# Kalman Filters: from Data Assimilation to Inverse Problems

Mark Asch [1]

[1]LAMFA - Université de Picardie Jules Verne

Journéé du LAMFA, Amiens, 28/06/2024

# Outline

# Outline of the Talk

- Direct vs. Inverse Problems
- Data Assimilation
- Kalman Filters: linear, nonlinear, ensemble.
- Ensemble Kalman Inversion

# Table of contents

# Introduction

# Direct and Inverse Problems

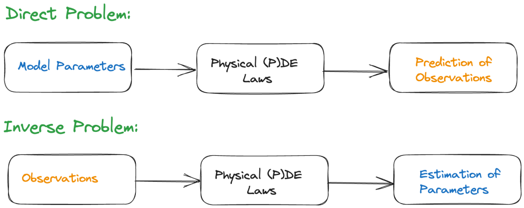
## Dynamical System

$$\frac{\mathrm{d}\mathbf{u}}{\mathrm{d}t} = g(t, \mathbf{u}; \boldsymbol{\theta}), \qquad \mathbf{u}(t_0) = \mathbf{u}_0,$$

with $g$ known, $\boldsymbol{\theta} \in \Theta$, $\mathbf{u}(t) \in \mathbb{R}^k$.

Direct: Given $\boldsymbol{\theta}$, $\mathbf{u}_0$, find $\mathbf{u}(t)$ for $t \geq t_0$.

Inverse: Given $\mathbf{u}(t)$ for $t \geq t_0$, find $\boldsymbol{\theta} \in \Theta$.



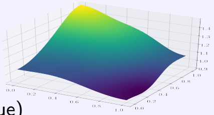
# Example: inference of the thermal conductivity in a plate

**Mathematical model:**

$$\begin{aligned} -\nabla \cdot \big(u(x)\nabla T(x)\big) &= f(x), & x &\in \Omega, \\ T(x) &= 0, & x &\in \partial\Omega. \end{aligned}$$

**Unknown parameter:**

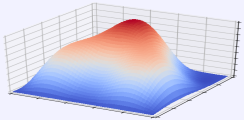
Thermal conductivity $u(x)$



(true)

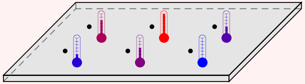
Forward problem

**Solution:**



Temperature field $T(x)$

**Data:**

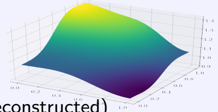


Noisy temperature measurements:

$$y = \big(T(x_1), \ldots, T(x_m)\big) + \eta.$$

Inverse problem

**MAP estimator:**



(reconstructed)

# Deterministic and Stochastic Problems

In reality, we have uncertainty (noise)

- in the model,
- in the parameters,
- in the observations.

The dynamical system becomes

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

where $\eta \sim \mathcal{N}(0, \Sigma)$.

# Inverse Problems - Deterministic Case

In the deterministic case ($\eta = 0$), because of the ill-posedness of the inverse problem, we replace it by the least-squares optimization problem,

$$\operatorname*{argmin}_{u \in X} \frac{1}{2} \|y - \mathcal{G}(u)\|_Y^2$$

that is usually regularized as

$$\operatorname*{argmin}_{u \in E} \frac{1}{2} \left( \|y - \mathcal{G}(u)\|_Y^2 + \frac{1}{2} \|u - m_0\|_E^2 \right)$$

for a given reference point $m_0 \in E$, with $E$, $X$, $Y$ Banach spaces. The optimization requires a gradient (or adjoint).

# Inverse Problems - Stochastic Case

In the stochastic case, the solution of the inverse problem is a posterior probability density function (ppdf).

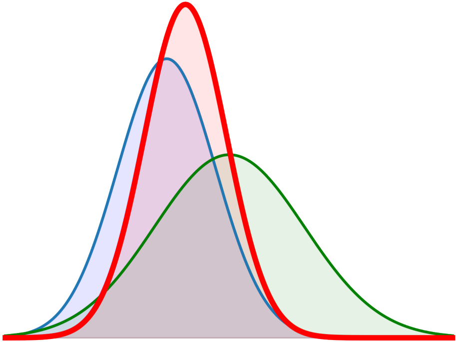
**Theorem (Bayes)**

$$p(u|y) = \frac{p(y|u)p(u)}{p(y)},$$

*or*

$$p(\text{parameter}|\text{data}) \propto p(\text{data}|\text{parameter})p(\text{parameter}).$$

# Bayes' Theorem



Bayes posterior for Gaussian prior and likelihood

# Data Assimilation - Definition

## Definition (DA)

Data assimilation concerns the estimation of the state of a dynamical system by optimally combining observed data with the underlying mathematical model.

*[Stuart] The problem of effectively combining data with a mathematical model constitutes a major challenge in applied mathematics. It is particular challenging for high-dimensional dynamical systems where data is received sequentially in time and the objective is to estimate the system state in an on-line fashion; this situation arises, for example, in weather forecasting. The sequential particle filter is then impractical and ad hoc filters, which employ some form of Gaussian approximation, are widely used. Prototypical of these ad hoc filters is the 3DVAR method.*

# Data Assimilation - Filters

- The 3DVAR method is the simplest Gaussian filter, relying on fixed (with respect to the data time-index increment) forecast and analysis model covariances, related through a Kalman update.

- A more sophisticated idea is to update the forecast covariance via the linearized dynamics, again computing the analysis covariance via a Kalman update, leading to the extended Kalman filter.

- In high dimensions computing the full linearized dynamics is not practical. For this reason the ensemble Kalman filter is widely used, in which the forecast covariance is estimated from an ensemble of particles, and each particle is updated in Kalman fashion.

# Kalman Filter - Navigation

## State Space Model

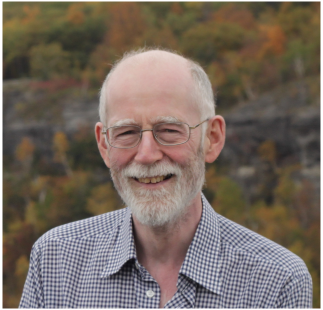
$$\begin{aligned}
\text{Dynamics:} \quad & v_{n+1} = M v_n + \xi_n, \quad n \in \mathbb{Z}^+, \\
\text{Data:} \quad & y_{n+1} = H v_{n+1} + \eta_{n+1}, \quad n \in \mathbb{Z}^+ \\
\text{Probability:} \quad & v_0 \sim \mathcal{N}(m_0, C_0), \quad \xi_n \sim \mathcal{N}(0, \Sigma), \quad \eta_n \sim \mathcal{N}(0, \Gamma) \\
\text{Probability:} \quad & v_0 \perp \{\xi_n\} \perp \{\eta_n\} \quad \text{independent}
\end{aligned}$$

- Rudolf Kálmán (1960), Apollo 11
- $\approx$ 44 000 citations (G-Scholar, 05/2024)
- Algorithm:
  - $v_n | Y_n \sim \mathcal{N}(m_n, C_n)$,
  - $(m_n, C_n) \mapsto (m_{n+1}, C_{n+1})$

# Kalman Filter - 3DVAR

## State Space Model

$$
\begin{aligned}
\text{Dynamics:} \quad & v_{n+1} = \Psi(v_n) + \xi_n, \quad n \in \mathbb{Z}^+, \\
\text{Data:} \quad & y_{n+1} = H v_{n+1} + \eta_{n+1}, \quad n \in \mathbb{Z}^+ \\
\text{Probability:} \quad & v_0 \sim \mathcal{N}(m_0, C_0), \quad \xi_n \sim \mathcal{N}(0, \Sigma), \quad \eta_n \sim \mathcal{N}(0, \Gamma) \\
\text{Probability:} \quad & v_0 \perp \{\xi_n\} \perp \{\eta_n\} \quad \text{independent}
\end{aligned}
$$



- Andrew Lorenc (1986)
- $\approx$ 2 000 citations (G-Scholar, 05/2024)
- Convergence: Stuart, et al (2012)
- Algorithm:
  - $v_n \mapsto v_{n+1}$,
  - $C$ fixed.

# Kalman Filter - Ensemble

## State Space Model

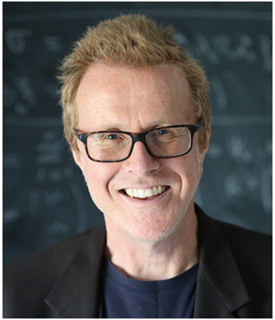
$$\begin{aligned} \text{Dynamics:} \quad & v_{n+1} = \Psi(v_n) + \xi_n, \quad n \in \mathbb{Z}^+, \\ \text{Data:} \quad & y_{n+1} = H v_{n+1} + \eta_{n+1}, \quad n \in \mathbb{Z}^+ \\ \text{Probability:} \quad & v_0 \sim \mathcal{N}(m_0, C_0), \quad \xi_n \sim \mathcal{N}(0, \Sigma), \quad \eta_n \sim \mathcal{N}(0, \Gamma) \\ \text{Probability:} \quad & v_0 \perp \{\xi_n\} \perp \{\eta_n\} \quad \text{independent} \end{aligned}$$

- Geir Evensen (1994)
- $\approx$ 15 000 citations (G-Scholar, 05/2024)
- Algorithm:
  - $X = \{x_e\}_{e=1}^{N_e}$, $\bar{x} = (1/N_e)\sum_{e=1}^{N_e} x_e$,
    $C_e = (1/(N_e - 1))\sum_{e=1}^{N_e}(x_e - \bar{x})(x_e - \bar{x})^{\mathrm{T}}$
  - $(m_n, C_n) \mapsto (m_{n+1}, C_{n+1})$

# Kalman Filter - Mean Field

## State Space Model

$$\begin{aligned} \text{Dynamics:} \quad & v_{n+1} = \Psi(v_n) + \xi_n, \quad n \in \mathbb{Z}^+, \\ \text{Data:} \quad & y_{n+1} = H(v_{n+1}) + \eta_{n+1}, \quad n \in \mathbb{Z}^+ \\ \text{Probability:} \quad & v_0 \sim \mathcal{N}(m_0, C_0), \quad \xi_n \sim \mathcal{N}(0, \Sigma), \quad \eta_n \sim \mathcal{N}(0, \Gamma) \\ \text{Probability:} \quad & v_0 \perp \{\xi_n\} \perp \{\eta_n\} \quad \text{independent} \end{aligned}$$



- Andrew Stuart (2022–)
- Convergence: Acta Numerica (2025)
- Algorithm:
  - $X = \{x_e\}_{e=1}^{N_e}$, $\bar{x} = (1/N_e)\sum_{e=1}^{N_e} x_e$,
    $C_e = (1/(N_e - 1))\sum_{e=1}^{N_e}(x_e - \bar{x})(x_e - \bar{x})^{\mathrm{T}}$
  - $(v_n, \mu_n) \mapsto (v_{n+1}, \mu_{n+1})$, $\mu_n \doteq \mathrm{Law}(v_n)$.

# Kalman Filters - Classical State Space

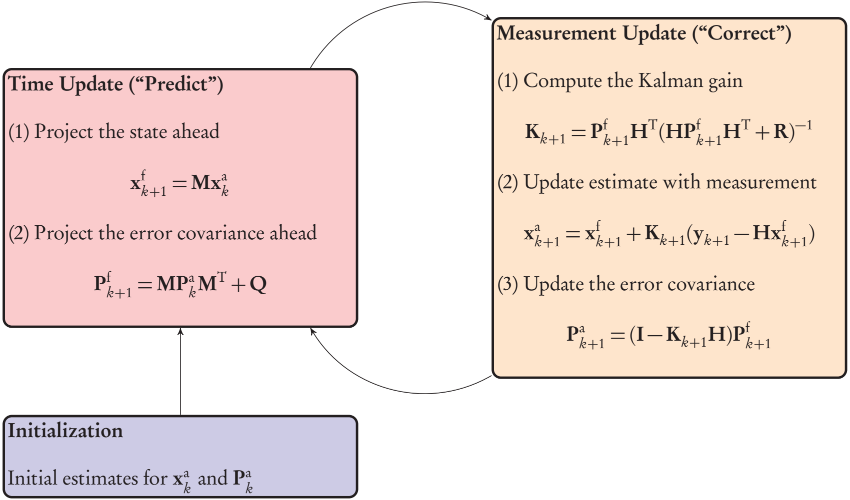
$$x_{k+1} = Mx_k + w_k, \tag{1}$$
$$y_{k+1} = Hx_k + v_k, \tag{2}$$

where[1]

- $M$ is the dynamics ,
- $H$ the observation operator,
- $w_k \sim \mathcal{N}(0, Q)$ the process noise,
- $v_k \sim \mathcal{N}(0, R)$ the observation noise.

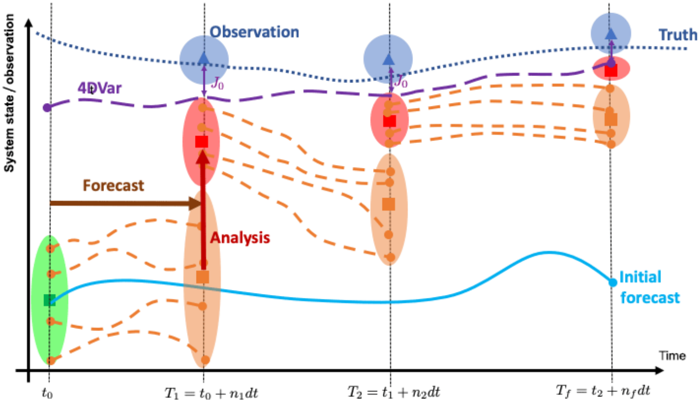
[1]Notation: classical state space uses $x$, $y$, $M$ and $H$, whereas probabilistically we will use $u$ or $v$, $y$, $\Psi$ or $\mathcal{G}$, and $H$. For covariance matrices $P$ or $C$, respectively.

# Kalman Filters - Flowchart



# Ensemble Kalman Filters

$$x_{k+1} = M(x_k) + w_k, \quad y_{k+1} = H(x_k) + v_k, \quad x = \{x_e\},\ e = 1, \ldots, N_e$$




[Credit:F. Gillet-Chaulet, 2020]

# Ensemble Kalman Filters - Algorithm

## Predict

$$\hat{v}^n_{k+1} = \Psi(v^n_k) + \xi^n_k,\;\; \hat{m}_{k+1} = \frac{1}{N_e}\sum_{i=1}^{N} \hat{v}^n_{k+1}, ; n = 1, \dots, N_e, \tag{3}$$

$$\hat{C}_{k+1} = \frac{1}{N_e - 1}\sum_{i=1}^{N_e} \left(\hat{v}^n_{k+1} - \hat{m}_{k+1}\right)\left(\hat{v}^n_{k+1} - \hat{m}_{k+1}\right)^{\mathrm{T}} \tag{4}$$

## Correct

$$K_{k+1} = \hat{C}_{k+1} H^{\mathrm{T}} S^{-1}_{k+1},\;\; S_{k+1} = H\hat{C}_{k+1}H^{\mathrm{T}} + \Gamma, \tag{5}$$

$$y^n_{k+1} = y_{k+1} + \eta^n_{k+1},\;\; n = 1, \dots, N_e, \tag{6}$$

$$v^n_{k+1} = \hat{v}^n_{k+1} + K_{k+1}d,\; d = y^n_{k+1} - H\hat{v}^n_{k+1},\;\; n = 1, \dots, N_e. \tag{7}$$

# Ensemble Kalman Filters - Algorithm

**Algorithm:** Ensemble KF

**Input** : $N_e$, $y$, process and noise covariances
**Output:** $v$, $C$
Choose $\{v_0^e\}_{e=1}^{N_e}$, $j = 0$ ;
**for** $k \leftarrow 1$ **to** $K$ **do**
    Predict $\{\hat{v}_{k+1}^e\}_{e=1}^{N_e}$, $\hat{C}_{k+1}$ from (3-4) ;
    Update $\{v_{k+1}^e\}_{e=1}^{N_e}$ from (5-7)

There are many ($> 20$) EnKF variants—stochastic and deterministic[2].

[2] S. Vetra-Carvalho, P.J. van Leeuwen, L. Nerger, A. Barth, M. Umer Altaf, P. Brasseur. "State-of-the-art stochastic data assimilation methods for high-dimensional non-Gaussian problems". Tellus A, 70, (2018)

# Ensemble Kalman Filters - Properties

- EnKF represents error statistics by ensembles of (nonlinear) model and (nonlinear) measurement realizations.
- EnKF performs sequential DA that processes measurements recursively in time.
- EnKF is suitable for weather-prediction or any other complex, chaotic dynamic systems.
- Error propagation is nonlinear.
- Filter update is linear and computed in the low rank, ensemble subspace.
- EnKF does not require any gradients, adjoints, linearizations.
- EnKF computations are embarassingly parallel.

# Example - EnKF for Lorenz63 System I

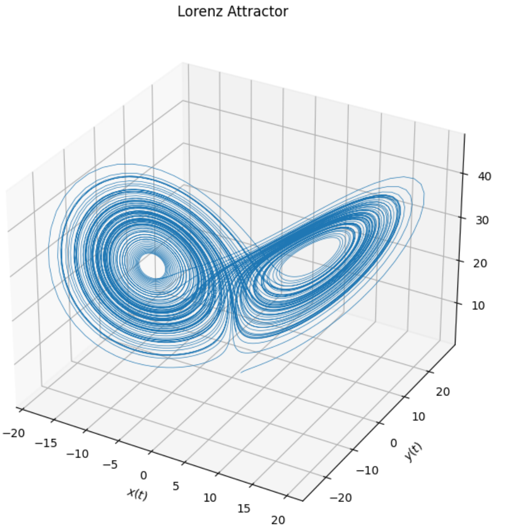
We apply the EnKF to the Lorenz systems of ordinary differential equations. These systems exhibit chaotic behavior and as such are considered as excellent toy models for complex phenomena, in particular for simulation of weather. The Lorenz-63 system is given by

$$\begin{aligned} \frac{\mathrm{d}x}{\mathrm{d}t} &= -\sigma(x-y), \\ \frac{\mathrm{d}y}{\mathrm{d}t} &= \rho x - y - xz, \\ \frac{\mathrm{d}z}{\mathrm{d}t} &= xy - \beta z, \end{aligned} \tag{8}$$

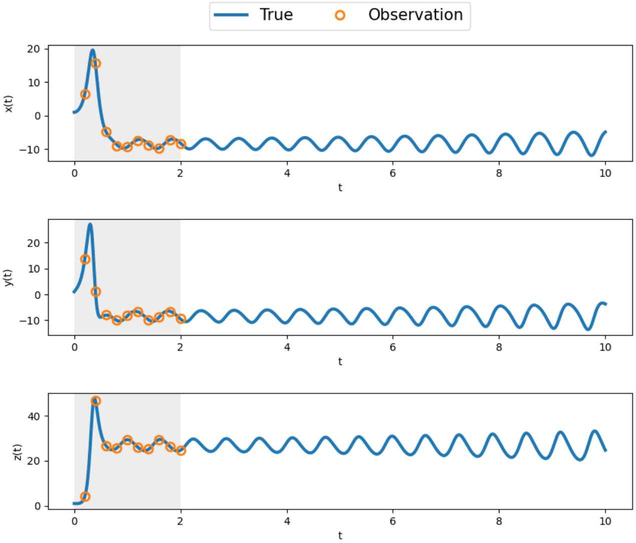
where $x = x(t)$, $y = y(t)$, $z = z(t)$ and $\sigma$ (ratio of kinematic viscosity divided by thermal diffusivity), $\rho$ (measure of stability) and $\beta$ (related to the wave number) are parameters.

# Example - EnKF for Lorenz63 System II

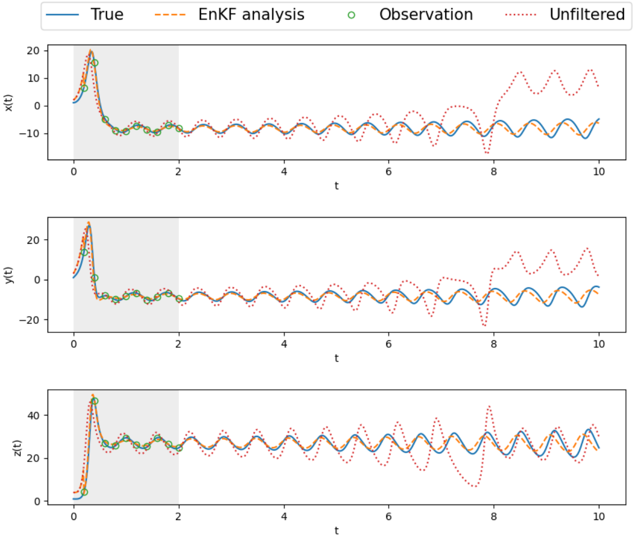
Chaotic behavior is obtained for
$\sigma = 10, \quad \rho = 28, \quad \beta = 8/3.$
The solution is very sensitive to the parameters and the initial conditions and a small difference in these values can lead to a very different solution. This equation is an excellent example of the lack of predictability. The solution switches between two stable orbits, around the points
$\left(\sqrt{\beta(\rho-1)}, \sqrt{\beta(\rho-1)}, \rho-1\right)$, and
$\left(-\sqrt{\beta(\rho-1)}, -\sqrt{\beta(\rho-1)}, \rho-1\right)$.



# Example - EnKF for Lorenz63 III



# Example - EnKF for Lorenz63 III



# Theory

# Ensemble Kalman Inversion - Setting I

- In 1960 R. Kalman[3] published the first paper to develop a systematic, principled approach to the use of data to improve the predictive capability of dynamical systems. As our ability to gather data grows at an enormous rate, the importance of this work continues to grow too. The paper is confined to linear dynamical systems subject to Gaussian noise.
- The work of Geir Evensen[4] in 1994 opened up far wider applicability of Kalman's ideas by introducing the ensemble Kalman filter. The EnKF applies to the setting in which nonlinear and noisy observations are used to make improved predictions of the state of a Markov chain. The algorithm results in an interacting particle system combining elements of the Markov chain and the observation process.

[3] R. Kalman, A new approach to linear filtering and prediction problems. J. Basic Eng., 82:35–45, 1960.

[4] G. Evensen, Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. J. Geophys. Research, 99(C5), 1994.

# Ensemble Kalman Inversion - Setting II

- A unifying mean-field perspective on the algorithm was very recently derived by Stuart et al[5] in the limit of an infinite number of interacting particles. This methodology can be used to study inverse problems, opening up diverse applications beyond prediction in dynamical systems.
- Analysis of the methodology, both in terms of accuracy and uncertainty quantification has been developed.
- Despite its widespread adoption in applications, a complete mathematical theory is lacking and there are many opportunities for analysis and applications in this area.

[5] Calvello, Edoardo, Sebastian Reich, and Andrew M. Stuart. 2025. "Ensemble Kalman Methods: A Mean Field Perspective." Acta Numerica. http://arxiv.org/abs/2209.11371.

# Iterative EKI

Conceptually, the idea of EKI is very simple. Just apply the standard EnKF to an augmented system of state plus parameters, for a given number of iterations, to invert for $\theta$. We begin by pairing the parameter-to-data map with a dynamical system for the parameter, and then employ techniques from filtering to estimate the parameter given the data.

## IDEA

Introduce a pseudo-time, and write down the augmented system

$$\begin{aligned} \theta_{n+1} &= \theta_n \\ y_{n+1} &= \mathcal{G}(\theta_{n+1}) + \eta_{n+1}, \end{aligned}$$

where the operator $\mathcal{G}$ contains both the system dynamics and the observation function.

# Mean-Field EKI

Consider the stochastic dynamical system

$$\begin{aligned} &\text{evolution:} && \theta_{n+1} = \theta_n + \omega_{n+1}, && \omega_{n+1} \sim \mathcal{N}(0, \Sigma_\omega), && (9)\\ &\text{observation:} && x_{n+1} = \mathcal{F}(\theta_{n+1}) + \nu_{n+1}, && \nu_{n+1} \sim \mathcal{N}(0, \Sigma_\nu). && (10) \end{aligned}$$

where

$$\mathcal{F}(\theta) = \begin{bmatrix} \mathcal{G}(\theta) \\ \theta \end{bmatrix}.$$

We seek the best Gaussian approximation of the posterior distribution of $\theta$ for ill-posed inverse problems, where the prior is a Gaussian, $\theta_0 \sim \mathcal{N}(r_0, \Sigma_0)$.

# Convergence

**Theorem (EKI Linear)**

*For a linear dynamic model, assume that the prior covariance matrix $\Sigma_0 \succ 0$ and initial covariance matrix $C_0 \succ 0$. Then iteration for the conditional mean $m_n$ and covariance matrix $C_n$ characterizing the distribution of $\theta_n | Y_n$ converges exponentially fast to the posterior mean, $m_{\text{post}}$, and covariance, $C_{\text{post}}$.*

**Theorem (EKI Near-Gaussian)**

*If the measure $\mu_n$ is $\epsilon$-close to Gaussian, then*

$$\sup_{0 \leq n \leq N_e} d_g(\mu_n, \mu_n^{\text{EK}}) \leq C\epsilon$$

*for bounded $\Psi, H$, where $d_g$ is a wieghted TV metric over measures $\mu$.*

# Conclusions and Perspectives

# Conclusions

- Kalman-based inversion can be widely used to construct derivative-free optimization and sampling methods for nonlinear inverse problems.
- Kalman-based inversion methods for Bayesian inference and UQ build on the work in both optimization and sampling.
- They propose a new method for Bayesian inference based on filtering a novel mean-field dynamical system subject to partial noisy observations, and which depends on the law of its own filtering distribution, together with application of the Kalman methodology.
- Theoretical guarantees are presented:
  - for linear inverse problems, the mean and covariance obtained by the method converge exponentially fast to the posterior mean and covariance;
  - for nonlinear inverse problems, numerical studies indicate the method delivers an excellent approximation of the posterior distribution for problems which are not too far from Gaussian. (proven 2024)

# Conclusions II

In terms of performance:

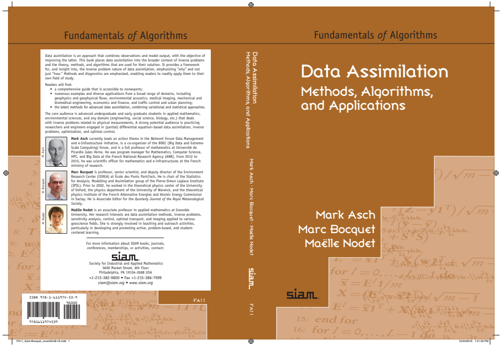
1. The methods are shown to be superior to existing coupling/transport methods, collectively known as iterative Kalman methods.
2. Deterministic, such as ETKF, rather than stochastic implementations of Kalman methodology are found to be favorable.

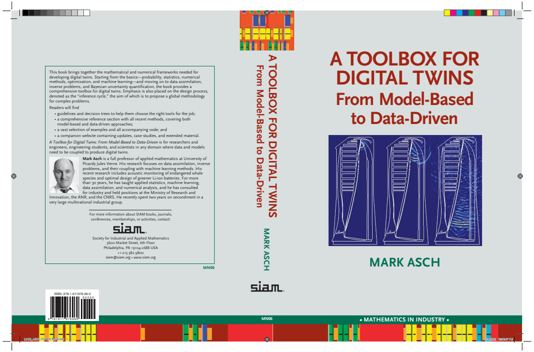
# Perspectives

- There are many open, interesting research questions both in theoretical aspects and for implementations.
- In the framework of the NumPeX[6] PEPR, INRIA-MAKUTU team proposes a PhD[7] on EKI for large-scale problems in wave propagation.
- High performance computing will be performed using the MELISSA framework of the INRIA-DATAMOVE[8] team.

[6] https://numpex.org/
[7] https://jobs.inria.fr/public/classic/fr/offres/2024-07451
[8] https://www.inria.fr/en/datamove

# 2 Books




Source and Codes: https://markasch.github.io/DT-tbx-v1/, https://github.com/markasch/DT-tbx-examples/, https://github.com/markasch/kfBIPq

# References - general


1. K. Law, A. Stuart, K. Zygalakis. *Data Assimilation. A Mathematical Introduction*. Springer, 2015.
2. G. Evensen. *Data assimilation, The Ensemble Kalman Filter*, 2nd ed., Springer, 2009.
3. A. Tarantola. *Inverse problem theory and methods for model parameter estimation.* SIAM. 2005.

# References - Inversion

1. Calvello, Edoardo, Sebastian Reich, and Andrew M. Stuart. 2022. "Ensemble Kalman Methods: A Mean Field Perspective." arXiv (to appear in Acta Numerica 2025). http://arxiv.org/abs/2209.11371.
2. Carrillo, J. A., F. Hoffmann, A. M. Stuart, and U. Vaes. 2024a. "Statistical Accuracy of Approximate Filtering Methods." https://arxiv.org/abs/2402.01593 .
3. ———. 2024b. "The Mean Field Ensemble Kalman Filter: Near-Gaussian Setting." https://arxiv.org/abs/2212.13239 .
4. Dashti, Masoumeh, and Andrew M. Stuart. 2015. "The Bayesian Approach to Inverse Problems." In Handbook of Uncertainty Quantification, edited by Roger Ghanem, David Higdon, and Houman Owhadi, 1–118. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-11259-6_7-1 .
5. Huang, Daniel Zhengyu, Jiaoyang Huang, Sebastian Reich, and Andrew M Stuart. 2022. "Efficient Derivative-Free Bayesian Inference for Large-Scale Inverse Problems." Inverse Problems 38 (12): 125006.
6. Iglesias, Marco A, Kody J H Law, and Andrew M Stuart. 2013. "Ensemble Kalman Methods for Inverse Problems." Inverse Problems 29 (4): 045001.