Toward consistent nonlinear filtering and smoothing via measure transport

Ricardo Baptista

Joint work with Alessio Spantini, Youssef Marzouk, Max Ramgraber, Mathieu Le Provost

1Computing + Mathematical Sciences
California Institute of Technology

2Center for Computational Science and Engineering
Massachusetts Institute of Technology

3Department of Computer Science
Long Island University

NCAR CISL Seminar

July 6, 2023
Sequential inference is ubiquitous

- **Goal**: Sequential state estimation in a Bayesian setting
- **Applications**: Weather prediction, oceanography, finance, population dynamics, pharmacology, robotics, aerodynamics, etc.

Wind forecast
Sequential inference is ubiquitous

- **Goal**: Sequential state estimation in a Bayesian setting
- **Applications**: Weather prediction, oceanography, finance, population dynamics, pharmacology, robotics, aerodynamics, etc.

Wind forecast

Epidemiological forecast
Sequential inference is ubiquitous

- **Goal**: Sequential state estimation in a Bayesian setting
- **Applications**: Weather prediction, oceanography, finance, population dynamics, pharmacology, robotics, aerodynamics, etc.

Vortex shedding around an aircraft wing

Baptista (rsb@caltech.edu)
Sequential inference is ubiquitous

- **Goal**: Sequential state estimation in a Bayesian setting
- **Applications**: Weather prediction, oceanography, finance, population dynamics, pharmacology, robotics, aerodynamics, etc.

Vortex shedding around an aircraft wing

Baptista (rsb@caltech.edu)
**Goal**: Sequential state estimation in a Bayesian setting

**Applications**: Weather prediction, oceanography, finance, population dynamics, pharmacology, robotics, aerodynamics, etc.

Vortex shedding around an aircraft wing
Sequential inference is ubiquitous

- **Goal**: Sequential state estimation in a Bayesian setting

- **Applications**: Weather prediction, oceanography, finance, population dynamics, pharmacology, robotics, aerodynamics, etc.

Vortex shedding around an aircraft wing
Sequential inference is ubiquitous

- **Goal**: Sequential state estimation in a Bayesian setting
- **Applications**: Weather prediction, oceanography, finance, population dynamics, pharmacology, robotics, aerodynamics, etc.

Vortex shedding around an aircraft wing

Baptista (rsb@caltech.edu)
Sequential inference is ubiquitous

- **Goal**: Sequential state estimation in a Bayesian setting
- **Applications**: Weather prediction, oceanography, finance, population dynamics, pharmacology, robotics, aerodynamics, etc.

Vortex shedding around an aircraft wing
Sequential inference is ubiquitous

- **Goal**: Sequential state estimation in a Bayesian setting
- **Applications**: Weather prediction, oceanography, finance, population dynamics, pharmacology, robotics, aerodynamics, etc.

Vortex shedding around an aircraft wing
Non-Gaussianity is ubiquitous

- Non-Gaussianity can include multi-modality and/or tail-heaviness

Lorenz-63 smoothing ensemble

\[(X, Y)\) distribution in additive manufacturing model [B et al., 2022]
Non-Gaussianity is ubiquitous

- Non-Gaussianity can include multi-modality and/or tail-heaviness

Lorenz-63 smoothing ensemble

\((\mathbf{X}, \mathbf{Y})\) distribution in additive manufacturing model [B et al., 2022]
Non-Gaussianity is ubiquitous

- Non-Gaussianity can include multi-modality and/or tail-heaviness

Lorenz-63 smoothing ensemble

\( (X, Y) \) distribution in additive manufacturing model [B et al., 2022]

**Takeaway**: Gaussian approximations under-predict data informativeness
Non-Gaussianity is ubiquitous

- Non-Gaussianity can include multi-modality and/or tail-heaviness

**Lorenz-63 smoothing ensemble**

\((X, Y)\) distribution in additive manufacturing model [B et al., 2022]

**Takeaway**: Gaussian approximations under-predict data informativeness

**Goal**: Develop consistent inference methods for non-Gaussian problems
State-space models

- States follow model dynamics $\pi_{X_t \mid X_{t-1}}$
- Observations follow likelihood function $\pi_{Y_t \mid X_t}$

Goal: Recursively sample distributions $\pi_{X_t \mid y_1^*, \ldots, y_t^*}$ or $\pi_{X_{1:t} \mid y_1^*, \ldots, y_t^*}$
Sequential Bayesian inference

State-space models

- States follow model dynamics $\pi_{X_t|X_{t-1}}$
- Observations follow likelihood function $\pi_{Y_t|X_t}$

Goal: Recursively sample distributions $\pi_{X_t|y_1^*,...,y_t^*}$ or $\pi_{X_{1:t}|y_1^*,...,y_t^*}$

Common challenges leading to non-Gaussianity

- Nonlinear dynamical models or observation operators
- Sparse observations in space and time


**Approach**: Approximate distributions using limited samples

- *Forecast step*

\[
\pi_{X_{t-1}|y_1^*, \ldots, y_{t-1}^*}
\]

- *Analysis step*

\[
\pi_{X_t|y_1^*, \ldots, y_{t-1}^*}
\]

Bayesian inference
**Ensemble filtering and smoothing**

**Approach:** Approximate distributions using limited samples

\[
\pi_{X_{t-1}|y_1^*,\ldots,y_{t-1}^*}
\]

**forecast step**

\[
\pi_{X_{t}|y_1^*,\ldots,y_{t-1}^*}
\]

**analysis step**

\[
\pi_{X_{t}|y_1^*,\ldots,y_{t}^*}
\]

Bayesian inference
**Approach:** Approximate distributions using limited samples

- **Forecast step:**
  \[ \pi X_{t-1} | y^*_1, \ldots, y^*_{t-1} \]

- **Analysis step:**
  \[ \pi X_t | y^*_1, \ldots, y^*_{t-1} \]
  \[ \pi X_t | y^*_1, \ldots, y^*_t \]

---

**Ensemble Kalman filters and smoothers**

- Analysis step updates particles by estimating a **linear transformation**
- **Inconsistent** for capturing Bayesian solution

---

**Bayesian inference**
**Ensemble filtering and smoothing**

**Approach:** Approximate distributions using limited samples

**Bayesian inference**

\[ \pi_{X_{t-1}|y_1^*,\ldots,y_{t-1}^*} \]

\[ \pi_{X_t|y_1^*,\ldots,y_{t-1}^*} \]

\[ \pi_{X_t|y_1^*,\ldots,y_t^*} \]

**Ensemble Kalman filters and smoothers**

- Analysis step updates particles by estimating a linear transformation
- **Inconsistent** for capturing Bayesian solution

**Goal:** Perform analysis **consistently and robustly** in non-Gaussian settings
Idea: Find map $T$ that take samples from prior $\pi_X$ to posterior $\pi_{X|Y}$.
**Prior-to-posterior transformations**

**Idea:** Find map $T$ that take samples from prior $\pi_X$ to posterior $\pi_{X|Y}$

![Diagram showing prior distribution $\pi_X$ and posterior distribution $\pi_{X|Y}$, with a sample $x^i$ transformed by $T(x^i)$ to the posterior distribution.]

**Plan for this talk:**

1. Maps for filtering $X = X_t$?
Idea: Find map $T$ that take samples from prior $\pi_X$ to posterior $\pi_{X|Y}$

Plan for this talk:
1. Maps for filtering $X = X_t$?
2. Maps for smoothing $X = X_{1:t}$?
**Idea:** Find map $T$ that take samples from prior $\pi_X$ to posterior $\pi_{X|Y}$

$$T(x^i)$$

**Plan for this talk:**

1. Maps for filtering $X = X_t$?
2. Maps for smoothing $X = X_{1:t}$?
3. Leveraging structure in $T$ to tackle high-dimensional problems
Transport maps characterize distributions

- Transport map $S$ induces a deterministic coupling between a target density $\pi$ and a reference density $\eta$ (e.g., standard normal)
  - Generate cheap and independent samples: $x \sim \pi \iff S(x) \sim \eta$
  - Evaluate the target density: $\pi(x) = S^\# \eta(x) := \eta \circ S(x) | \text{det} \nabla S(x) |$

![Diagram showing transport maps and their effects on samples and densities](#)
Transport maps characterize distributions

- **Transport map** $S$ induces a deterministic coupling between a target density $\pi$ and a reference density $\eta$ (e.g., standard normal)
  - Generate cheap and independent samples: $z \sim \eta \iff S^{-1}(z) \sim \pi$
  - Evaluate the target density: $\pi(x) = S^\#\eta(x) := \eta \circ S(x)|\det \nabla S(x)|$
Monotone triangular maps

As a building block, consider the **Knothe-Rosenblatt rearrangement**

\[
S(x) = \begin{bmatrix}
  S_1(x_1) \\
  S_2(x_1, x_2) \\
  \vdots \\
  S_d(x_1, x_2, \ldots, x_d)
\end{bmatrix}
\]

1. Unique under mild assumptions on \(\pi\) and \(\eta\)
2. Invertibility is guaranteed by one-dimensional monotonicity \(\frac{\partial k}{\partial S_k} > 0\)
3. \(S^{-1}(z)\) and \(\det \nabla S(x)\) are simple to evaluate
4. Each component \(S_k\) characterizes one marginal conditional \(\pi_{X_k} = \pi_{X_k \mid X_{k-1}, \ldots, X_1} \cdots \pi_{X_k \mid X_1} \cdots \pi_{X_k \mid X_1, \ldots, X_{d-1}}\)
Monotone triangular maps

As a building block, consider the Knothe-Rosenblatt rearrangement

\[
S(x) = \begin{bmatrix}
S_1(x_1) \\
S_2(x_1, x_2) \\
\vdots \\
S_d(x_1, x_2, \ldots, x_d)
\end{bmatrix}
\]

1. **Unique** under mild assumptions on \( \pi \) and \( \eta \)
2. Invertibility is guaranteed by **one-dimensional monotonicity** \( \partial_k S_k > 0 \)
3. \( S^{-1}(z) \) and \( \det \nabla S(x) \) are simple to evaluate
4. Each component \( S_k \) characterizes one **marginal conditional**

\[
\pi_x = \pi_{x_1} \pi_{x_2|x_1} \cdots \pi_{x_d|x_1,\ldots,x_{d-1}}
\]
Learning expressive triangular maps from samples

Given target density $\pi$ and standard Gaussian $\eta$,

$$\min_S D_{KL}(\pi||S^\# \eta)$$

$$\Leftrightarrow \min_{\{s: \partial_k s > 0\}} \mathbb{E}_\pi \left[ \frac{1}{2} s(x_{1:k})^2 - \log |\partial_k s(x_{1:k})| \right] \forall k$$

Target density approximation: $\hat{\pi}(x) := \hat{S}^\# \eta(x)$.
Given target density $\pi$ and standard Gaussian $\eta$, 

$$\min_S D_{KL}(\pi||S^\# \eta) \iff \min_{\{s: \partial_k s > 0\}} \mathbb{E}_\pi \left[ \frac{1}{2} s(x_{1:k})^2 - \log |\partial_k s(x_{1:k})| \right] \forall k$$

Given samples $\{x^i\}_{i=1}^n \sim \pi$, find $\hat{S}_k$ via

$$\arg \min_{\{s: \partial_k s > 0\}} \frac{1}{n} \sum_{i=1}^n \left[ \frac{1}{2} s(x^i_{1:k})^2 - \log |\partial_k s(x^i_{1:k})| \right]$$

**Target density approximation:** $\hat{\pi}(x) := \hat{S}^\# \eta(x)$
Consider the triangular map pushing forward $\pi_{Y,X}$ to $\eta_{Z_1,Z_2}$:

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

- $S^Y$ pushes forward $\pi_Y$ to $\eta_{Z_1}$
- $S^X(y,\cdot)$ pushes forward $\pi_{X|y}$ to $\eta_{Z_2}$ for any $y$
Triangular maps enable conditional sampling

Consider the triangular map pushing forward $\pi_{Y,X}$ to $\eta_{Z_1,Z_2}$:

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

- $S^Y$ pushes forward $\pi_Y$ to $\eta_{Z_1}$
- $S^X(y, \cdot)$ pushes forward $\pi_{X|y}$ to $\eta_{Z_2}$ for any $y$

**Recipe for amortized inference:**
To characterize posterior $\pi_{X|y^*} \propto \pi_{Y^*|X} \pi_X$ given an observation $y^*$:

- Simulate from the model: $x^i \sim \pi_X$, $y^i \sim \pi_{Y|x^i}$
- Estimate $S^X$ from joint samples $(x^i, y^i) \sim \pi_{X,Y}$
- Simulate $\tilde{S}^X(y^*, \cdot)^{-1}|_{z^i} \sim \pi_{X|y^*}$ for $z^i \sim \eta_{Z_2}$

**Related Work:** Simulation-based or likelihood-free inference [Papamakarios & Murray, 2016; Lueckmann et al., 2017; Greenberg et al., 2019]
Numerical example: image in-painting [Kovachki, B, et al., 2021]

- **Goal**: Reconstruct image after removing its center section
- Use map to sample from the conditional distribution for the $14 \times 14$ center pixels of a $28 \times 28$ MNIST handwritten digit
- Estimate conditional mean and variance and classify digit probability

Note: Prior distributions in imaging problems have no analytic form
Will this always work well?

**Lorenz-63 model**

- Infer the hidden state given noisy point-wise observations
- With $N = 50$ samples, we can *at best* estimate linear maps
- Measure root-mean-squared error (RMSE) of ensemble mean

![Graph showing RMSE over time and number of training samples]

**Takeaway**: This approach yields large errors with limited samples.
Will this always work well?

**Lorenz-63 model**

- Infer the hidden state given noisy point-wise observations
- With $N = 50$ samples, we can *at best* estimate linear maps
- Measure root-mean-squared error (RMSE) of ensemble mean

**Takeaway**: This approach yields large errors with limited samples
Another approach: compose maps for sampling

For $\pi_{Y,X}$ and $\eta_{Z_1,Z_2}$, consider the triangular map

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

- $S^X(y, \cdot)$ pushes forward $\pi_{X|Y}$ to $\eta_{Z_2}$ for any $y$
- $S^X(y, x)$ pushes forward $\pi_{Y,X}$ to $\eta_{Z_2}$
Another approach: compose maps for sampling

For $\pi_{Y,X}$ and $\eta_{Z_1,Z_2}$, consider the triangular map

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

- $S^X(y, \cdot)$ pushes forward $\pi_{X|y}$ to $\eta_{Z_2}$ for any $y$
- $S^X(y, x)$ pushes forward $\pi_{Y,X}$ to $\eta_{Z_2}$
Another approach: compose maps for sampling

For $\pi_{Y,X}$ and $\eta_{Z_1,Z_2}$, consider the triangular map

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

- $S^X(y, \cdot)$ pushes forward $\pi_{X|Y}$ to $\eta_{Z_2}$ for any $y$
- $S^X(y, x)$ pushes forward $\pi_{Y,X}$ to $\eta_{Z_2}$
Another approach: compose maps for sampling

The prior-to-posterior map that pushes $\pi_{Y,X}$ to $\pi_{X|y^*}$ is

$$T_{y^*}(y, x) = S^X(y^*, \cdot)^{-1} \circ S^X(y, x)$$
Another approach: compose maps for sampling

The prior-to-posterior map that pushes $\pi_{Y,X}$ to $\pi_{X|y^*}$ is

$$T_{y^*}(y, x) = S_{X}(y^*, \cdot)^{-1} \circ S_{X}(y, x)$$

**Stochastic map algorithm:**

1. Estimate $S_{X}$ using $(y^i, x^i) \sim \pi_{Y,X}$
2. Evaluate composed map $T_{y^*}(y, x)$ to approximately sample posterior
Stochastic map algorithm for filtering

Forecast step
1. Apply dynamics to generate forecast ensemble \((x^f_t)^i \sim \pi_{x_t|x_{t-1}}\)

Analysis step
1. Sample observations \(y^i_t \sim \pi_{Y_t|(x^f_t)^i}\) using forecast samples
2. Estimate lower-triangular map \(S\) that couples \(\pi_{Y_t,x_t}\) and \(N(0, I)\)
   \[
   S(y_t, x_t) = \begin{bmatrix}
   S^Y(y_t) \\
   S^X(y_t, x_t)
   \end{bmatrix}
   \]
3. Compose maps \(T^*_y(y_t, x_t) = S^X(y^*_t, \cdot)^{-1} \circ S^X(y_t, x_t)\)
4. Generate analysis ensemble \(x^i_t = T^*_y(y^i_t, x^i_t)\) for \(i = 1, \ldots, N\)
Composed maps are stable for tracking

**Lorenz-63 model**

- Infer the hidden state given noisy point-wise observations
- With $N = 50$ samples, we can *at best* estimate linear maps
- Measure root-mean-squared error (RMSE) of ensemble mean

**Takeaway**: Composed maps have stable RMSE with limited samples
Composed maps are stable for tracking

**Lorenz-63 model**

- Infer the hidden state given noisy point-wise observations
- With $N = 50$ samples, we can *at best* estimate linear maps
- Measure root-mean-squared error (RMSE) of ensemble mean

**Takeaway**: Composed maps have **stable RMSE with limited samples**
Numerical details of the stochastic map algorithm

Generalization of the EnKF

- Restricting $S^X$ to be affine in $x_t, y_t$, we recover the transformation

\[ T_{y_t^*}(y_t, x_t) = x_t - \sum_{x_t, y_t} \sum_{y_t^-1}(y_t - y_t^*), \]

- Transport maps allow for the gradual introduction of nonlinear terms
- Nonlinear maps $T_{y_t^*}$ capture non-Gaussian structure of $\pi_{Y_t, X_t}$
Numerical details of the stochastic map algorithm

Generalization of the EnKF
- Restricting $S^X$ to be affine in $x_t, y_t$, we recover the transformation

$$T_{y^*} (y_t, x_t) = x_t - \sum_{x_t, y_t} \Sigma_{y_t}^{-1} (y_t - y^*_t),$$

- Transport maps allow for the gradual introduction of nonlinear terms
- Nonlinear maps $T_{y^*_t}$ capture non-Gaussian structure of $\pi_{Y_t, X_t}$

Example map parameterization
- Each component is the sum of nonlinear univariate functions

$$S_k(z_1, \ldots, z_k) = u_1(z_1) + \cdots + u_k(z_k),$$

where $u_i(z) = u_{i,0} z + \sum_{j=1}^{p} u_{ij} \mathcal{N}(z; \xi_j, \sigma_j^2)$ and $u_k(z_k)$ is monotone.
Nonlinear maps capture filtering distribution

Lorenz-63 model

- $d = 3$ with $\Delta t_{obs} = 0.1$ and fully-observed state
- Observations follow $y_t = x_t + \eta_t$ with $\eta_t \sim \mathcal{N}(0, 4I)$
- Measure root-mean-squared-error $RMSE(t) = \|x_t^* - \mathbb{E}[x_t|y_{1:t}^*]\|_2^2 / \sqrt{d}$
- Compare statistics to a particle filter (PF) with 1M samples

Improved posterior estimates is also stable with increasing $\Delta t_{obs}$
Nonlinear maps improve tracking

Lorenz-96 model: chaotic dynamics

- 40 states, 20 observations, and $\Delta t_{obs} = 0.4$ (large!)
- Measure average RMSE (left) over 2000 assimilation cycles
- Parametrize maps with increasing nonlinearity using RBFs

Nonlinear maps also improve estimates of posterior moments
Nonlinear maps better capture uncertainty in true state

- Tracking two marginals of Lorenz-96 system at two assimilation times
- Compare ensemble distribution from EnKF and nonlinear maps
Nonlinear maps better capture uncertainty in true state

- Tracking two marginals of Lorenz-96 system at two assimilation times
- Compare ensemble distribution from EnKF and nonlinear maps
Extension to smoothing

**Goal:** Characterize full smoothing distribution $\pi_{X_1:T|y_1:T}$ or a marginal

- Consider update for all states given a single observation at time $T$

![Diagram showing states $X_1, X_2, \ldots, X_{T-1}, X_T$ and observation $Y_T$.]

**Ensemble Transport Smoother:** Apply stochastic map algorithm on joint states over time:

$$T_{y_T^*}(y_T, x_{1:T}) = S^X(y_T^*, \cdot)^{-1} \circ S^X(y_T, x_{1:T})$$

- Ordering of states in $S^X$ defines different smoothing algorithms
- Exploiting the Markov structure of the states yields sparse maps
Transport maps exploit conditional independence.

Theorem: Sparsity of triangular maps [Spantini et al., 2018]

Conditional independence of target distribution $\pi$ (encoded by graph) defines functional dependence of $S$ such that $S^{\#} \eta = \pi$

Markov structure of 5-dimensional distribution

Markov structure of hidden Markov model

Sparsity of $\partial_j S_k$

Sparsity of $\partial_j S_k$
Transport maps exploit conditional independence

Theorem: Sparsity of triangular maps [Spantini et al., 2018]

Conditional independence of target distribution $\pi$ (encoded by graph) defines functional dependence of $S$ such that $S^\# \eta = \pi$

$$
\begin{bmatrix}
S_1(x_1) \\
S_2(x_1, x_2) \\
S_3(x_1, x_2, x_3) \\
S_4(x_1, x_2, x_3, x_4)
\end{bmatrix} \rightarrow 
\begin{align*}
\pi(x_1) \\
\pi(x_2|x_1) \\
\pi(x_3|x_1, x_2) = \pi(x_3|x_2) & \quad X_3 \perp \!\!\!\perp X_1 | X_2 \\
\pi(x_4|x_1, x_2, x_3) = \pi(x_4|x_3) & \quad X_4 \perp \!\!\!\perp (X_1, X_2) | X_3
\end{align*}
$$

$X_1$ $X_2$ $X_3$ $X_4$
Two new classes of smoothers [Ramgraber, B et al., 2022]

**Backwards-in-time**: uses the ordering $\mathbf{x}_T, \ldots, \mathbf{x}_1$

$$S^\mathcal{X}(\mathbf{y}_T, \mathbf{x}_{1:T}) \overset{\text{CI}}{=} \begin{bmatrix} S_T(\mathbf{y}_T, \mathbf{x}_T) \\ S_{T-1}(\mathbf{x}_T, \mathbf{x}_{T-1}) \\ \vdots \\ S_1(\mathbf{x}_2, \mathbf{x}_1) \end{bmatrix}$$

(\text{CI}) exploits chain structure: $\mathbf{x}_{1:T-1} \perp \perp \mathbf{y}_T | \mathbf{x}_T$ and $\mathbf{x}_{1:s-1} \perp \perp \mathbf{x}_{s+1:T} | \mathbf{x}_s$
Two new classes of smoothers [Ramgraber, B et al., 2022]

**Backwards-in-time:** uses the ordering $x_T, \ldots, x_1$

$$S^x(y_T, x_{1:T})^{CI} = \begin{bmatrix}
S_T(y_T, x_T) \\
S_{T-1}(x_T, x_{T-1}) \\
\vdots \\
S_1(x_2, x_1)
\end{bmatrix}$$

(CI) exploits chain structure: $x_{1:T-1} \perp \perp y_T | x_T$ and $x_{1:s-1} \perp \perp x_{s+1:T} | x_s$

**Forwards-in-time:** uses the ordering $x_1, \ldots, x_T$

$$S^x(y_T, x_{1:T})^{CI} = \begin{bmatrix}
S_1(y_T, x_1) \\
S_2(y_T, x_1, x_2) \\
\vdots \\
S_T(y_T, x_{T-1}, x_T)
\end{bmatrix}$$

(CI) exploits chain structure: $x_s \perp \perp x_{1:s-2} | x_{s-1}$ for $s \geq 2$
Two new classes of smoothers [Ramgraber, B et al., 2022]

**Backwards-in-time**: uses the ordering \( \mathbf{x}_T, \ldots, \mathbf{x}_1 \)

\[
S^{\mathcal{X}}(\mathbf{y}_T, \mathbf{x}_{1:T}) \equiv \begin{bmatrix} S_T(\mathbf{y}_T, \mathbf{x}_T) \\ S_{T-1}(\mathbf{x}_T, \mathbf{x}_{T-1}) \\ \vdots \\ S_1(\mathbf{x}_2, \mathbf{x}_1) \end{bmatrix}
\]

(CI) exploits chain structure: \( \mathbf{x}_{1:T-1} \perp \perp \mathbf{y}_T | \mathbf{x}_T \) and \( \mathbf{x}_{1:s-1} \perp \perp \mathbf{x}_{s+1:T} | \mathbf{x}_s \)

**Forwards-in-time**: uses the ordering \( \mathbf{x}_1, \ldots, \mathbf{x}_T \)

\[
S^{\mathcal{X}}(\mathbf{y}_T, \mathbf{x}_{1:T}) \equiv \begin{bmatrix} S_1(\mathbf{y}_T, \mathbf{x}_1) \\ S_2(\mathbf{y}_T, \mathbf{x}_1, \mathbf{x}_2) \\ \vdots \\ S_T(\mathbf{y}_T, \mathbf{x}_{T-1}, \mathbf{x}_T) \end{bmatrix}
\]

(CI) exploits chain structure: \( \mathbf{x}_s \perp \perp \mathbf{x}_{1:s-2} | \mathbf{x}_{s-1} \) for \( s \geq 2 \)

- Empirical results suggest backward-in-time accumulates less errors
- Forwards smoother constrains state trajectories by dynamics
Focusing on backwards smoother

**Sequential context:** The joint decomposition simplifies

\[
\pi(x_1:T | y_1^*:T) = \pi(x_T | y_1^*:T) \prod_{s=1}^{T-1} \pi(x_s | x_{s+1}, y_1^*:T)
\]

\[
= \pi(x_T | y_1^*:T) \prod_{s=1}^{T-1} \pi(x_s | x_{s+1}, y_1^*:s)
\]

- Component \(S_s\) samples \(\pi(x_s | x_{s+1}, y_1^*:s)\)
- We estimate \(S_s\) using filtering ensemble \((x^i_s, x^i_{s+1}) \sim \pi(x_s, x_{s+1} | y_1^*:s)\)
Focusing on backwards smoother

**Sequential context:** The joint decomposition simplifies

\[
\pi(x_{1:T} | y_{1:T}^*) = \pi(x_T | y_{1:T}^*) \prod_{s=1}^{T-1} \pi(x_s | x_{s+1}, y_{1:T}^*)
\]

\[
= \pi(x_T | y_{1:T}^*) \prod_{s=1}^{T-1} \pi(x_s | x_{s+1}, y_{1:s}^*)
\]

▶ Component \( S_s \) samples \( \pi(x_s | x_{s+1}, y_{1:s}^*) \)

▶ We estimate \( S_s \) using filtering ensemble \( (x_s^i, x_{s+1}^i) \sim \pi(x_s, x_{s+1} | y_{1:s}^*) \)

**Generalization of the Ensemble RTS smoother**

▶ Restricting \( S^X \) to be affine in \( y_t, x_{1:t} \), we recover the transformation

\[
T_{y_T^*}(y_T, x_T) = x_T - \Sigma_{x_T,y_T} \Sigma_{y_T}^{-1} (y_T - y_T^*)
\]

\[
T_{x_{s+1}^*}(x_s, x_{s+1}) = x_s - \Sigma_{x_s,x_{s+1}} \Sigma_{x_{s+1}}^{-1} (x_{s+1} - x_{s+1}^*), \quad s < t
\]

**Takeaway:** Non-linear transport maps generalize linear smoothers
Nonlinear smoothers capture bimodal distributions

- Sinusoidal state $x_t$ with observation $y_t = |x_t + \gamma|$ for $\gamma \sim \mathcal{N}(0, 0.1)$
- Infer state using random walk model without knowing true dynamics
- Backward smoother is initialized from nonlinear transport filter
Nonlinear smoothers improve state estimation

Lorenz-63 model

A: Lorenz-63 EnTF and EnTS results

B: Lorenz-63 iEnKS results

Baptista (rsb@caltech.edu)
So far: Transport maps are consistent for sampling non-Gaussian filtering and smoothing distributions without requiring importance weights
Tackling high-dimensional inference problems

So far: Transport maps are consistent for sampling non-Gaussian filtering and smoothing distributions without requiring importance weights.

How do we compute transport maps given small ensemble sizes?

1. Localize estimators with approximate Markov structure
2. Targeted non-linearity using hybrid nonlinear+linear maps
3. Restrict inference to relevant low-dimensional subspaces
Many spatial fields satisfy approximate Markov properties

Inverse covariance matrix for Lorenz-96 model forecast is **sparse**
1. Transport maps are easy to “localize” in high dimensions

Many spatial fields satisfy approximate Markov properties

Inverse covariance matrix for Lorenz-96 model forecast is \textit{sparse}
Many spatial fields satisfy approximate Markov properties

**Idea:** Regularize the estimation of $S$ by imposing sparsity:

$$
\hat{S}(x) = 
\begin{bmatrix}
\hat{S}^1(x_1) \\
\hat{S}^2(x_1, x_2) \\
\hat{S}^3(\ldots, x_2, x_3) \\
\hat{S}^4(\ldots, x_3, x_4)
\end{bmatrix}
$$

**Heuristic:** Let $\hat{S}^k$ depend on neighboring variables $(x_j)_{j<k}$ that are physically close to $x_k$:

$$
\hat{S}^k(x_1, \ldots, x_k) \approx \hat{S}^k(x_{N(k)}, x_k)
$$
2. Structured hybrid linear and nonlinear maps

**Local-likelihood models**: Scalar observation $y \sim \pi_{Y|x_1}$

\[
T(y, x) = \begin{bmatrix}
T_1(y, x_1) \\
\vdots \\
T_l(x_1, \ldots, x_l) \\
L_{l+1}(x_1, \ldots, x_{l+1}) \\
\vdots \\
L_d(x_1, \ldots, x_d)
\end{bmatrix}
\]

**Idea**: For conditionally Gaussian models, use nonlinear updates $T_k$ for state variables $x_{1:l}$ and use linear updates $L_k$ for $x_{l+1:d}$
2. Structured hybrid linear and nonlinear maps

**Local-likelihood models:** Scalar observation $y \sim \pi_{Y|X_1}$

**Idea:** For conditionally Gaussian models, use nonlinear updates $T_k$ for state variables $x_{1:l}$ and use linear updates $L_k$ for $x_{l+1:d}$

**Special cases:**

- $l = 1$: Nonlinear $T_1$ and keeping all other components affine is related to the rank histogram filter [Andersen 2010]
- With decay in correlation, $L_{l+1}, \ldots, L_d$ reverts to an identity map
3. Low-rank updates via an example in turbulent flows

Inference problem:
- States $x_t$: Positions and strengths of point vortices
- Observations $y_t$: Pressure observations along airfoil

Challenges:
- High-dimensional states and observations $d = 180$ and $m = 50$
- Observations are non-local: $y_t$ is related to all $x_t$ by Poisson equation
- Limited ensemble of size $N = \mathcal{O}(100)$
Low-rank stochastic map filter

Main ideas

- Only part of the state $x_r = U_r^T x$ is informed by the observations
- Only part of the observation $y_s = V_s^T y$ is relevant to the states
## Main ideas

- Only part of the state $\mathbf{x}_r = U_r^T \mathbf{x}$ is informed by the observations
- Only part of the observation $\mathbf{y}_s = V_s^T \mathbf{y}$ is relevant to the states

- Consider the posterior approximation at each assimilation step

$$
\hat{\pi}_{\mathbf{x}|\mathbf{y}}(\mathbf{x}|\mathbf{y}) = \hat{\pi}_{\mathbf{x}_r|\mathbf{y}_s}(\mathbf{x}_r|\mathbf{y}_s) \pi_{\mathbf{x}_\perp|x_r}(\mathbf{x}_\perp|x_r)
$$

- **Approach:** Find $U_r, V_s$ with small $r$ and $s$ from prior ensemble and observation operator such that $\pi_{\mathbf{x}|\mathbf{y}} \approx \hat{\pi}_{\mathbf{x}|\mathbf{y}}$ [B, Marzouk et al., 2022]
Low-rank stochastic map filter

Main ideas

- Only part of the state \( x_r = U_r^T x \) is informed by the observations
- Only part of the observation \( y_s = V_s^T y \) is relevant to the states

- Consider the posterior approximation at each assimilation step

\[
\hat{\pi}_{X|Y}(x|y) = \hat{\pi}_{X_r|Y_s}(x_r|y_s)\pi_{X_\perp|x_r}(x_\perp|x_r)
\]

- **Approach:** Find \( U_r, V_s \) with small \( r \) and \( s \) from prior ensemble and observation operator such that \( \pi_{X|Y} \approx \hat{\pi}_{X|Y} \) [B, Marzouk et al., 2022]

- **Result:** Prior-to-posterior map only acts on low-dimensional variables

\[
T_{y^*}(y, x) = U_r T_{y^*_s}(V_s^T y, U_r^T x) + U_\perp U_\perp^T x
\]

- \( T_r \) can be linear [Le Provost, B et al., 2022] or non-linear
Low-rank filter is stable for small ensemble sizes

Observations:
- RMSE is stable for small $N$ for different energy ratios
- Adaptive reduced dimensions $r, s$ do not increase over time
Low-rank EnkF is stable with model error

High-fidelity numerical simulation at Reynolds number 1000

Baptista (rsb@caltech.edu)
Low-rank EnkF is stable with model error

High-fidelity numerical simulation at Reynolds number 1000

Inviscid vortex model with EnKF
Low-rank EnkF is stable with model error

High-fidelity numerical simulation at Reynolds number 1000

Inviscid vortex model with EnKF

Inviscid vortex model with LR-EnKF
Conclusions and outlook

Central idea: consistent data assimilation using measure transport

- Composed transport maps generalize ensemble filters and smoothers
- Nonlinear maps improve state estimation for chaotic systems
- Exploit (approximate) conditional independence structure for scaling to high-dimensional inference problems

Ongoing work

- Square-root versions of nonlinear filters and smoothers
- Connections to other nonlinear filters, e.g., conjugate transform filter


Thank You

Supported by the U.S. Department of Energy and NSERC

Baptista (rsb@caltech.edu) Nonlinear ensemble filtering & smoothing
Central idea: consistent data assimilation using measure transport

- **Composed transport maps** generalize ensemble filters and smoothers
- Nonlinear maps **improve state estimation** for chaotic systems
- **Exploit (approximate) conditional independence structure** for scaling to high-dimensional inference problems

Thank You
Conclusions and outlook

Central idea: consistent data assimilation using measure transport

- **Composed transport maps** generalize ensemble filters and smoothers
- Nonlinear maps **improve state estimation** for chaotic systems
- **Exploit (approximate) conditional independence structure** for scaling to high-dimensional inference problems

Ongoing work

- Square-root versions of nonlinear filters and smoothers
- Connections to other nonlinear filters, e.g., conjugate transform filter [Chipilski 2023]


Thank You

Supported by the U.S. Department of Energy and NSERC

Baptista (rsb@caltech.edu)  Nonlinear ensemble filtering & smoothing
Conclusions and outlook

Central idea: consistent data assimilation using measure transport

- **Composed transport maps** generalize ensemble filters and smoothers
- Nonlinear maps **improve state estimation** for chaotic systems
- **Exploit (approximate) conditional independence structure** for scaling to high-dimensional inference problems

Ongoing work

- Square-root versions of nonlinear filters and smoothers
- Connections to other nonlinear filters, e.g., conjugate transform filter [Chipilski 2023]


Thank You

Supported by the U.S. Department of Energy and NSERC


References II


