Preserving linear invariants in ensemble filtering methods

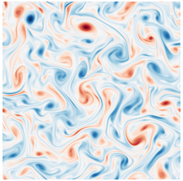
Mathieu Le Provost
Department of Aeronautics and Astronautics
Massachusetts Institute of Technology

Joint work with Jan Glaubitz & Youssef Marzouk (MIT)

Applied and Computational Mathematics Seminar
Dartmouth College
April 23, 2024

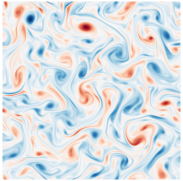
Motivation: Sequential state estimation

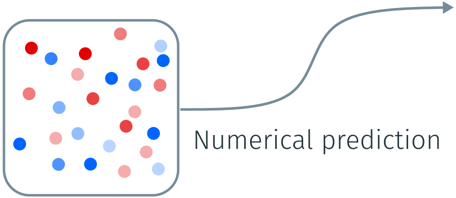
Physical system



Motivation: Sequential state estimation

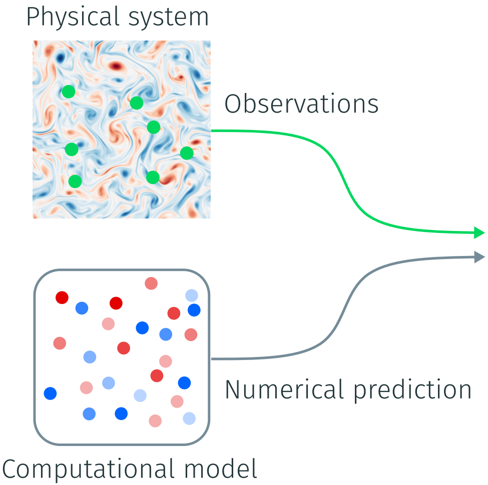
Physical system



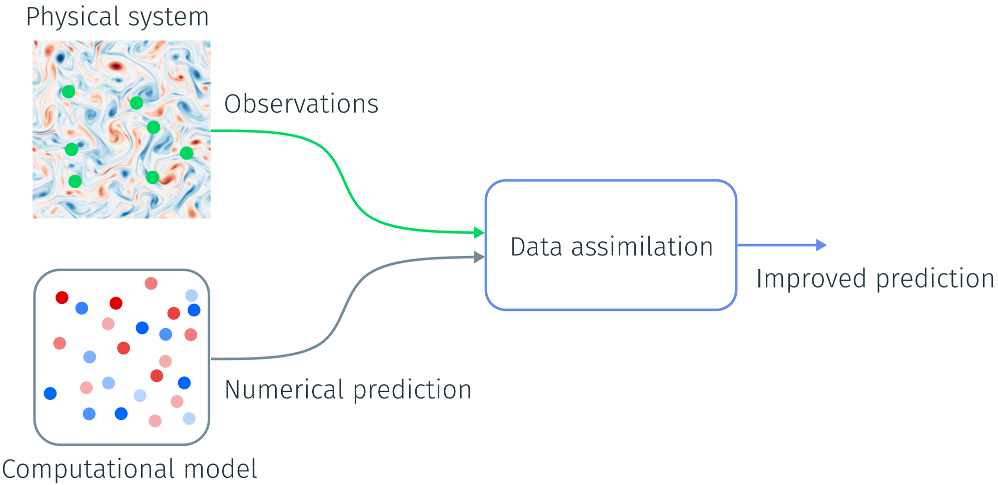


Computational model

Motivation: Sequential state estimation



Motivation: Sequential state estimation



Filtering problem

State-space model:

$$\begin{aligned}
&\text{Dynamical model – transition kernel:} && x_{t+1} \sim \pi_{\mathsf{X}_{t+1} \mid \mathsf{X}_t}(\cdot \mid x_t) \\
&\text{Observation model – likelihood model:} && y_t \sim \pi_{\mathsf{Y}_t \mid \mathsf{X}_t}(\cdot \mid x_t)
\end{aligned}$$

Filtering problem

State-space model:

Dynamical model – transition kernel: $x_{t+1} \sim \pi_{\mathsf{X}_{t+1} \mid \mathsf{X}_t}(\cdot \mid x_t)$

Observation model – likelihood model: $y_t \sim \pi_{\mathsf{Y}_t \mid \mathsf{X}_t}(\cdot \mid x_t)$

Goal: Characterize the filtering distribution

Sequentially estimate the state X_t given the observations $y_1^\star, y_2^\star, \ldots, y_t^\star$. i.e. **the filtering density** $\pi_{t \mid t} = \pi_{\mathsf{X}_t \mid \mathsf{Y}_{1:t} := y_{1:t}^\star}$

Filtering problem

State-space model:

Dynamical model – transition kernel: $x_{t+1} \sim \pi_{X_{t+1} \mid X_t}(\cdot \mid x_t)$

Observation model – likelihood model: $y_t \sim \pi_{Y_t \mid X_t}(\cdot \mid x_t)$

Goal: Characterize the filtering distribution

Sequentially estimate the state X_t given the observations $y_1^\star, y_2^\star, \ldots, y_t^\star$. i.e. **the filtering density** $\pi_{t \mid t} = \pi_{X_t \mid Y_{1:t} := y_{1:t}^\star}$

Challenges:

- Nonlinear state-space model $\rightarrow$ non-Gaussian transition kernel and likelihood model
- High-dimensions
- Sparsity in space/time

Generic ensemble filtering algorithm

Ensemble filters approximate $\pi_{t \mid t}$ by updating a set of M state realizations $\{\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(M)}\}$.

At each assimilation cycle, they apply

1. Forecast step: Filtering dist. at time $t-1$, $\pi_{t-1 \mid t-1} \rightarrow$ Forecast dist. $\pi_{t \mid t-1}$
   Samples are propagated through the dynamical model.
   We obtain samples $\{\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(M)}\} \sim \pi_{t \mid t-1}$

2. Analysis step: Forecast dist. $\pi_{t \mid t-1} \rightarrow$ Filtering dist. $\pi_{t \mid t}$
   Update the forecast samples with the new observation $\boldsymbol{y}_t^\star$.
   We obtain samples $\{\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(M)}\} \sim \pi_{t \mid t}$

Generic ensemble filtering algorithm

Ensemble filters approximate $\pi_{t \mid t}$ by updating a set of M state realizations $\{\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(M)}\}$.

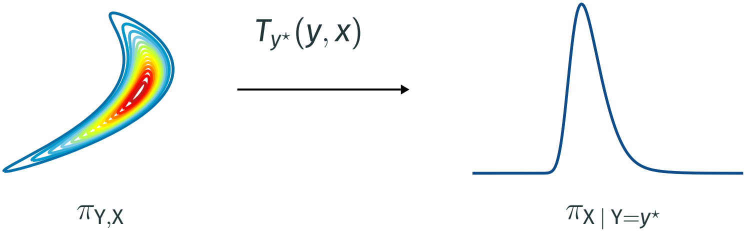
At each assimilation cycle, they apply

1. Forecast step: Filtering dist. at time $t-1$, $\pi_{t-1 \mid t-1} \rightarrow$ Forecast dist. $\pi_{t \mid t-1}$
   Samples are propagated through the dynamical model.
   We obtain samples $\{\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(M)}\} \sim \pi_{t \mid t-1}$

2. Analysis step: Forecast dist. $\pi_{t \mid t-1} \rightarrow$ Filtering dist. $\pi_{t \mid t}$
   Update the forecast samples with the new observation $\boldsymbol{y}_t^\star$.
   We obtain samples $\{\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(M)}\} \sim \pi_{t \mid t}$

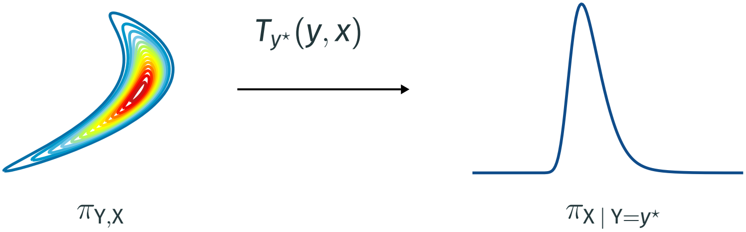
$\rightarrow$ **This talk will focus on the analysis step.**

A “transformative” view of the analysis step



Analysis step: **Analysis map $T_{y^\star}$** that maps $\pi_{(\mathsf{Y}_t,\mathsf{X}_t) \mid \mathsf{Y}_{1:t-1}=y^\star_{1:t-1}}$ to $\pi_{\mathsf{X}_t \mid \mathsf{Y}_{1:t}=y^\star_{1:t}}$

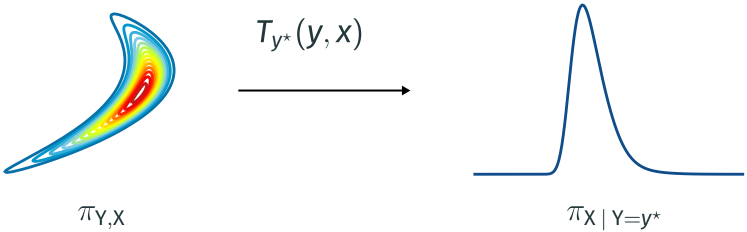
A “transformative” view of the analysis step



Analysis step: **Analysis map** $T_{y^\star}$ that maps $\pi_{(Y_t, X_t) \mid Y_{1:t-1} = y^\star_{1:t-1}}$ to $\pi_{X_t \mid Y_{1:t} = y^\star_{1:t}}$

The analysis map of the Kalman filter $T_{y^\star, \mathrm{KF}}$:

$$T_{y^\star, \mathrm{KF}}(y, x) = x - \boldsymbol{\Sigma}_{X_t, Y_t} \boldsymbol{\Sigma}_{Y_t}^{-1}(y - y^\star) = x - K_t(y - y^\star)$$



Analysis step: **Analysis map** $T_{y^\star}$ that maps $\pi_{(\mathsf{Y}_t,\mathsf{X}_t) \mid \mathsf{Y}_{1:t-1}=y^\star_{1:t-1}}$ to $\pi_{\mathsf{X}_t \mid \mathsf{Y}_{1:t}=y^\star_{1:t}}$

The analysis map of the Kalman filter $T_{y^\star,\text{KF}}$:

$$T_{y^\star,\text{KF}}(y, x) = x - \boldsymbol{\Sigma}_{\mathsf{X}_t,\mathsf{Y}_t}\boldsymbol{\Sigma}_{\mathsf{Y}_t}^{-1}(y - y^\star) = x - K_t(y - y^\star)$$

The ensemble Kalman filter (EnKF) (Evensen, 1994) estimates $K_t \in \mathbb{R}^{n \times d}$ from samples $\{x^1, \dots, x^M\}$ of the forecast distribution $\pi_{t \mid t-1}$.

Preservation of invariants

Physical systems have important **invariants**, i.e., preserved quantities, $\mathsf{H}\colon \mathbb{R}^n \to \mathbb{R}^r$:

- Mass, $\mathsf{H}(x) = U_\perp^\top x$
- Energy, $\mathsf{H}(x) = x^\top A x$
- Hamiltonian, e.g., $\mathsf{H}(x) = 0.5m||x||^2 + V(x)$
- Stoichiometric balance of chemical species, $\mathsf{H}(x) = U_\perp^\top x$

Preservation of invariants

Physical systems have important **invariants**, i.e., preserved quantities, $\mathsf{H}\colon \mathbb{R}^n \to \mathbb{R}^r$:

- Mass, $\mathsf{H}(x) = \boldsymbol{U}_\perp^\top \boldsymbol{x}$
- Energy, $\mathsf{H}(x) = \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x}$
- Hamiltonian, e.g., $\mathsf{H}(x) = 0.5m||\boldsymbol{x}||^2 + V(\boldsymbol{x})$
- Stoichiometric balance of chemical species, $\mathsf{H}(x) = \boldsymbol{U}_\perp^\top \boldsymbol{x}$

Fact: Modern solvers ensure that **discrete solutions preserve invariants** of the system.

To update or to not update invariants?

Do we want to update the invariants of the system with incoming observations?

To update or to not update invariants?

Do we want to update the invariants of the system with incoming observations?

Scenario 1: The value of the invariant H is known, i.e., $H_\sharp \pi_X$ is a Dirac centered at $C \in \mathbb{R}^r$.

To update or to not update invariants?

Do we want to update the invariants of the system with incoming observations?

Scenario 1: The value of the invariant H is known, i.e., $\mathsf{H}_\sharp \pi_\mathsf{X}$ is a Dirac centered at $\mathsf{C} \in \mathbb{R}^r$.
$\rightarrow$ Bayes' rule should not modify the value of the invariant.

To update or to not update invariants?

Do we want to update the invariants of the system with incoming observations?

Scenario 1: The value of the invariant H is known, i.e., $\mathsf{H}_{\sharp}\pi_{\mathsf{X}}$ is a Dirac centered at $\mathsf{C} \in \mathbb{R}^r$.
$\rightarrow$ Bayes' rule should not modify the value of the invariant.

Scenario 2: The value of the invariant H is uncertain, i.e., $\mathsf{H}_{\sharp}\pi_{\mathsf{X}}$ is not singular.

To update or to not update invariants?

Do we want to update the invariants of the system with incoming observations?

Scenario 1: The value of the invariant H is known, i.e., $H_\sharp \pi_X$ is a Dirac centered at $C \in \mathbb{R}^r$.
$\rightarrow$ Bayes' rule should not modify the value of the invariant.

Scenario 2: The value of the invariant H is uncertain, i.e., $H_\sharp \pi_X$ is not singular.
$\rightarrow$ We want to update the invariant as we are gathering information about the true system.

A preservation property of Bayes' rule

Theorem

- *Consider a prior* π_{X}*, a likelihood model* $\pi_{\mathsf{Y}\,|\,\mathsf{X}}$*, and an invariant* $\boldsymbol{H}\colon \mathbb{R}^n \to \mathbb{R}^r$.
- *Assume that the invariant is* ***constant over the prior*** π_{X}*, i.e.,* $\boldsymbol{H}(\boldsymbol{x}) = \boldsymbol{C} \in \mathbb{R}^r$ *for any realization* $\boldsymbol{x}$ *of* X.
- *Then the invariant is preserved by Bayes' rule and* ***constant over the posterior*** $\pi_{\mathsf{X}\,|\,\mathsf{Y}}$

A preservation property of Bayes' rule

Theorem

- *Consider a prior* π_{X}*, a likelihood model* $\pi_{\mathsf{Y} \mid \mathsf{X}}$*, and an invariant* $\boldsymbol{H} \colon \mathbb{R}^n \to \mathbb{R}^r$*.*
- *Assume that the invariant is* ***constant over the prior*** π_{X}*, i.e.,* $H(x) = C \in \mathbb{R}^r$ *for any realization* x *of* X*.*
- *Then the invariant is preserved by Bayes' rule and* ***constant over the posterior*** $\pi_{\mathsf{X} \mid \mathsf{Y}}$

Proof. $\operatorname{supp}(\pi_{\mathsf{X} \mid \mathsf{Y}}) \subseteq \operatorname{supp}(\pi_{\mathsf{X}}) \subseteq \{x \in \mathbb{R}^n | \mathrm{H}(x) = C\}$.

A preservation property of Bayes' rule

Theorem

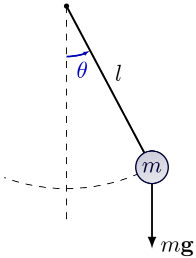
- *Consider a prior π_{X}, a likelihood model $\pi_{\mathsf{Y} \mid \mathsf{X}}$, and an invariant $\boldsymbol{H}\colon \mathbb{R}^n \to \mathbb{R}^r$.*
- *Assume that the invariant is* ***constant over the prior*** *π_{X}, i.e., $\boldsymbol{H}(\boldsymbol{x}) = \boldsymbol{C} \in \mathbb{R}^r$ for any realization $\boldsymbol{x}$ of X.*
- *Then the invariant is preserved by Bayes' rule and* ***constant over the posterior*** *$\pi_{\mathsf{X} \mid \mathsf{Y}}$*

Proof. $\mathrm{supp}(\pi_{\mathsf{X} \mid \mathsf{Y}}) \subseteq \mathrm{supp}(\pi_{\mathsf{X}}) \subseteq \{\boldsymbol{x} \in \mathbb{R}^n | \mathrm{H}(\boldsymbol{x}) = \boldsymbol{C}\}$.

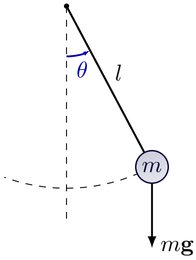
Takeaway

If the invariants are constant over the prior, violations of invariants can be fully attributed to the **discrete approximation** of Bayes' rule.

Oscillating pendulum

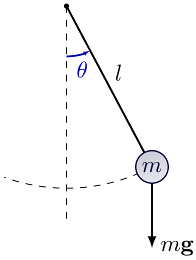


Oscillating pendulum



Hamiltonian structure: $\mathsf{H}(\theta, \dot{\theta}) = \frac{ml^2\dot{\theta}^2}{2} + mgl(1 - \cos(\theta))$

Oscillating pendulum



Hamiltonian structure: $\mathsf{H}(\theta, \dot{\theta}) = \frac{ml^2\dot{\theta}^2}{2} + mgl(1 - \cos(\theta))$

The Hamiltonian H is preserved over time:

$$\frac{\mathrm{d}\mathsf{H}(\theta, \dot{\theta})}{\mathrm{d}t} = 0$$

Oscillating pendulum

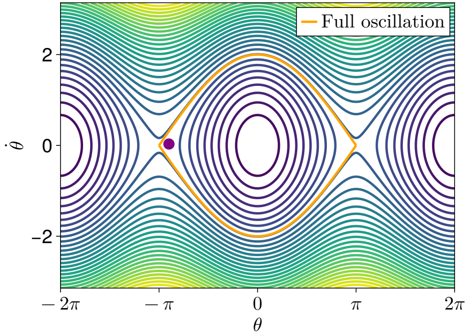


Figure 1: Level sets of $\mathrm{H}(\theta, \dot{\theta})$

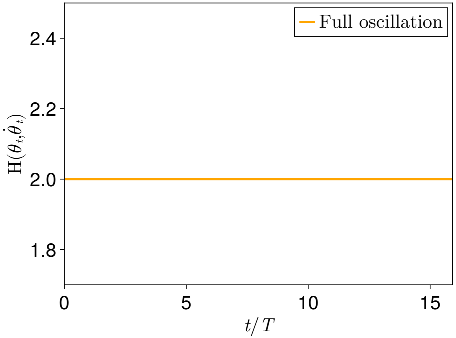


Figure 2: Time evolution of $\mathrm{H}(\theta_t, \dot{\theta}_t)$

- Closed level sets correspond to oscillations.
- Open levels sets correspond to full rotations.

Oscillating pendulum

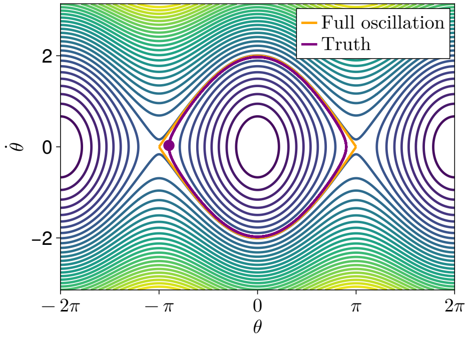


Figure 1: Level sets of $\mathsf{H}(\theta, \dot{\theta})$

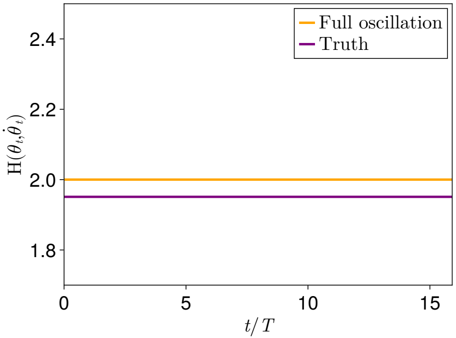


Figure 2: Time evolution of $\mathsf{H}(\theta_t, \dot{\theta}_t)$

- Closed level sets correspond to oscillations.
- Open levels sets correspond to full rotations.
- Symplectic solver preserves H.

Data assimilation for the oscillating pendulum

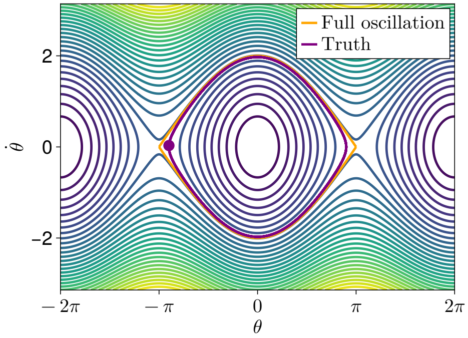


Figure 3: Level sets of $\mathrm{H}(\theta, \dot{\theta})$

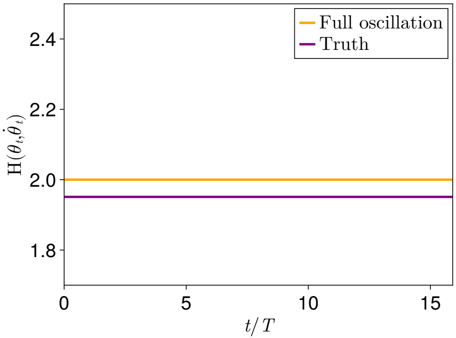


Figure 4: Time evolution of $\mathrm{H}(\theta_t, \dot{\theta}_t)$

$\rightarrow$ Perform data assimilation with EnKF.

Data assimilation for the oscillating pendulum

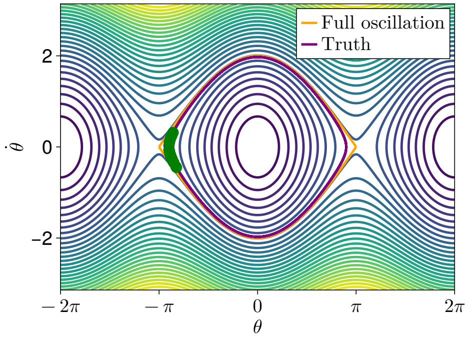


Figure 3: Level sets of $\mathrm{H}(\theta, \dot\theta)$

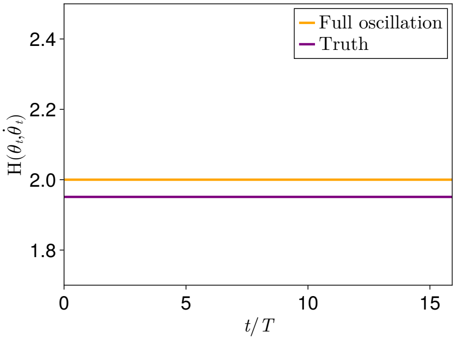


Figure 4: Time evolution of $\mathrm{H}(\theta_t, \dot\theta_t)$

$\rightarrow$ Perform data assimilation with EnKF.

- Initialize ensemble with true Hamiltonian.

Data assimilation for the oscillating pendulum

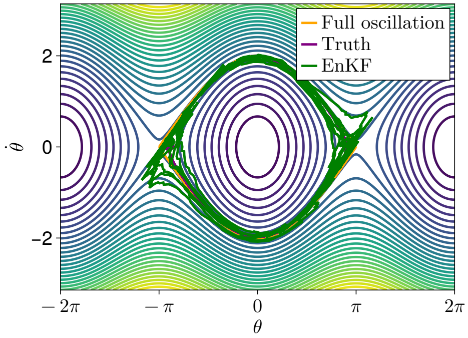


Figure 3: Phase portrait for $\mathrm{H}(\theta, \dot{\theta})$

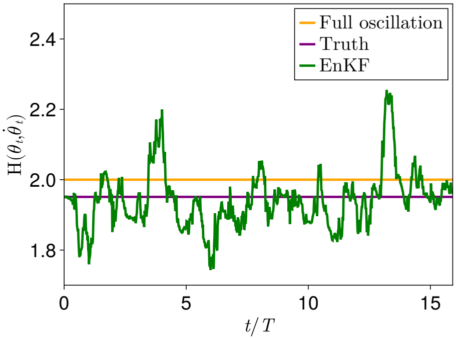


Figure 4: Time evolution of $\mathrm{H}(\theta_t, \dot{\theta}_t)$

→ Perform data assimilation with EnKF.

- Initialize ensemble with true Hamiltonian.
- The EnKF does not preserve **H**.

To update or not update invariants?

Takeaway

Discrete approximations of Bayes' rule can cause **spurious updates** or **break known invariants.**

To update or not update invariants?

Takeaway

Discrete approximations of Bayes' rule can cause **spurious updates** or **break known invariants.**

- We advocate for a **conservative view** on the update of invariants.
- We want to design **discrete algorithms** that respect this preservation property of Bayes' rule.

Preservation of linear invariants

In this talk, we focus on the **preservation** of **linear invariants**, i.e., $\mathrm{H}(x) \colon \mathbb{R}^n \to \mathbb{R}^r, x \mapsto U_\perp^\top x$.

Linear invariants are omnipresent in science and engineering, e.g.,

- Stochiometric balance of chemical reactions
- Mass conservation in conservation laws
- Divergence-free condition in incompressible fluid mechanics
- Kirchhoff's current laws in electrical networks

Example 1: Chemical reaction

Consider the reversible chemical reaction

$$O + \mathrm{NO} \rightleftharpoons \mathrm{NO}_2, \text{ with reaction rates } (k_+, k_-)$$

The associated ODE system is

$$\begin{aligned}
\frac{\mathrm{d}[O]}{\mathrm{d}t} &= -k_+[O][\mathrm{NO}] + k_-[\mathrm{NO}_2] \\
\frac{\mathrm{d}[\mathrm{NO}]}{\mathrm{d}t} &= -k_+[O][\mathrm{NO}] + k_-[\mathrm{NO}_2] \\
\frac{\mathrm{d}[\mathrm{NO}_2]}{\mathrm{d}t} &= k_+[O][\mathrm{NO}] - k_-[\mathrm{NO}_2]
\end{aligned}$$

Conservation of nitrogen and oxygen elements: $\mathrm{H}(\boldsymbol{x}) = \boldsymbol{U}_\perp^\top \boldsymbol{x}$ with $\boldsymbol{U}_\perp = \begin{bmatrix} 0 & 1 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}$,

Preservation of linear invariants

Objective

Introduce a class of analysis maps preserving linear invariants (Lin-PAMs) in the strong sense, i.e.,

If $(\mathsf{y}^{(i)}, \mathsf{x}^{(i)}) \sim \pi_{(\mathsf{Y}_t, \mathsf{X}_t) \mid \mathsf{Y}_{1:t-1} = \boldsymbol{y}^\star_{1:t-1}}$ with $\mathsf{H}(\mathsf{x}^{(i)}) = \mathsf{C}_i \in \mathbb{R}^r$,

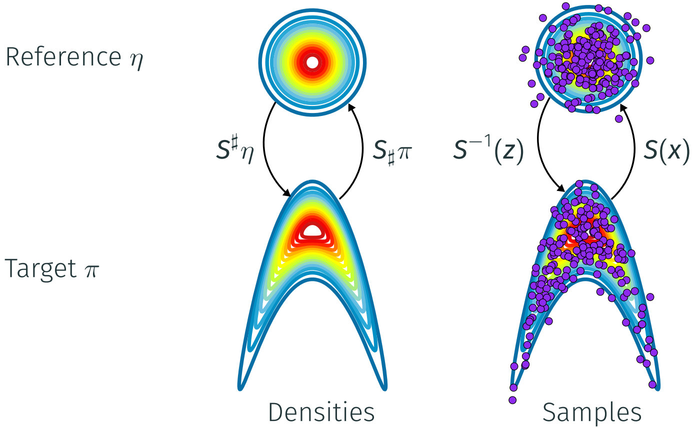
then we want $\mathsf{x}_a{}^{(i)} = \widetilde{T}_{\boldsymbol{y}^\star_t}(\mathsf{y}^{(i)}, \mathsf{x}^{(i)}) \sim \pi_{\mathsf{X}_t \mid \mathsf{Y}_{1:t} = \boldsymbol{y}^\star_{1:t}}$ such that $\mathsf{H}(\mathsf{x}_a{}^{(i)}) = \mathsf{C}_i$.

Idea: Use tools from measure transport

Transport map between two probability distributions

Idea

- Target dist. π = Transformation of a reference dist. η by a map S, i.e., $S_\sharp \pi = \eta$.
- With S, sampling and density estimation are easy.



Looking for a map suited for conditional inference (Marzouk et al., 2016)

We consider the **Knothe-Rosenblatt (KR) rearrangement** S between π and η, defined as the unique lower triangular and monotone map s.t. $S_\sharp \pi = \eta$.

$$S(\boldsymbol{x}) = S(x_1, x_2, \cdots, x_m) = \begin{bmatrix} S^1(x_1) \\ S^2(x_1, x_2) \\ \vdots \\ S^m(x_1, x_2, \ldots, x_m) \end{bmatrix}.$$

Looking for a map suited for conditional inference (Marzouk et al., 2016)

We consider the **Knothe-Rosenblatt (KR) rearrangement** S between π and η, defined as the unique lower triangular and monotone map s.t. $S_\sharp \pi = \eta$.

$$S(\boldsymbol{x}) = S(x_1, x_2, \cdots, x_m) = \begin{bmatrix} S^1(x_1) \\ S^2(x_1, x_2) \\ \vdots \\ S^m(x_1, x_2, \dots, x_m) \end{bmatrix}.$$

The KR has nice features for **Bayesian inference**:

- The 1D map $\xi \mapsto S^k(x_1, x_2, \dots, x_{k-1}, \xi)$ characterizes the **marginal conditional** $\pi_{\mathsf{X}_k \mid \mathsf{X}_{1:k-1} = \boldsymbol{x}_{1:k-1}}(\xi)$.
- S is **easy to invert** and $\det \nabla S(\boldsymbol{x})$ is **fast to evaluate**.

Looking for a map suited for conditional inference (Marzouk et al., 2016)

We consider the **Knothe-Rosenblatt (KR) rearrangement** S between π and η, defined as the unique lower triangular and monotone map s.t. $S_\sharp \pi = \eta$.

$$S(\mathbf{x}) = S(x_1, x_2, \cdots, x_m) = \begin{bmatrix} S^1(x_1) \\ S^2(x_1, x_2) \\ \vdots \\ S^m(x_1, x_2, \dots, x_m) \end{bmatrix}.$$

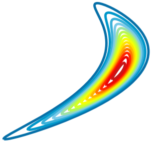
The KR has nice features for **Bayesian inference**:

- The 1D map $\xi \mapsto S^k(x_1, x_2, \dots, x_{k-1}, \xi)$ characterizes the **marginal conditional** $\pi_{\mathsf{X}_k \mid \mathsf{X}_{1:k-1} = \mathbf{x}_{1:k-1}}(\xi)$.
- S is **easy to invert** and $\det \nabla S(\mathbf{x})$ is **fast to evaluate**.

Gaussian case

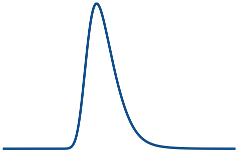
Consider $\mathsf{X} \sim \pi = \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and let $LL^\top = \boldsymbol{\Sigma}^{-1}$ be the Cholesky factorization of $\boldsymbol{\Sigma}^{-1}$. Then $S(\mathbf{x}) = L(\mathbf{x} - \boldsymbol{\mu})$ is the KR that pushes forward π to $\eta = \mathcal{N}(\mathbf{0}_n, \boldsymbol{I}_n)$.

Construction of the analysis map (Spantini et al., 2022)



$\pi_{Y,X}$

$T_{y^\star}(y, x)$

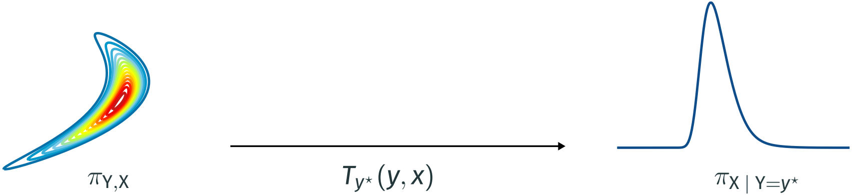


$\pi_{X \mid Y = y^\star}$

Construction of the analysis map (Spantini et al., 2022)

Consider the KR rearrangement S s.t. $S_\sharp \pi_{\mathsf{Y},\mathsf{X}} = \eta_{\mathsf{Y}} \otimes \eta_{\mathsf{X}}$

$$S(y, x) = \begin{bmatrix} S^{\mathcal{Y}}(y) \\ S^{\mathcal{X}}(y, x) \end{bmatrix},$$

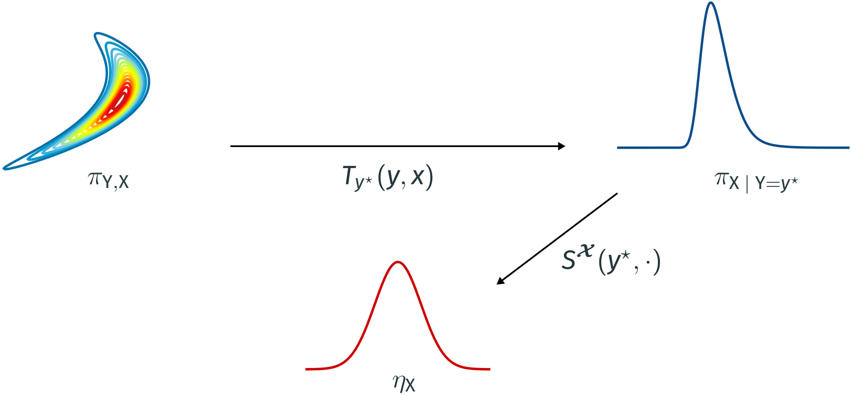


Construction of the analysis map (Spantini et al., 2022)

Consider the KR rearrangement S s.t. $S_\sharp \pi_{Y,X} = \eta_Y \otimes \eta_X$

$$S(y, x) = \begin{bmatrix} S^{\mathcal{Y}}(y) \\ S^{\mathcal{X}}(y, x) \end{bmatrix},$$

- The map $\xi \mapsto S^{\mathcal{X}}(y, \xi)$ pushes forward $\pi_{X \mid Y}(\cdot \mid y)$ to η_X for any y.

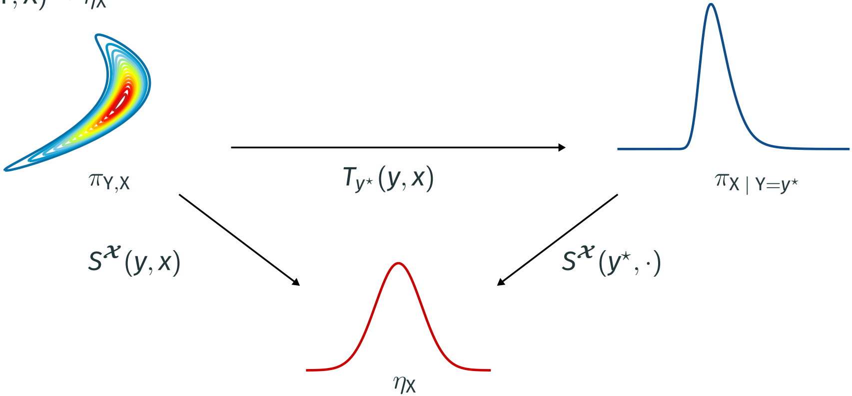


Construction of the analysis map (Spantini et al., 2022)

Consider the KR rearrangement S s.t. $S_\sharp \pi_{\mathsf{Y},\mathsf{X}} = \eta_{\mathsf{Y}} \otimes \eta_{\mathsf{X}}$

$$S(y,x) = \begin{bmatrix} S^{\mathcal{Y}}(y) \\ S^{\mathcal{X}}(y,x) \end{bmatrix},$$

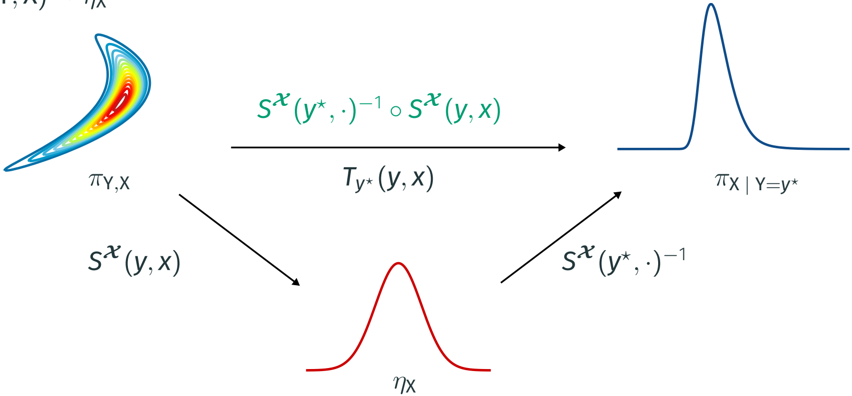
- The map $\xi \mapsto S^{\mathcal{X}}(y,\xi)$ pushes forward $\pi_{\mathsf{X}\,|\,\mathsf{Y}}(\cdot \mid y)$ to η_{X} for any y.
- $S^{\mathcal{X}}(\mathsf{Y},\mathsf{X}) \sim \eta_{\mathsf{X}}$



Consider the KR rearrangement S s.t. $S_\sharp \pi_{\mathsf{Y},\mathsf{X}} = \eta_{\mathsf{Y}} \otimes \eta_{\mathsf{X}}$

$$S(y,x) = \begin{bmatrix} S^{\mathcal{Y}}(y) \\ S^{\mathcal{X}}(y,x) \end{bmatrix},$$

- The map $\xi \mapsto S^{\mathcal{X}}(y,\xi)$ pushes forward $\pi_{\mathsf{X} \mid \mathsf{Y}}(\cdot \mid y)$ to η_{X} for any y.
- $S^{\mathcal{X}}(\mathsf{Y},\mathsf{X}) \sim \eta_{\mathsf{X}}$



A broad class of ensemble filters (Le Provost et al., 2023)

Analysis map $T_{y^\star}$: $T_{y^\star}(y, x) = S^{\boldsymbol{\mathcal{X}}}(y^\star, \cdot)^{-1} \circ S^{\boldsymbol{\mathcal{X}}}(y, x)$

A broad class of ensemble filters (Le Provost et al., 2023)

Analysis map $T_{y^\star}$: $T_{y^\star}(y, x) = S^{\mathcal{X}}(y^\star, \cdot)^{-1} \circ S^{\mathcal{X}}(y, x)$

This formulation can represent a **broad class** of ensemble filters by choosing

A broad class of ensemble filters (Le Provost et al., 2023)

Analysis map $T_{y^\star}$: $T_{y^\star}(y, x) = S^{\mathcal{X}}(y^\star, \cdot)^{-1} \circ S^{\mathcal{X}}(y, x)$

This formulation can represent a **broad class** of ensemble filters by choosing

- the reference density η_{X},

A broad class of ensemble filters (Le Provost et al., 2023)

Analysis map $\boldsymbol{T}_{y^\star}$: $\boldsymbol{T}_{y^\star}(y, x) = \boldsymbol{S}^{\boldsymbol{\mathcal{X}}}(y^\star, \cdot)^{-1} \circ \boldsymbol{S}^{\boldsymbol{\mathcal{X}}}(y, x)$

This formulation can represent a **broad class** of ensemble filters by choosing

- the reference density η_{X},
- the class of functions to approximate $\boldsymbol{S}^{\boldsymbol{\mathcal{X}}}$ or $\boldsymbol{T}_{y^\star}$,

A broad class of ensemble filters (Le Provost et al., 2023)

Analysis map $T_{y^\star}$: $T_{y^\star}(y, x) = S^{\mathcal{X}}(y^\star, \cdot)^{-1} \circ S^{\mathcal{X}}(y, x)$

This formulation can represent a **broad class** of ensemble filters by choosing

- the reference density η_{X},
- the class of functions to approximate $S^{\mathcal{X}}$ or $T_{y^\star}$,
- and the estimation of $S^{\mathcal{X}}$ or $T_{y^\star}$ from samples

A broad class of ensemble filters (Le Provost et al., 2023)

Analysis map $\boldsymbol{T}_{\boldsymbol{y}^\star}$: $\boldsymbol{T}_{\boldsymbol{y}^\star}(\boldsymbol{y}, \boldsymbol{x}) = \boldsymbol{S}^{\boldsymbol{\mathcal{X}}}(\boldsymbol{y}^\star, \cdot)^{-1} \circ \boldsymbol{S}^{\boldsymbol{\mathcal{X}}}(\boldsymbol{y}, \boldsymbol{x})$

This formulation can represent a **broad class** of ensemble filters by choosing

- the reference density η_{X},
- the class of functions to approximate $\boldsymbol{S}^{\boldsymbol{\mathcal{X}}}$ or $\boldsymbol{T}_{\boldsymbol{y}^\star}$,
- and the estimation of $\boldsymbol{S}^{\boldsymbol{\mathcal{X}}}$ or $\boldsymbol{T}_{\boldsymbol{y}^\star}$ from samples

Stochastic EnKF (Evensen, 1994)

- $\eta_{\mathsf{X}} = \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$
- Linear $\boldsymbol{S}^{\boldsymbol{\mathcal{X}}}$
- (Localized) sample covariance estimator $\widehat{\boldsymbol{\Sigma}}_{\mathsf{X}_t} = \rho \circ \left(\frac{1}{M} \sum_{i=1}^{M} (\boldsymbol{x}^{(i)} - \widehat{\boldsymbol{\mu}}_X)(\boldsymbol{x}^{(i)} - \widehat{\boldsymbol{\mu}}_X)^\top\right)$

How to construct analysis maps $T_{y^\star}$ preserving linear invariants $x \mapsto U_\perp^\top x$?

How to construct analysis maps $T_{y^\star}$ preserving linear invariants $x \mapsto U_\perp^\top x$?

Idea: Formulate the analysis map in the right coordinate system.

A state decomposition

Up to a thin QR factorization of $\boldsymbol{U}_\perp$, assume that $\boldsymbol{U}_\perp \in \mathbb{R}^{n \times r}$ is sub-unitary, i.e., $\boldsymbol{U}_\perp^\top \boldsymbol{U}_\perp = \boldsymbol{I}_r$.

A state decomposition

Up to a thin QR factorization of $\boldsymbol{U}_\perp$, assume that $\boldsymbol{U}_\perp \in \mathbb{R}^{n\times r}$ is sub-unitary, i.e., $\boldsymbol{U}_\perp^\top \boldsymbol{U}_\perp = \boldsymbol{I}_r$.

We choose $\boldsymbol{U}_{||} \in \mathbb{R}^{n\times(n-r)}$ such that $\boldsymbol{U} = [\boldsymbol{U}_\perp, \boldsymbol{U}_{||}] \in \mathbb{R}^{n\times n}$ form an orthonormal basis of $\mathbb{R}^n$.

A state decomposition

Up to a thin QR factorization of $\boldsymbol{U}_\perp$, assume that $\boldsymbol{U}_\perp \in \mathbb{R}^{n\times r}$ is sub-unitary, i.e., $\boldsymbol{U}_\perp^\top \boldsymbol{U}_\perp = \boldsymbol{I}_r$.

We choose $\boldsymbol{U}_{||} \in \mathbb{R}^{n\times(n-r)}$ such that $\boldsymbol{U} = [\boldsymbol{U}_\perp, \boldsymbol{U}_{||}] \in \mathbb{R}^{n\times n}$ form an orthonormal basis of $\mathbb{R}^n$.

We decompose the state $\boldsymbol{x} \in \mathbb{R}^n$ as

$$\begin{aligned} \boldsymbol{x} &= \boldsymbol{U}_\perp \boldsymbol{U}_\perp^\top \boldsymbol{x} \oplus \boldsymbol{U}_{||} \boldsymbol{U}_{||}^\top \boldsymbol{x}, \\ &= \boldsymbol{U}_\perp \boldsymbol{x}_\perp \oplus \boldsymbol{U}_{||} \boldsymbol{x}_{||}, \end{aligned}$$

with $\boldsymbol{x}_\perp = \boldsymbol{U}_\perp^\top \boldsymbol{x} \in \mathbb{R}^r$ and $\boldsymbol{x}_{||} = \boldsymbol{U}_{||}^\top \boldsymbol{x} \in \mathbb{R}^{n-r}$.

Lower triangular maps characterize conditional distributions

Consider the rotation $(\mathsf{Y}, \mathsf{X}) \mapsto (\mathsf{Y}, [U_\perp, U_\parallel]^\top \mathsf{X}) = (\mathsf{Y}, \mathsf{X}_\perp, \mathsf{X}_\parallel)$.

Lower triangular maps characterize conditional distributions

Consider the rotation $(\mathsf{Y}, \mathsf{X}) \mapsto (\mathsf{Y}, [U_\perp, U_{\|}]^\top \mathsf{X}) = (\mathsf{Y}, \mathsf{X}_\perp, \mathsf{X}_{\|})$.

We have the following factorization of $\pi_{\mathsf{Y},\mathsf{X}}$:

$$\pi_{\mathsf{Y},\mathsf{X}}(y, x) = \pi_{\mathsf{Y},\mathsf{X}_\perp,\mathsf{X}_{\|}}(y, x_\perp, x_{\|}) = \pi_{\mathsf{Y}}(y)\pi_{\mathsf{X}_\perp \mid \mathsf{Y}}(x_\perp \mid y)\pi_{\mathsf{X}_{\|} \mid \mathsf{Y},\mathsf{X}_\perp}(x_{\|} \mid y, x_\perp)$$

Lower triangular maps characterize conditional distributions

Consider the rotation $(\mathsf{Y}, \mathsf{X}) \mapsto (\mathsf{Y}, [U_\perp, U_\parallel]^\top \mathsf{X}) = (\mathsf{Y}, \mathsf{X}_\perp, \mathsf{X}_\parallel)$.

We have the following factorization of $\pi_{\mathsf{Y},\mathsf{X}}$:

$$\pi_{\mathsf{Y},\mathsf{X}}(y, x) = \pi_{\mathsf{Y},\mathsf{X}_\perp,\mathsf{X}_\parallel}(y, x_\perp, x_\parallel) = \pi_{\mathsf{Y}}(y)\pi_{\mathsf{X}_\perp \mid \mathsf{Y}}(x_\perp \mid y)\pi_{\mathsf{X}_\parallel \mid \mathsf{Y},\mathsf{X}_\perp}(x_\parallel \mid y, x_\perp)$$

If S pushes forward $\pi_{\mathsf{Y},\mathsf{X}_\perp,\mathsf{X}_\parallel}$ to $\eta_{\mathsf{Y}} \otimes \eta_{\mathsf{X}_\perp} \otimes \eta_{\mathsf{X}_\parallel}$ and S is lower triangular, i.e.,

$$S(y, x_\perp, x_\parallel) = \begin{bmatrix} S^{\mathcal{Y}}(y) \\ S^{\mathcal{X}_\perp}(y, x_\perp) \\ S^{\mathcal{X}_\parallel}(y, x_\perp, x_\parallel) \end{bmatrix}, \qquad \text{then} \qquad \begin{aligned} S^{\mathcal{Y}}{}_\sharp \pi_{\mathsf{Y}} &= \eta_{\mathsf{Y}}, \\ S^{\mathcal{X}_\perp}{}_\sharp \pi_{\mathsf{X}_\perp \mid \mathsf{Y}} &= \eta_{\mathsf{X}_\perp}, \\ S^{\mathcal{X}_\parallel}{}_\sharp \pi_{\mathsf{X}_\parallel \mid \mathsf{Y},\mathsf{X}_\perp} &= \eta_{\mathsf{X}_\parallel} \end{aligned}$$

See (Baptista et al., 2020) for the proof.

Analysis map $T_{y^\star}$ in the rotated space $(\mathsf{Y}, \mathsf{X}_\perp, \mathsf{X}_{||})$

The following analysis map $T^{\perp}_{y^\star}$ pushes forward $\pi_{\mathsf{Y},\mathsf{X}_\perp}$ to $\pi_{\mathsf{X}_\perp \mid \mathsf{Y}=y^\star}$,

$$T^{\perp}_{y^\star}(y, x_\perp) = S^{\mathcal{X}_\perp}(y^\star, \cdot)^{-1} \circ S^{\mathcal{X}_\perp}(y, x_\perp).$$

The following analysis map $T^{||}_{y^\star, x_{\perp,a}}$ pushes forward $\pi_{\mathsf{Y},\mathsf{X}_\perp,\mathsf{X}_{||}}$ to $\pi_{\mathsf{X}_{||} \mid \mathsf{Y}=y^\star, \mathsf{X}_\perp = x_{\perp,a}}$,

$$T^{||}_{y^\star, x_{\perp,a}}(y, x_\perp, x_{||}) = S^{\mathcal{X}_{||}}(y^\star, T^{\perp}_{y^\star}(y, x_\perp), \cdot)^{-1} \circ S^{\mathcal{X}_{||}}(y, x_\perp, x_{||}).$$

Perform inference in the rotated space by recursive updates:

1. Update coordinate $x_\perp$ with $T^{\perp}_{y^\star}$
2. Update coordinate $x_{||}$ with $T^{||}_{y^\star, x_{\perp,a}}$

Analysis map formulated in the original space

$$T_{y^\star}(y, x) = U_\perp T^{\perp}_{y^\star}(y, U_\perp^\top x) + U_{||} T^{||}_{y^\star}(y, U_\perp^\top x, U_{||}^\top x)$$

Formulation of linear invariant-preserving analysis map (Lin-PAM)

In the rotated space, the invariants are given by the coordinates $\mathbf{x}_\perp$.

Formulation of linear invariant-preserving analysis map (Lin-PAM)

In the rotated space, the invariants are given by the coordinates $x_\perp$.

To preserve invariants, we set the analysis map $\widetilde{T}^{\perp}_{y^\star}$ to the identity, i.e., $\widetilde{T}^{\perp}_{y^\star}(y, x_\perp) = x_\perp$.

Formulation of linear invariant-preserving analysis map (Lin-PAM)

In the rotated space, the invariants are given by the coordinates $x_\perp$.

To preserve invariants, we set the analysis map $\widetilde{T}^{\perp}_{y^\star}$ to the identity, i.e., $\widetilde{T}^{\perp}_{y^\star}(y, x_\perp) = x_\perp$.

We obtain the constrained analysis map $\widetilde{T}^{||}_{y^\star}$ as

$$\widetilde{T}^{||}_{y^\star}(y, x_\perp, x_{||}) = S^{\mathcal{X}_{||}}(y^\star, x_\perp, \cdot)^{-1} \circ S^{\mathcal{X}_{||}}(y, x_\perp, x_{||}).$$

Formulation of linear invariant-preserving analysis map (Lin-PAM)

In the rotated space, the invariants are given by the coordinates $\boldsymbol{x}_\perp$.

To preserve invariants, we set the analysis map $\widetilde{\boldsymbol{T}}^{\perp}_{\boldsymbol{y}^\star}$ to the identity, i.e., $\widetilde{\boldsymbol{T}}^{\perp}_{\boldsymbol{y}^\star}(\boldsymbol{y}, \boldsymbol{x}_\perp) = \boldsymbol{x}_\perp$.

We obtain the constrained analysis map $\widetilde{\boldsymbol{T}}^{||}_{\boldsymbol{y}^\star}$ as

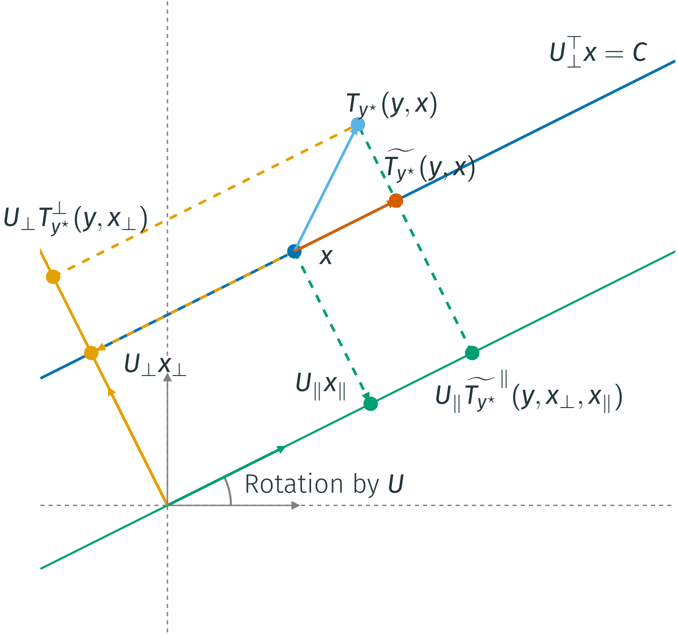
$$\widetilde{\boldsymbol{T}}^{||}_{\boldsymbol{y}^\star}(\boldsymbol{y}, \boldsymbol{x}_\perp, \boldsymbol{x}_{||}) = \boldsymbol{S}^{\boldsymbol{\mathcal{X}}_{||}}(\boldsymbol{y}^\star, \boldsymbol{x}_\perp, \cdot)^{-1} \circ \boldsymbol{S}^{\boldsymbol{\mathcal{X}}_{||}}(\boldsymbol{y}, \boldsymbol{x}_\perp, \boldsymbol{x}_{||}).$$

The Lin-PAM $\widetilde{\boldsymbol{T}}_{\boldsymbol{y}^\star}$ formulated in the original space reads

$$\begin{aligned}\widetilde{\boldsymbol{T}}_{\boldsymbol{y}^\star}(\boldsymbol{y}, \boldsymbol{x}) &= \boldsymbol{U}_\perp \widetilde{\boldsymbol{T}}^{\perp}_{\boldsymbol{y}^\star}(\boldsymbol{y}, \boldsymbol{U}_\perp^\top \boldsymbol{x}) + \boldsymbol{U}_{||} \widetilde{\boldsymbol{T}}^{||}_{\boldsymbol{y}^\star}(\boldsymbol{y}, \boldsymbol{U}_\perp^\top \boldsymbol{x}, \boldsymbol{U}_{||}^\top \boldsymbol{x}) \\ &= \boldsymbol{U}_\perp \boldsymbol{U}_\perp^\top \boldsymbol{x} + \boldsymbol{U}_{||} \widetilde{\boldsymbol{T}}^{||}_{\boldsymbol{y}^\star}(\boldsymbol{y}, \boldsymbol{U}_\perp^\top \boldsymbol{x}, \boldsymbol{U}_{||}^\top \boldsymbol{x}).\end{aligned}$$

Le Provost, M., Glaubitz, J., and Marzouk Y. (2024), **"Preserving linear invariants in ensemble filtering methods."**, arXiv:2404.14328

A schematic summary



Empirical approximations of a Lin-PAM preserve linear invariants

In practice, we use an empirical estimator $\widehat{T}_{y^\star}$ of the form

$$\widehat{T}_{y^\star}(y, x) = U_\perp U_\perp^\top x + U_\parallel \widehat{T}^{\parallel}_{y^\star}(y, U_\perp^\top x, U_\parallel^\top x)$$

Empirical approximations of a Lin-PAM preserve linear invariants

In practice, we use an empirical estimator $\widehat{T}_{y^\star}$ of the form

$$\widehat{T}_{y^\star}(y, x) = U_\perp U_\perp^\top x + U_\parallel \widehat{T}_{y^\star}^{\parallel}(y, U_\perp^\top x, U_\parallel^\top x)$$

where the map $\widehat{T}_{y^\star}^{\parallel}$ is imperfect due to

Empirical approximations of a Lin-PAM preserve linear invariants

In practice, we use an empirical estimator $\widehat{T}_{y^\star}$ of the form

$$\widehat{T}_{y^\star}(y, x) = U_\perp U_\perp^\top x + U_\| \widehat{T}^{\|}_{y^\star}(y, U_\perp^\top x, U_\|^\top x)$$

where the map $\widehat{T}^{\|}_{y^\star}$ is imperfect due to

- Choice of an approximation class, e.g., radial basis functions, polynomials, NN

Empirical approximations of a Lin-PAM preserve linear invariants

In practice, we use an empirical estimator $\widehat{T}_{y^\star}$ of the form

$$\widehat{T}_{y^\star}(y, x) = U_\perp U_\perp^\top x + U_\| \widehat{T}_{y^\star}^{\|}(y, U_\perp^\top x, U_\|^\top x)$$

where the map $\widehat{T}_{y^\star}^{\|}$ is imperfect due to

- Choice of an approximation class, e.g., radial basis functions, polynomials, NN
- Estimation from samples $\{(y^{(i)}, x^{(i)})\}$ of an approximation of the forecast distribution $\widehat{\pi}_{(\mathsf{Y}_t, \mathsf{X}_t) \mid \mathsf{Y}_{1:t-1} = y^\star_{1:t-1}}$

Empirical approximations of a Lin-PAM preserve linear invariants

In practice, we use an empirical estimator $\widehat{T}_{y^\star}$ of the form

$$\widehat{T}_{y^\star}(y, x) = U_\perp U_\perp^\top x + U_\| \widehat{T}^{\|}_{y^\star}(y, U_\perp^\top x, U_\|^\top x)$$

where the map $\widehat{T}^{\|}_{y^\star}$ is imperfect due to

- Choice of an approximation class, e.g., radial basis functions, polynomials, NN
- Estimation from samples $\{(y^{(i)}, x^{(i)})\}$ of an approximation of the forecast distribution $\widehat{\pi}_{(\mathsf{Y}_t, \mathsf{X}_t) \mid \mathsf{Y}_{1:t-1} = y^\star_{1:t-1}}$
- Observation to assimilate $y^\star_t \sim \widehat{\pi}_{\mathsf{Y}_t}$

Takeway: Independently of the quality of $\widehat{T}^{\|}_{y^\star}$, $\widehat{T}_{y^\star}$ still preserves the invariants $x \to U_\perp^\top x$.

Preservation of linear invariants for the Kalman filter and EnKF

Fact: The vanilla Kalman filter and EnKF preserve linear invariants if they are constant over the prior π_{X}

Why do we need Lin-PAMs in the Gaussian case?

Preservation of linear invariants for the Kalman filter and EnKF

Fact: The vanilla Kalman filter and EnKF preserve linear invariants if they are constant over the prior π_{X}

Why do we need Lin-PAMs in the Gaussian case?

This result no longer holds when the EnKF is regularized.

Preservation of linear invariants for the Kalman filter and EnKF

Fact: The vanilla Kalman filter and EnKF preserve linear invariants if they are constant over the prior π_X

Why do we need Lin-PAMs in the Gaussian case?

This result no longer holds when the EnKF is regularized.

Two opposing mechanisms:

- Regularization such as covariance tapering based on the **local conditional structure** of $\pi_{Y,X}$.
  $\rightarrow$ Essentially discard updates at long distances.

Preservation of linear invariants for the Kalman filter and EnKF

Fact: The vanilla Kalman filter and EnKF preserve linear invariants if they are constant over the prior π_X

Why do we need Lin-PAMs in the Gaussian case?

This result no longer holds when the EnKF is regularized.

Two opposing mechanisms:

- Regularization such as covariance tapering based on the **local conditional structure** of $\pi_{Y,X}$.
  $\rightarrow$ Essentially discard updates at long distances.
- Most invariants are **global**, i.e., $H(x)$ depends on all the state components.

$\rightarrow$ We show how to reconcile them.

Lin-PAM in the Gaussian case (i)

Let (Y, X) be jointly Gaussian distributed with

$$\begin{bmatrix} \mathsf{Y} \\ \mathsf{X} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu}_{\mathsf{X}} \\ \boldsymbol{\mu}_{\mathsf{Y}} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{\mathsf{Y}} & \boldsymbol{\Sigma}_{\mathsf{X},\mathsf{Y}}^{\top} \\ \boldsymbol{\Sigma}_{\mathsf{X},\mathsf{Y}} & \boldsymbol{\Sigma}_{\mathsf{X}} \end{bmatrix} \right).$$

The KR rearrangement S that pushes forward $\pi_{\mathsf{Y},\mathsf{X}_\perp,\mathsf{X}_\parallel}$ to $\eta\mathsf{Y} \otimes \eta\mathsf{X}_\perp \otimes \eta\mathsf{X}_\parallel$ is given by

$$S(y, x_\perp, x_\parallel) = \begin{bmatrix} S^{\mathcal{Y}}(y) \\ S^{\mathcal{X}_\perp}(y, x_\perp) \\ S^{\mathcal{X}_\parallel}(y, x_\perp, x_\parallel) \end{bmatrix} = \begin{bmatrix} L_{\mathsf{Y}}(y - \boldsymbol{\mu}_{\mathsf{Y}}) \\ L_{\mathsf{X}_\perp \mid \mathsf{Y}} \left(x_\perp - \boldsymbol{\mu}_{\mathsf{X}_\perp \mid \mathsf{Y}} \right) \\ L_{\mathsf{X}_\parallel \mid \mathsf{Y},\mathsf{X}_\perp} (x_\parallel - \boldsymbol{\mu}_{\mathsf{X}_\parallel \mid \mathsf{Y},\mathsf{X}_\perp}) \end{bmatrix}.$$

For $\mathsf{Z} \sim \mathcal{N}(\boldsymbol{\mu}_{\mathsf{Z}}, \boldsymbol{\Sigma}_{\mathsf{Z}})$, $\boldsymbol{\Sigma}_{\mathsf{Z}}^{-1} = L_{\mathsf{Z}} L_{\mathsf{Z}}^{\top}$ is the Cholesky factorization of $\boldsymbol{\Sigma}_{\mathsf{Z}}^{-1}$.

Lin-PAM in the Gaussian case (ii)

We obtain the unconstrained analysis maps $T^{\perp}_{y^\star}$ and $T^{\|}_{y^\star}$

$$T^{\perp}_{y^\star}(y, x_\perp) = S^{\mathcal{X}_\perp}(y^\star, \cdot)^{-1} \circ S^{\mathcal{X}_\perp}(y, x_\perp) = x_\perp - \boldsymbol{\Sigma}_{\mathrm{X}_\perp,\mathrm{Y}} \boldsymbol{\Sigma}_{\mathrm{Y}}^{-1}(y - y^\star),$$

$$T^{\|}_{y^\star}(y, x_\perp, x_\|) = S^{\mathcal{X}_\|}(y^\star, T^{\perp}_{y^\star}(y, x_\perp), \cdot)^{-1} \circ S^{\mathcal{X}_\|}(y, x_\perp, x_\|) = x_\| - \boldsymbol{\Sigma}_{\mathrm{X}_\|,\mathrm{Y}} \boldsymbol{\Sigma}_{\mathrm{Y}}^{-1}(y - y^\star).$$

Lin-PAM in the Gaussian case (ii)

We obtain the unconstrained analysis maps $T^{\perp}_{y^\star}$ and $T^{\|}_{y^\star}$

$$T^{\perp}_{y^\star}(y, x_\perp) = S^{\mathcal{X}_\perp}(y^\star, \cdot)^{-1} \circ S^{\mathcal{X}_\perp}(y, x_\perp) = x_\perp - \boldsymbol{\Sigma}_{\mathsf{X}_\perp,\mathsf{Y}} \boldsymbol{\Sigma}_{\mathsf{Y}}^{-1}(y - y^\star),$$

$$T^{\|}_{y^\star}(y, x_\perp, x_\|) = S^{\mathcal{X}_\|}(y^\star, T^{\perp}_{y^\star}(y, x_\perp), \cdot)^{-1} \circ S^{\mathcal{X}_\|}(y, x_\perp, x_\|) = x_\| - \boldsymbol{\Sigma}_{\mathsf{X}_\|,\mathsf{Y}} \boldsymbol{\Sigma}_{\mathsf{Y}}^{-1}(y - y^\star).$$

- Kalman filter update in $\mathrm{span}(U_\perp)$ and $\mathrm{span}(U_\|)$

Lin-PAM in the Gaussian case (ii)

We obtain the unconstrained analysis maps $T^{\perp}_{y^\star}$ and $T^{||}_{y^\star}$

$$T^{\perp}_{y^\star}(y, x_\perp) = S^{\mathcal{X}_\perp}(y^\star, \cdot)^{-1} \circ S^{\mathcal{X}_\perp}(y, x_\perp) = x_\perp - \boldsymbol{\Sigma}_{\mathsf{X}_\perp, \mathsf{Y}} \boldsymbol{\Sigma}_{\mathsf{Y}}^{-1}(y - y^\star),$$

$$T^{||}_{y^\star}(y, x_\perp, x_{||}) = S^{\mathcal{X}_{||}}(y^\star, T^{\perp}_{y^\star}(y, x_\perp), \cdot)^{-1} \circ S^{\mathcal{X}_{||}}(y, x_\perp, x_{||}) = x_{||} - \boldsymbol{\Sigma}_{\mathsf{X}_{||}, \mathsf{Y}} \boldsymbol{\Sigma}_{\mathsf{Y}}^{-1}(y - y^\star).$$

- Kalman filter update in $\mathrm{span}(U_\perp)$ and $\mathrm{span}(U_{||})$
- Despite the recursive update, $T^{||}_{y^\star}(y, x_\perp, x_{||})$ does not depend on $x_\perp$.

Lin-PAM in the Gaussian case (ii)

We obtain the unconstrained analysis maps $T^{\perp}_{y^\star}$ and $T^{\|}_{y^\star}$

$$T^{\perp}_{y^\star}(y, x_\perp) = S^{\mathcal{X}_\perp}(y^\star, \cdot)^{-1} \circ S^{\mathcal{X}_\perp}(y, x_\perp) = x_\perp - \boldsymbol{\Sigma}_{\mathsf{X}_\perp,\mathsf{Y}} \boldsymbol{\Sigma}_{\mathsf{Y}}^{-1}(y - y^\star),$$

$$T^{\|}_{y^\star}(y, x_\perp, x_\|) = S^{\mathcal{X}_\|}(y^\star, T^{\perp}_{y^\star}(y, x_\perp), \cdot)^{-1} \circ S^{\mathcal{X}_\|}(y, x_\perp, x_\|) = x_\| - \boldsymbol{\Sigma}_{\mathsf{X}_\|,\mathsf{Y}} \boldsymbol{\Sigma}_{\mathsf{Y}}^{-1}(y - y^\star).$$

- Kalman filter update in $\mathrm{span}(U_\perp)$ and $\mathrm{span}(U_\|)$
- Despite the recursive update, $T^{\|}_{y^\star}(y, x_\perp, x_\|)$ does not depend on $x_\perp$.

$\rightarrow$ Update of $x_\perp$ and $x_\|$ can be decoupled.

Lin-PAM in the Gaussian case (iii)

The analysis map in the original space $T_{y^\star}$ reads

$$
\begin{aligned}
T_{y^\star}(y, x) &= U_\perp T^\perp_{y^\star}(y, x_\perp) + U_\| T^\|_{y^\star}(y, x_\perp, x_\|) \\
&= x - \boldsymbol{\Sigma}_{X,Y}\boldsymbol{\Sigma}_Y^{-1}(y - y^\star),
\end{aligned}
$$

Lin-PAM in the Gaussian case (iii)

The analysis map in the original space $T_{y^\star}$ reads

$$\begin{aligned} T_{y^\star}(y, x) &= U_\perp T^\perp_{y^\star}(y, x_\perp) + U_{||} T^{||}_{y^\star}(y, x_\perp, x_{||}) \\ &= x - \mathbf{\Sigma}_{\mathrm{X},\mathrm{Y}} \mathbf{\Sigma}_{\mathrm{Y}}^{-1}(y - y^\star), \end{aligned}$$

We recover the **Kalman filter's update**.

Lin-PAM in the Gaussian case (iii)

The analysis map in the original space $T_{y^\star}$ reads

$$\begin{aligned} T_{y^\star}(y,x) &= U_\perp T^{\perp}_{y^\star}(y,x_\perp) + U_{||} T^{||}_{y^\star}(y,x_\perp,x_{||}) \\ &= x - \boldsymbol{\Sigma}_{\mathrm{X,Y}}\boldsymbol{\Sigma}_{\mathrm{Y}}^{-1}(y-y^\star), \end{aligned}$$

We recover the **Kalman filter's update**.

The (constrained) analysis map $\widetilde{T}_{y^\star}$ preserving the invariant $\mathrm{H}(x) = U_\perp^\top x$ reads

$$\begin{aligned} \widetilde{T}_{y^\star}(y,x) &= U_\perp x_\perp + U_{||} T^{||}_{y^\star}(y,x_\perp,x_{||}) \\ &= x - (I - U_\perp U_\perp^\top)\boldsymbol{\Sigma}_{\mathrm{X,Y}}\boldsymbol{\Sigma}_{\mathrm{Y}}^{-1}(y-y^\star). \end{aligned}$$

Lin-PAM in the Gaussian case (iii)

The analysis map in the original space $T_{y^\star}$ reads

$$T_{y^\star}(y, x) = U_\perp T^\perp_{y^\star}(y, x_\perp) + U_{||} T^{||}_{y^\star}(y, x_\perp, x_{||})$$
$$= x - \boldsymbol{\Sigma}_{\mathsf{X},\mathsf{Y}} \boldsymbol{\Sigma}_{\mathsf{Y}}^{-1}(y - y^\star),$$

We recover the **Kalman filter's update**.

The (constrained) analysis map $\widetilde{T}_{y^\star}$ preserving the invariant $\mathsf{H}(x) = U_\perp^\top x$ reads

$$\widetilde{T}_{y^\star}(y, x) = U_\perp x_\perp + U_{||} T^{||}_{y^\star}(y, x_\perp, x_{||})$$
$$= x - (I - U_\perp U_\perp^\top) \boldsymbol{\Sigma}_{\mathsf{X},\mathsf{Y}} \boldsymbol{\Sigma}_{\mathsf{Y}}^{-1}(y - y^\star).$$

We recover a **projected formulation of the Kalman filter** (Simon, 2010).

Lin-PAM in the Gaussian case (iii)

The analysis map in the original space $T_{y^\star}$ reads

$$\begin{aligned} T_{y^\star}(y, x) &= U_\perp T_{y^\star}^\perp(y, x_\perp) + U_{||} T_{y^\star}^{||}(y, x_\perp, x_{||}) \\ &= x - \mathbf{\Sigma}_{X,Y}\mathbf{\Sigma}_Y^{-1}(y - y^\star), \end{aligned}$$

We recover the **Kalman filter's update**.

The (constrained) analysis map $\widetilde{T}_{y^\star}$ preserving the invariant $\mathsf{H}(x) = U_\perp^\top x$ reads

$$\begin{aligned} \widetilde{T}_{y^\star}(y, x) &= U_\perp x_\perp + U_{||} T_{y^\star}^{||}(y, x_\perp, x_{||}) \\ &= x - (I - U_\perp U_\perp^\top)\mathbf{\Sigma}_{X,Y}\mathbf{\Sigma}_Y^{-1}(y - y^\star). \end{aligned}$$

We recover a **projected formulation of the Kalman filter** (Simon, 2010).

Two equivalent treatments in the Gaussian case

For linear constraints in the Gaussian case:
Inference in rotated space = Projection of the Kalman's update.

A synthetic linear problem with an arbitrary number of invariants

Consider the linear dynamical model

$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} = \boldsymbol{A}_r\boldsymbol{x},\ \ \boldsymbol{x}(0) = \boldsymbol{x}_0,$$

where $\boldsymbol{A}_r$ has rank $n - r$ and eigendecomposition $\boldsymbol{A}_r = \boldsymbol{U}\boldsymbol{\Lambda}_r\boldsymbol{U}^{-1}$ where

$$\boldsymbol{\Lambda}_r = \mathrm{Diag}([\boldsymbol{0}_r, -\lambda_{r+1}, \ldots, -\lambda_n]),\ \ \text{with } \lambda_k > 0 \text{ for } k > r.$$

- $\mathrm{span}(\boldsymbol{U}[:, 1:r])$ is an invariant subspace of the dynamical system.
- Parametric study over the ratio of invariants r/n and the ensemble size M

Filter	Category	Preserve linear invariants
EnKF with tapering	Linear	✗
Constrained EnKF with tapering	Linear	✓

RMSE results for the synthetic linear problem

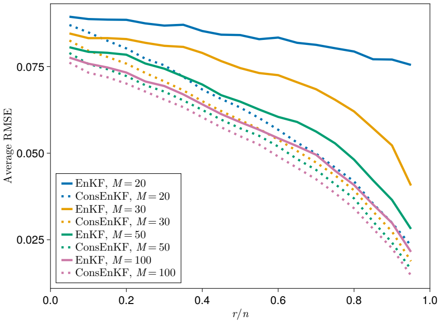


Figure 5: Evolution of the RMSE with the ratio r/n for the EnKF and the constrained EnKF (ConsEnKF) for $M = 20, 30, 50, 100$ samples.

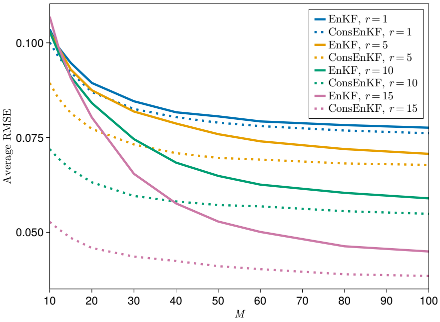


Figure 6: Evolution of the RMSE with the ensemble size M for the EnKF and the constrained EnKF (ConsEnKF) for $r = 1, 5, 10, 15$ linear invariants.

Takeway: Preserving linear invariants is most beneficial when M is small and r/n is large.

Linear advection equation

Consider the linear advection problem on the periodic domain $\Omega = [0, 1)$:

$$\frac{\partial u(s,t)}{\partial t} + \nabla \cdot (cu(s,t)) = 0, \qquad s \in \Omega,\ t > 0,$$
$$u(s,0) = u_0(s), \qquad s \in \Omega,$$

Discrete mass is preserved, i.e., $\boldsymbol{x} \to \boldsymbol{U}_\perp^\top \boldsymbol{x}$ with $\boldsymbol{U}_\perp = [1, \ldots, 1]^\top / \sqrt{n} \in \mathbb{R}^n$.

Filter	Category	Preserve linear invariants
EnKF with tapering	Linear	✗
Constrained EnKF with tapering	Linear	✓

From previous example, we don't expect much improvement on global tracking metrics (such as RMSE) for a small ratio r/n.

Evolution of the invariant $U_\perp^\top x_t$

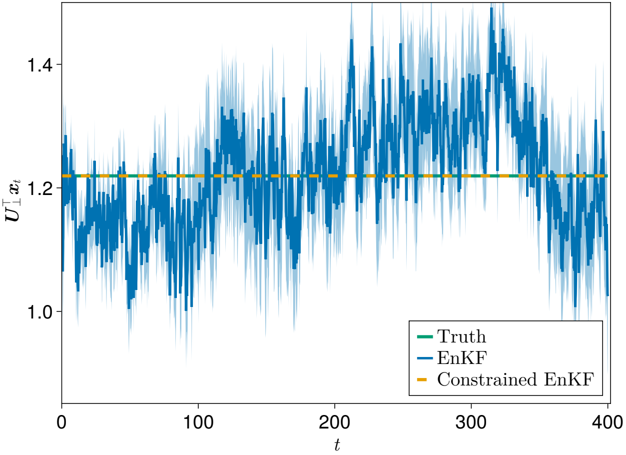


Figure 7: Time evolution of $U_\perp^\top x_t$ for the true state process (green) and the posterior mean obtained with the EnKF (blue) and the constrained EnKF (dashed yellow) for $M = 40$.

A embedded Lorenz-63 model

We embed the Lorenz-63 model in $\mathbb{R}^4$ to create a dynamical system with a linear invariant,i.e.,

$$\frac{\mathrm{d}\widetilde{\boldsymbol{x}}}{\mathrm{d}t} = \widetilde{\mathfrak{F}}(\widetilde{\boldsymbol{x}}, t) = \begin{bmatrix} \sigma(\widetilde{x}_2 - \widetilde{x}_1) \\ \widetilde{x}_1(\rho - \widetilde{x}_2) - \widetilde{x}_2 \\ \widetilde{x}_1\widetilde{x}_2 - \beta\widetilde{x}_3 \\ 0 \end{bmatrix},$$

where $\widetilde{x}_4$ has zero dynamic. We apply a random rotation $\boldsymbol{\Theta} \in O(4)$ to define $\boldsymbol{x} = \boldsymbol{\Theta}\widetilde{\boldsymbol{x}}$

$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} = \frac{\mathrm{d}\boldsymbol{\Theta}\widetilde{\boldsymbol{x}}}{\mathrm{d}t} = \boldsymbol{\Theta}\widetilde{\mathfrak{F}}(\boldsymbol{\Theta}^{-1}\boldsymbol{x}, t) = \boldsymbol{\Theta}\widetilde{\mathfrak{F}}(\boldsymbol{\Theta}^{\top}\boldsymbol{x}, t),$$

By construction, $\boldsymbol{x} \to \boldsymbol{U}_{\perp}^{\top}\boldsymbol{x}$ is preserved where $\boldsymbol{U}_{\perp} = \boldsymbol{\Theta}\boldsymbol{e}_4 \in \mathbb{R}^4$.

Details on the setting

We compare three filters:

- EnKF with optimal multiplicative inflation (OMI)
- A stochastic map filter (SMF) based on radial basis functions with OMI
- A constrained stochastic map filter with OMI

For this low-dimensional problem, tapering is not beneficial.

Filter	Category	Preserve linear ininvariants
EnKF	Linear	✓
SMF	Nonlinear	✗
Constrained SMF	Nonlinear	✓

Results for the embedded Lorenz-63 model

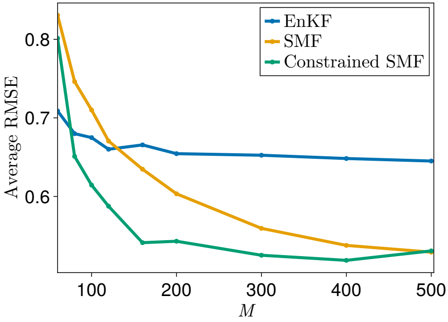


Figure 8: Evolution of the RMSE with the ensemble size M for the EnKF (blue), the SMF (yellow), and the constrained SMF (green).

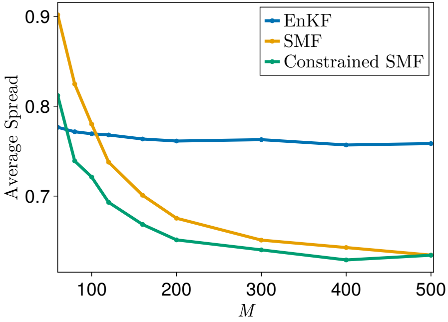


Figure 9: Evolution of the spread with the ensemble size M.

Takeway: Constrained SMF exploits structure + nonlinear update

Evolution of the invariant $U_{\perp}^{\top} x_t$

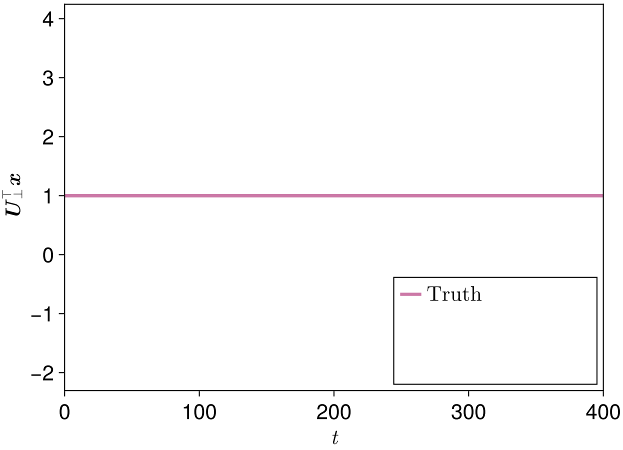


Figure 10: Evolution of $\boldsymbol{U}_{\perp}^{\top} \boldsymbol{x}_t$ for $M = 120$.

Evolution of the invariant $U_\perp^\top x_t$

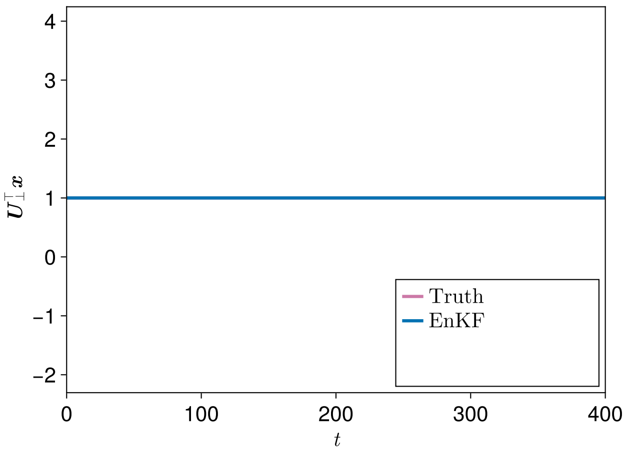


Figure 10: Evolution of $\boldsymbol{U}_\perp^\top \boldsymbol{x}_t$ for $M = 120$.

Evolution of the invariant $U_\perp^\top x_t$

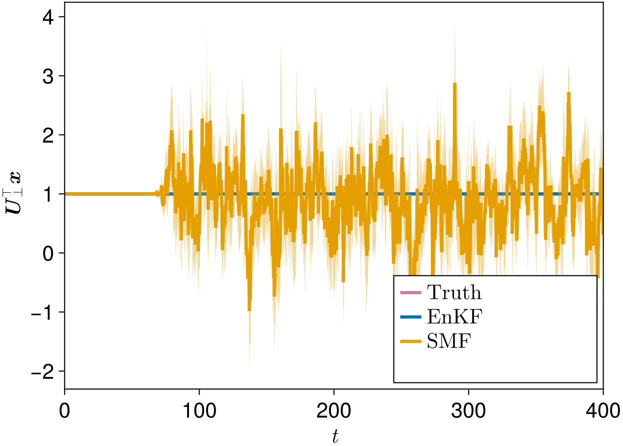


Figure 10: Evolution of $\boldsymbol{U}_\perp^\top \boldsymbol{x}_t$ for $M = 120$.

Evolution of the invariant $U_\perp^\top x_t$

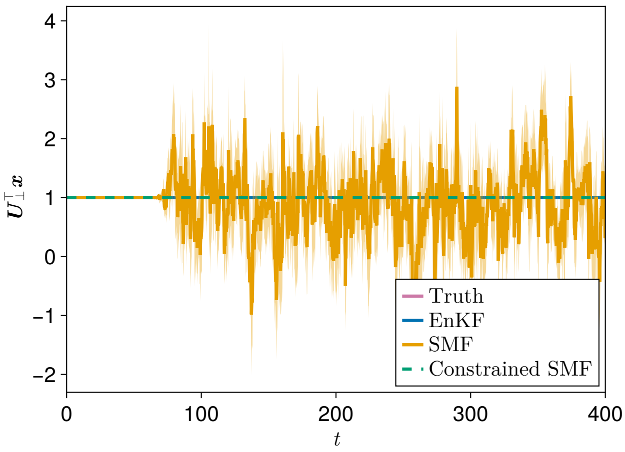


Figure 10: Evolution of $\boldsymbol{U}_\perp^\top \boldsymbol{x}_t$ for $M = 160$.

Future work and Acknowledgements

Summary:

- We introduced a class of linear invariant-preserving analysis maps for non-Gaussian filtering problems
- In the Gaussian case, we recovered a constrained formulation of the Kalman filter
- Assessed the benefits of preserving linear invariants for linear /nonlinear ensemble filters.

Future work:

- Extension to nonlinear invariants, e.g., Hamiltonian, energy, entropy
- Weak preservation of invariants in non-Gaussian settings

Main reference with Github repo:

Le Provost, M., Glaubitz, J., and Marzouk Y. (2024), **"Preserving linear invariants in ensemble filtering methods."**, arXiv:2404.14328

Funding: National Science Foundation (Grant PHY-2028125).

References

Baptista, R., Hosseini, B., Kovachki, N. B., and Marzouk, Y. (2020). Conditional sampling with monotone GANs: from generative models to likelihood-free inference. *arXiv preprint arXiv:2006.06755*.

Evensen, G. (1994). Sequential Data Assimilation with a Nonlinear Quasi-Geostrophic Model Using Monte Carlo Methods to Forecast Error Statistics. *Journal of Geophysical Research: Oceans*, 99(C5):10143–10162.

Le Provost, M., Baptista, R., Eldredge, J. D., and Marzouk, Y. (2023). An adaptive ensemble filter for heavy-tailed distributions: tuning-free inflation and localization. *arXiv preprint arXiv:2310.08741*.

Marzouk, Y., Moselhy, T., Parno, M., and Spantini, A. (2016). Sampling via measure transport: An introduction. *Handbook of Uncertainty Quantification*, 1:2.

Simon, D. (2010). Kalman filtering with state constraints: a survey of linear and nonlinear algorithms. *IET Control Theory & Applications*, 4(8):1303–1318.

Spantini, A., Baptista, R., and Marzouk, Y. (2022). Coupling techniques for nonlinear ensemble filtering. *SIAM Review*, 64(4):921–953.