrt{\mathsf{G}}\,dW_t - dZ_t)$ |

# The Key Assumptions: Detectability & Controllability

Given $M \in \mathbb{R}^{q\times q}$, we define the spectral radius

$$\rho(M) := \max\{|\lambda| : \lambda \text{ is an eigenvalue of } M\}$$

and spectral abscissa,

$$\alpha(M) := \max\{\mathrm{Re}(\lambda) : \lambda \text{ is an eigenvalue of } M\}.$$

## Definitions

- Stable $M$: A square matrix $M$ is $c$-stable if $\alpha(M) < 0$; we say $M$ is $d$-stable if $\rho(M) < 1$.
- Detectable $(H, A)$: some $K$ makes $A - KH$ stable
  $\rightarrow$ *observations see (and correct) every unstable direction of the dynamics;*
- Controllable $(A, \Sigma^{1/2})$: $A, \Sigma \in \mathbb{R}^{d\times d}$ and
  $$\mathrm{rank}[\Sigma^{1/2}, A\Sigma^{1/2}, A^2\Sigma^{1/2}, \cdots, A^{d-1}\Sigma^{1/2}] = d.$$
  $\rightarrow$ *The noise excites every direction of the state space; automatically holds when* $\Sigma \succ 0$.

- **Classical**:implies asymptotic stability of the Kalman filter **[Wonham 1968]**;
- **New**: implies MFEnKF 'forgets' non-Gaussian initial data.

# Contribution 1: Variational Formulation of MFEnKF

Bayes' rule is variational:

$$\pi_{\rm post} \in \arg\min_{\mu} \mathrm{KL}(\mu \| \pi_{\rm prior}) + \tfrac{1}{2} \int |Hx - y|_\Gamma^2 d\mu \,.$$

- Explicitly solvable for Gaussian $\pi_{\rm prior}$; needs approximation in general, a Langrangian perspective is desirable.
- **Idea**: replace $\mathrm{KL}$ by the covariance-weighted 2-Wasserstein distance $W_2(\cdot, \cdot\,; M)$, $|v|_M^2 := v^\top M^{-1} v$;
- Explicit bounds via Talagrand / HWI inequalities [Otto, Villani 2000]: if $D^2 \widehat{U}(x) \geq M^{-1} \succ 0$, then

$$\frac{1}{2} W_2^2(\mu, \widehat{\pi}; M) \leq \mathrm{KL}((\|\mu)||\widehat{\pi}) \leq \frac{1}{2} W_2(\mu, \widehat{\pi}; M) \left(2\sqrt{I_M(\mu||\widehat{\pi})} - W_2(\mu, \widehat{\pi}; M)\right);$$

# Contribution 1: Variational Formulation of MFEnKF

$\tilde{\pi}_{post}(\varphi) := \mathbf{E}^{\eta}[\tilde{\pi}_{post,\eta}(\varphi)]$, where

$$\tilde{\pi}_{post,\eta} \in \underset{\mu \in \mathcal{P}_2(\mathbb{R}^d)}{\arg\min} \frac{1}{2} W_2^2(\mu, \pi_{prior}; \mathrm{Cov}(\pi_{prior})) + \frac{1}{2} \int_{\mathbb{R}^d} |Hx + \eta - y|_{\Gamma}^2 \, d\mu(u);$$

- If $\pi_{\text{prior}}$ is Gaussian, then $\tilde{\pi}_{post} = \pi_{\text{post}}$.

## Theorem (informal)

For $\pi_{prior} = \widehat{\rho}_{n+1}^{EK,\mu}$, this variational problem has a **unique** minimizer, and coincides **exactly** with the MFEnKF solution $\rho_{n+1}^{EK,\mu}$.

Take-Away: covariance-weighted optimal transport is the natural geometry for EnKF.

# Contribution 2: Contraction to the Gaussian Manifold

Let $G\mu$ = Gaussian projection of $\mu$. Assume detectability and controllability hold.

## Theorem (discrete time, informal)

$$W_2^2\big(\widehat{\rho}_{n+1}^{EK,\mu}, \widehat{\rho}_{n+1}^{EK,G\mu}\,; \mathrm{Cov}(\widehat{\rho}_{n+1}^{EK,\mu})\big) \;\le\; (1-\beta_n)\, W_2^2\big(\widehat{\rho}_{n}^{EK,\mu}, \widehat{\rho}_{n}^{EK,G\mu}\,; \mathrm{Cov}(\widehat{\rho}_{n}^{EK,\mu})\big)$$

with explicit rates $\beta_n \in (0,1)$, uniformly bounded away from $0$ in $n$.

## Theorem (continuous time, informal)

$$W_2\big(\rho_t^{EK,\mu}, \rho_t^{EK,G\mu}\,; \mathrm{Cov}(\rho_t^{EK,\mu})\big) \;\le\; \exp\Big(-\tfrac{1}{2}\beta_t t\Big)\, W_2\big(\mu, G\mu\,; \mathrm{Cov}(\mu)\big)$$

with $\beta_t > 0$ uniformly bounded away from $0$, for **every** sample path of the observations.

- **Why this is hard**: with $\Phi_n := A - \mathrm{Cov}(\widehat{\rho}_n^{EK,\mu})H^T(H\,\mathrm{Cov}(\widehat{\rho}_n^{EK,\mu})H^T + \Gamma)^{-1}H$, detectability & controllability only give $\Phi_n \to \Phi_\infty$ with spectral radius $< 1$ — **not** $\|\Phi_\infty\|_2 < 1$, so naive Euclidean contraction can fail transiently;
- **Key tool**: $\widehat{X}^n \mapsto \Phi_n^T\,\mathrm{Cov}(\widehat{X}^n)\Phi_n$ is a Lyapunov function in the Loewner order.

# Contribution 2: Proof via Matrix Riccati Equations

**Key observation:**

- **Discrete Time:** The covariance matrices $(\widehat{C}^n)_{n\geq 1}$ with $\widehat{C}^n := \mathrm{Cov}(\widehat{X}^n)$ satisfy

$$\widehat{C}^{n+1} = \mathrm{Ricc}_d(\widehat{C}^n),$$

  where the discrete Riccati operator $\mathrm{Ricc}_d : \mathbb{R}_+^{d\times d} \to \mathbb{R}_+^{d\times d}$ is defined by

$$\mathrm{Ricc}_d(C) := ACA^T + \Sigma - ACH^T(HCH^T + \Gamma)^{-1}HCA^T.$$

- **Continuous Time:** Covariance matrix $C_t := \mathrm{Cov}(X_t)$ satisfies the Riccati differential equation

$$\dot{C}_t = \mathrm{Ricc}_c(C_t),$$

  where the continuous Riccati operator $\mathrm{Ricc}_c : \mathbb{R}_+^{d\times d} \to \mathbb{R}_+^{d\times d}$ is defined by

$$\mathrm{Ricc}_c(C) := C\mathsf{F}^T + \mathsf{F}C + \mathbf{\Sigma} - C\mathsf{H}^T\mathsf{G}^{-1}\mathsf{H}C.$$

**Remarks:**

- Satisfied by MFEnKF regardless of whether the initial data is Gaussian [Del Moral, Tugaut 2018]; this is **not** true for the optimal filter.
- In both cases, the covariance evolution is deterministic and independent of the observations.

# Contribution 3: Stability & Accuracy of the MFEnKF

- On the Gaussian manifold: MFEnKF = Kalman filter = optimal filter.

## Theorem (asymptotic stability, continuous time)

Assume $(H, F)$ is a detectable pair, and $\Sigma, G \succ 0$. For any $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ with nondegenerate covariance,

$$\limsup_{t\to\infty} \frac{1}{t} \log W_2(\rho_t^{EK,\mu}, \rho_t^{EK,\nu}) < 0.$$

## Theorem (accuracy, continuous time)

Under the same assumptions, given any bounded and uniformly continuous $\varphi : \mathbb{R}^d \to \mathbb{R}$ or $\varphi = \mathrm{id}$, for almost every sample path

$$\rho_t^{EK,\mu}(\varphi) - \rho_t^{\nu}(\varphi) \to 0 \text{ as } t \to \infty,$$

(i.e. MFEnKF and the optimal filter agree in the limit).

- Proof: combine (C2) with classical stability of the optimal filter **[Ocone, Pardoux 1996]**.
- **MFEnKF is asymptotically exact** — even though any single step may be a poor Bayes approximation.

# Summary & Open Questions: Large-Time Behavior

In Part 3,

- **variational** derivation of the MFEnKF via covariance-weighted OT;
- **exponential contraction** to the Gaussian manifold (discrete & continuous time);
- **stability** and **accuracy** of the MFEnKF, with no near-Gaussian assumption.

**Perspectives for future work**:

- **Finite-$J$**: combine with Part 2 for uniform-in-time, finite-particle guarantees;
- stronger **optimal filter** stability results;
- beyond linear observations $h(x) = Hx$;
- large-time behavior of the optimal filter in the **nonlinear** setting.

| | **Setting** | **Result** |
|---|---|---|
| **Part 1** | EnKF, near-Gaussian, $J = \infty$ | finite-time, local accuracy |
| **Part 2** | TEFs, $J \to \infty$ | quantitative PoC |
| **Part 3** | EnKF, $J = \infty$ | all-time, global accuracy |

# References (1/3)

**Parts 1 & 2: EnKF, Propagation of Chaos**


- G. Evensen. Ocean Dynamics, 2003
- J. Mandel, L. Cobb, and J. D. Beezley. Applications of Mathematics, 2011
- F. Le Gland, V. Monbet, and V.-D. Tran. In The Oxford handbook of nonlinear filtering. Oxford Univ. Press, Oxford, 2011
- E. Kwiatkowski and J. Mandel. SIAM/ASA Journal on Uncertainty Quantification, 2015
- E. Calvello, P. Monmarché, A. M. Stuart, and U. Vaes. SIAM Journal on Numerical Analysis, 2026
- O. Al-Ghattas and D. Sanz-Alonso. Information and Inference: A Journal of the IMA, 2024
- N. Fournier and A. Guillin. Probability Theory and Related Fields, 2015
- A. Spantini, R. Baptista, and Y. Marzouk. SIAM Review, 2022
- A.-S. Sznitman. In P.-L. Hennequin, editor, École d'Été de Probabilités de Saint-Flour XIX—1989, Lecture Notes in Mathematics. Springer, Berlin, 1991

# References (2/3)

**Parts 1 & 2: EnKF, Propagation of Chaos (cont.)**


- R. Vershynin. Cambridge University Press, 2018
- M. Kac. In J. Neyman, editor, Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, Berkeley and Los Angeles. University of California Press, 1956
- H. P. McKean Jr. In Stochastic Differential Equations (Lecture Series in Differential Equations, Session 7, Catholic University, 1967). Air Force Office of Scientific Research, Arlington, VA, 1967
- T.-V. Hoang, S. Krumscheid, H. G. Matthies, and R. Tempone. Foundations of Data Science, 2023
- F. J. N. Jorgensen and Y. M. Marzouk. 2025. arXiv preprint arXiv:2510.00158
- E. Bach, R. Baptista, E. Calvello, B. Chen, and A. M. Stuart. Journal of Computational Physics, 2026
- D. T. Kelly, K. J. Law, and A. M. Stuart. Nonlinearity, 2014
- X. T. Tong, A. J. Majda, and D. Kelly. Communications in Mathematical Sciences, 2016

# References (3/3)

**Part 3: Large-Time Behavior**


- R. Kalman. Journal of Basic Engineering, 1960
- R. Kalman and R. Bucy. Journal of Basic Engineering, 1961
- W. M. Wonham. SIAM Journal on Control, 1968
- D. Ocone and E. Pardoux. SIAM Journal on Control and Optimization, 1996
- A. Garbuno-Inigo, F. Hoffmann, W. Li, and A. M. Stuart. SIAM Journal on Applied Dynamical Systems, 2020
- E. Calvello, S. Reich, and A. M. Stuart. Acta Numerica, 2025
- M. Burger, M. Erbar, F. Hoffmann, D. Matthes, and A. Schlichting. Archive for Rational Mechanics and Analysis, 2025
- F. Otto and C. Villani. Journal of Functional Analysis, 2000