An ensemble filter for heavy-tailed distributions

Mathieu Le Provost[1], Ricardo Baptista[2], Youssef Marzouk[1], and Jeff D. Eldredge[3]
[1] Massachusetts Institute of Technology, [2] California Institute of Technology,
[3] University of California, Los Angeles

Motivations

- ***Departure from Gaussian tails*** *is a common feature of geophysical inference problems due to the* ***nonlinear dynamical and observation processes*** *and the uncertainty from the physical sensors.*
- *Many filters* ***like the EnKF*** *assume that the tails of the forecast distribution are Gaussian and* ***not suited*** *for* ***heavy-tailed filtering problems****.*
- *How can we do* ***consistent inference*** *in this setting?*

Filtering problem and measure transport

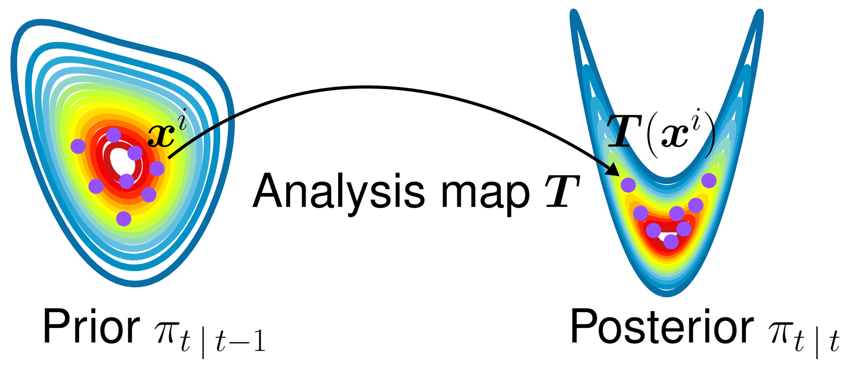
Setting : Nonlinear state-space model for $(\boldsymbol{x}_t, \boldsymbol{y}_t)$:
Dynamical model $\boldsymbol{x}_{t+1} = \boldsymbol{f}(\boldsymbol{x}_t) + \boldsymbol{w}_t \in \mathbb{R}^n,\ \boldsymbol{w}_t \perp \boldsymbol{x}_t$
Observation model $\boldsymbol{y}_t = \boldsymbol{h}(\boldsymbol{x}_t) + \boldsymbol{\epsilon}_t \in \mathbb{R}^d,\ \boldsymbol{\epsilon}_t \perp \boldsymbol{x}_t$

Filtering problem

Sequentially estimate the distribution for X_t given all the observations available up to that time $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_t$.

Ensemble filters propagate a set of M particles $\{\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(M)}\}$ to form an empirical approximation for the filtering density $\pi_{\mathsf{X}_t \mid \boldsymbol{y}_{1:t}} = \pi_{t \mid t}$.

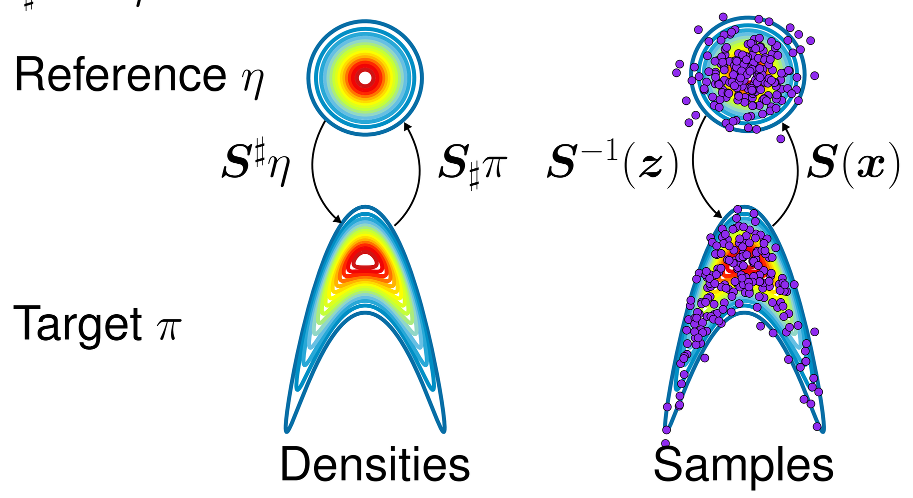
Analysis step: mapping the prior $\pi_{t \mid t-1}$ to posterior $\pi_{t \mid t}$



Analysis map of the Kalman filter : $\boldsymbol{T}_{KF}(\boldsymbol{y}, \boldsymbol{x}) = \boldsymbol{x} - \boldsymbol{\Sigma}_{\mathsf{X}_t, \mathsf{Y}_t} \boldsymbol{\Sigma}_{\mathsf{Y}_t}^{-1} (\boldsymbol{y} - \boldsymbol{y}^\star)$

How to build the analysis map $\boldsymbol{T}$? Use measure transport theory

A transport map $\boldsymbol{S}$ between 2 distributions π (target) and η (reference) is a transformation s.t. if $\boldsymbol{x} \sim \pi \Rightarrow \boldsymbol{S}(\boldsymbol{x}) \sim \eta$. We say that $\boldsymbol{S}$ pushes forward π to η, i.e., $\boldsymbol{S}_\sharp \pi = \eta$.



Aim: find a transport map suited for conditional inference

We use the Knothe-Rosenblatt (KR) rearrangement $\boldsymbol{S}$ btw π and η:
The unique lower triangular and strictly monotone map s.t. $\boldsymbol{S}_\sharp \pi = \eta$

$$\boldsymbol{S}(\boldsymbol{z}) = \boldsymbol{S}(z_1, z_2, \cdots, z_m) = \begin{bmatrix} S^1(z_1) \\ S^2(z_1, z_2) \\ \vdots \\ S^m(z_1, z_2, \ldots, z_m) \end{bmatrix}.$$

The KR has many nice features for conditional inference:

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

Gaussian case

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

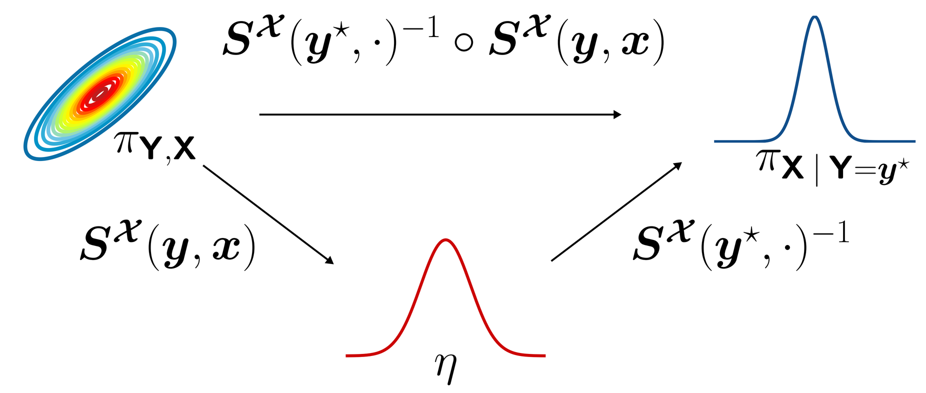
Takeaways

- Using **measure transport theory**, we introduce a new ensemble filter, called **ensemble robust filter** (EnRF), for **heavy-tailed filtering problems**.
- The EnRF can **adapt** its **prior-to-posterior update** to the **tail-heaviness of the data**.
- The EnRF features **tuning-free inflation** and **localization**.

Construction of the analysis map $\boldsymbol{T}$, [1]

Let's partition the KR rearrangement $\boldsymbol{S}$ s.t. $\boldsymbol{S}_\sharp \pi_{\mathsf{Y},\mathsf{X}} = \eta$: $\boldsymbol{S}(\boldsymbol{y}, \boldsymbol{x}) = \begin{bmatrix} \boldsymbol{S}^{\mathcal{Y}}(\boldsymbol{y}) \\ \boldsymbol{S}^{\mathcal{X}}(\boldsymbol{y}, \boldsymbol{x}) \end{bmatrix}$

- The map $\xi \mapsto \boldsymbol{S}^{\mathcal{X}}(\boldsymbol{y}^\star, \xi)$ pushes forward $\pi_{\mathsf{X} \mid \mathsf{Y}}(\cdot \mid \boldsymbol{y}^\star)$ to η • $\boldsymbol{S}^{\mathcal{X}}(\mathsf{Y}, \mathsf{X}) \sim \eta$
- **Analysis map** $\boldsymbol{T}(\boldsymbol{y}, \boldsymbol{x}) = \boldsymbol{S}^{\mathcal{X}}(\boldsymbol{y}^\star, \cdot)^{-1} \circ \boldsymbol{S}^{\mathcal{X}}(\boldsymbol{y}, \boldsymbol{x})$



A taxonomy of ensemble filters

Ensemble filters differ in the choice of

- The reference density η
- The class of approximating functions for $\boldsymbol{S}^{\mathcal{X}}$ or $\boldsymbol{T}$
- The estimation of $\boldsymbol{S}^{\mathcal{X}}$ or $\boldsymbol{T}$ from samples

Stochastic EnKF: $\eta = \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$ + Linear $\boldsymbol{S}^{\mathcal{X}}$ + Sample-based covariance estimator

t–distributions have tunable tail-heaviness

t–distributions are parameterized by a mean $\boldsymbol{\mu}_{\mathsf{X}} \in \mathbb{R}^n$, a scale matrix (PD) $\boldsymbol{C}_{\mathsf{X}} \in \mathbb{R}^{n \times n}$, and a degree of freedom (DOF) $\nu \in [1, \infty[$.

The degree of freedom ν characterizes the tail-heaviness:

- For $\nu = 1$, we recover the Cauchy distribution (algebraic tails)
- For $\nu = \infty$, we recover the Gaussian distribution (exponential tails)

Our contribution

- We choose a *t*–distributed reference distribution η_ν with tunable DOF ν to adapt to the tail-heaviness of the data.
- We restrict $\boldsymbol{S}^{\mathcal{X}}$ to be linear

Construction of $\boldsymbol{T}_\nu$

For a joint *t*–distribution $\pi_{\mathsf{Y},\mathsf{X}}$ with DOF ν, let $\boldsymbol{S}_\nu = [\boldsymbol{S}_\nu^{\mathcal{Y}}; \boldsymbol{S}_\nu^{\mathcal{X}}]$ be the KR rearrangement s.t. $\boldsymbol{S}_{\nu\sharp} \pi_{\mathsf{Y},\mathsf{X}} = \eta_\nu = St(\boldsymbol{0}, \boldsymbol{I}, \nu)$.

We obtain the analysis map $\boldsymbol{T}_\nu$ by partial inversion of $\boldsymbol{S}_\nu^{\mathcal{X}}$:

$$\boldsymbol{T}_\nu(\boldsymbol{y}, \boldsymbol{x}) = \boldsymbol{\mu}_{\mathsf{X}} + \boldsymbol{C}_{\mathsf{X},\mathsf{Y}} \boldsymbol{C}_{\mathsf{Y}}^{-1} (\boldsymbol{y}^\star - \boldsymbol{\mu}_{\mathsf{Y}}) + \sqrt{\frac{\alpha_{\mathsf{Y}}(\boldsymbol{y}^\star)}{\alpha_{\mathsf{Y}}(\boldsymbol{y})}} \left[(\boldsymbol{x} - \boldsymbol{\mu}_{\mathsf{X}}) - \boldsymbol{C}_{\mathsf{X},\mathsf{Y}} \boldsymbol{C}_{\mathsf{Y}}^{-1} (\boldsymbol{y} - \boldsymbol{\mu}_{\mathsf{Y}}) \right],$$

$$\text{with } \alpha_{\mathsf{Y}}(\boldsymbol{y}) = \frac{\nu + (\boldsymbol{y} - \boldsymbol{\mu}_{\mathsf{Y}})^\top \boldsymbol{C}_{\mathsf{Y}}^{-1} (\boldsymbol{y} - \boldsymbol{\mu}_{\mathsf{Y}})}{\nu + d}$$

Appealing features of $\boldsymbol{T}_\nu$

- $\boldsymbol{T}_\nu$ converges to Kalman filter for $\nu \to \infty$.
- $\boldsymbol{T}_\nu$ is robust to outlying synthetic observations, i.e., $\boldsymbol{y} \sim \pi_{\mathsf{Y}}$ with large $\alpha_{\mathsf{Y}}(\boldsymbol{y})$.
- $\boldsymbol{T}_\nu$ embeds an adaptive (ν) and data-dependent ($\boldsymbol{y}^\star$) multiplicative inflation.

Estimate $\boldsymbol{T}_\nu$ from samples

Goal : Estimate $\boldsymbol{T}_\nu$ from joint forecast samples $\{\boldsymbol{y}_t^{(i)}, \boldsymbol{x}_t^{(i)}\} \sim \pi_{\mathsf{Y}_t,\mathsf{X}_t}$
→ ensemble robust filter (EnRF).

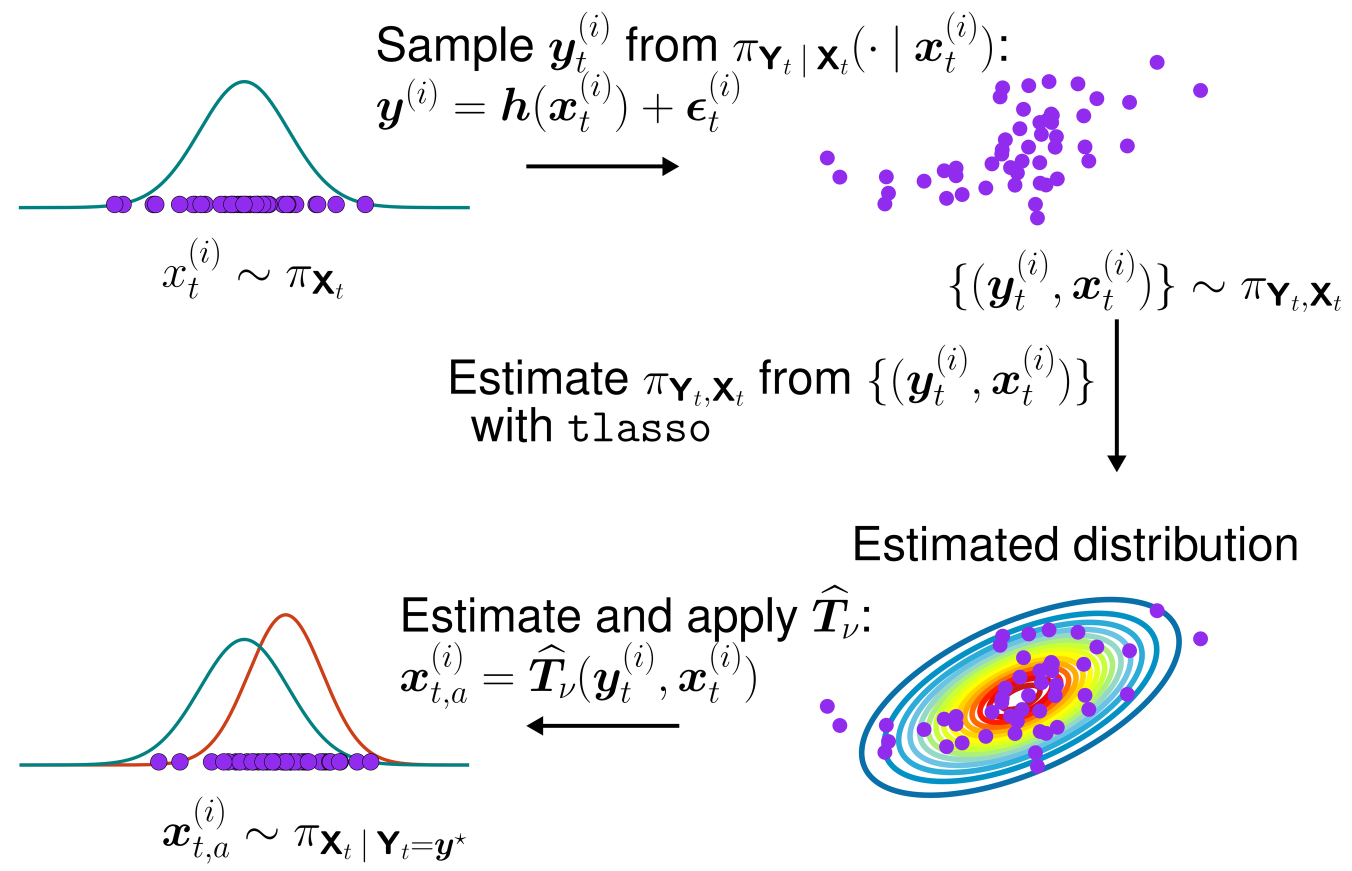
Our strategy:

- Leverage the conditional independence of $\pi_{\mathsf{Y}_t,\mathsf{X}_t} \sim$ sparsity of $\boldsymbol{C}_{\mathsf{Y}_t,\mathsf{X}_t}^{-1}$
- Use the `tlasso` [2]: a $\ell 1$-regularized expectation-maximization algorithm for *t*–distributions to estimate the statistics of $\pi_{\mathsf{Y}_t,\mathsf{X}_t}$:

Plug and play

The EnRF features tuning-free inflation and localization.

Analysis step of the ensemble robust filter (EnRF)



Results

Lorenz-63 setting

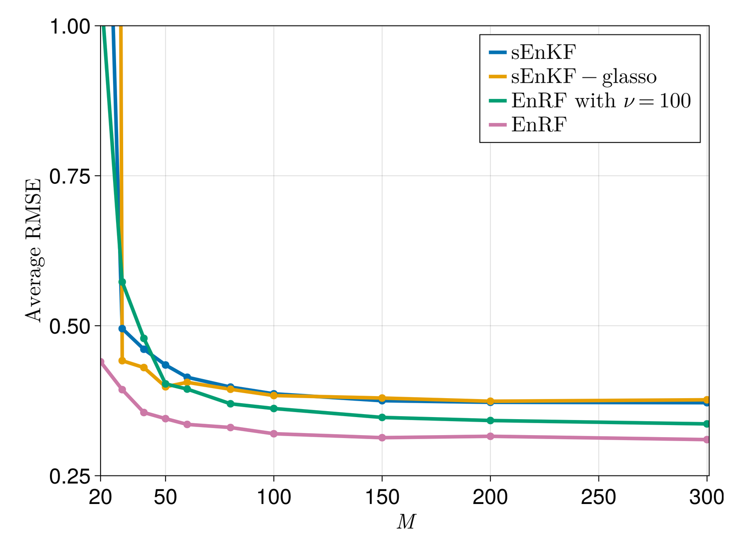


Figure: Evolution of the RMSE with the ensemble size M for the Lorenz-63 model with *t*–distributed observation noise with $\nu = 3.0$.
Reduction of the RMSE by 27% without tuning of the EnRF.

Lorenz-96 setting

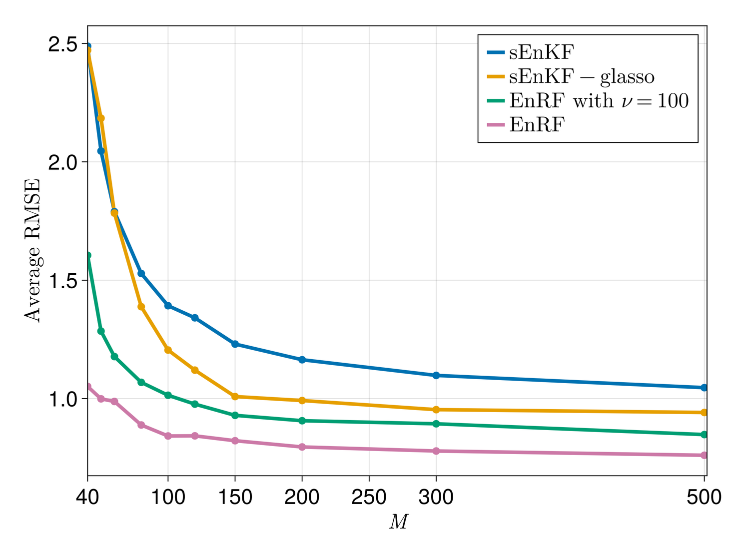


Figure: Evolution of the RMSE with the ensemble size M for the Lorenz-96 model with *t*–distributed observation noise with $\nu = 3.0$.
Reduction of the RMSE by 25% without tuning of the EnRF.

Links

- ArXiv print: https://arxiv.org/abs/2310.08741
- Github repository: https://github.com/mleprovost/Paper-Ensemble-Robust-Filter.jl
- Correspondence email: mleprovo@mit.edu

References

1. Spantini, A., Baptista, R. & Marzouk, Y. Coupling Techniques for Nonlinear Ensemble Filtering. *SIAM Review* **64.** Publisher: Society for Industrial and Applied Mathematics, 921–953 (Nov. 2022).
2. Finegold, M. A. & Drton, M. Robust graphical modeling with t-distributions. *arXiv preprint arXiv:1408.2033* (2014).

Acknowledgements

MLP and YM acknowledge support of the National Science Foundation under Grant PHY-2028125.