An ensemble filter for heavy-tailed t–distributions

Mathieu Le Provost[1], Ricardo Baptista[2],
Youssef Marzouk[3] and Jeff D. Eldredge[4]

[1]Long Island University, [2]California Institute of Technology,
[3]Massachusetts Institute of Technology, [4]University of California, Los Angeles

18th EnKF workshop
May 4, 2023

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 **at least** that the **tails of the forecast distribution are Gaussian** and not suited for heavy-tailed distributions.

Objective: How can do consistent inference in heavy-tailed filtering problems?

Problem setting

We consider a generic state-space model:

The evolution of the state $(\mathbf{X}_t)_{t \geq 0}$ is fully described by the initial distribution $\pi_{\mathbf{X}_0}$ and the dynamical model:

$$\mathbf{X}_t = \boldsymbol{f}(\mathbf{X}_{t-1}) + \mathbf{W}_t$$

We collect observations $(\mathbf{Y}_t)_{t>0}$ at every time step according to the observation model:

$$\mathbf{Y}_t = \boldsymbol{h}(\mathbf{X}_t) + \boldsymbol{\mathcal{E}}_t$$

Objective: Sequentially estimate **the filtering density** $\pi_{t \mid t} := \pi_{\mathbf{X}_t \mid \mathbf{Y}_{1:t} = \boldsymbol{y}_{1:t}}$

Generic ensemble filtering algorithm

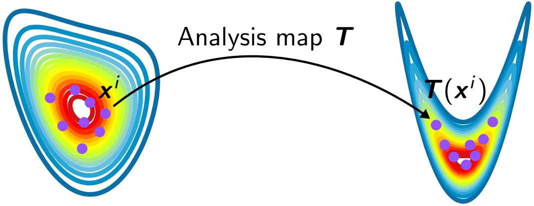
Ensemble filtering methods propagate a set of M particles $\{\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(M)}\}$ to form an empirical approximation for the filtering density $\pi_{t \mid t}$.

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

Ensemble filtering algorithms share the same forecast step but differ in the analysis step.

A "transformative" view of the analysis step



Analysis step: application of the **analysis map** $\boldsymbol{T}$: Prior $\pi_{t \mid t-1} \rightarrow$ Posterior $\pi_{t \mid t}$

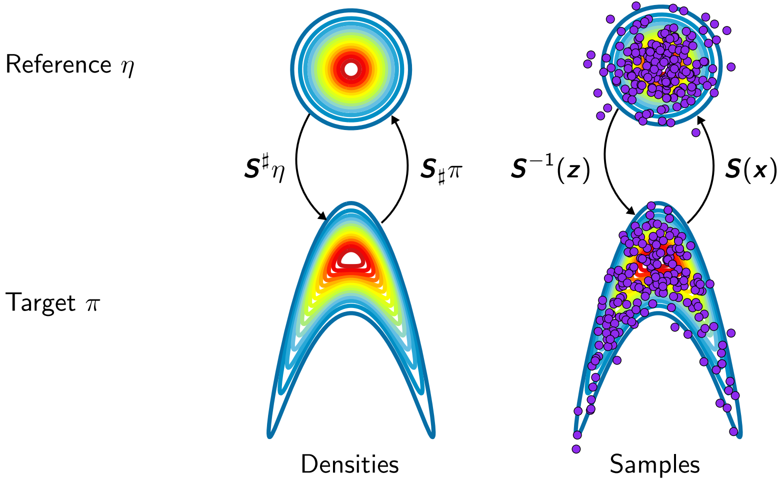
The analysis map of the Kalman filter $\boldsymbol{T}_{KF}$:

$$\boldsymbol{T}_{KF}(\boldsymbol{y}, \boldsymbol{x}) = \boldsymbol{x} - \boldsymbol{\Sigma}_{\mathsf{X},\mathsf{Y}} \boldsymbol{\Sigma}_{\mathsf{Y}}^{-1} (\boldsymbol{y} - \boldsymbol{y}^\star) = \boldsymbol{x} - \boldsymbol{K}(\boldsymbol{y} - \boldsymbol{y}^\star)$$

The ensemble Kalman filter (EnKF) [Evensen, 1994] constructs an estimate $\widehat{\boldsymbol{K}} \in \mathbb{R}^{n \times d}$ from limited samples $\{\boldsymbol{x}^1, \ldots, \boldsymbol{x}^M\}$ of the forecast distribution.

Transport map between two probability measures

- Seek a transport map $\boldsymbol{S}$ that pushes forward π to η, i.e. $\boldsymbol{S}_\sharp \pi = \eta$.
- Generate cheap and independent samples $\boldsymbol{x} \sim \pi \Rightarrow \boldsymbol{S}(\boldsymbol{x}) \sim \eta$.



Looking for a good map [Marzouk et al., 2016]

Consider the **Knothe-Rosenblatt (KR) rearrangement** $\boldsymbol{S}$ s.t. $\boldsymbol{S}_\sharp \pi = \eta$

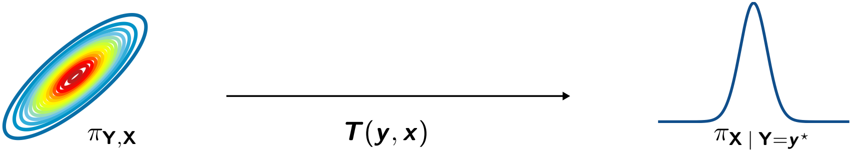
$$\boldsymbol{S}(\boldsymbol{z}) = \boldsymbol{S}(z_1, z_2, \cdots, z_m) = \begin{bmatrix} S^1(z_1) \\ S^2(z_1, z_2) \\ \vdots \\ S^m(z_1, z_2, \ldots, z_m) \end{bmatrix}.$$

- The KR has many nice features for Bayesian inference, e.g. easily invertible and $\det \nabla \boldsymbol{S}(\boldsymbol{x})$ is simple to evaluate [Marzouk et al., 2016, Baptista et al., 2020].
- The 1D map $\xi \mapsto S^k(x_1, x_2, \ldots, x_{k-1}, \xi)$ characterizes the marginal conditional $\pi_{\mathsf{X}_k \mid \mathbf{X}_{1:k-1} = \boldsymbol{x}_{1:k-1}}(\xi)$.

Gaussian case

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

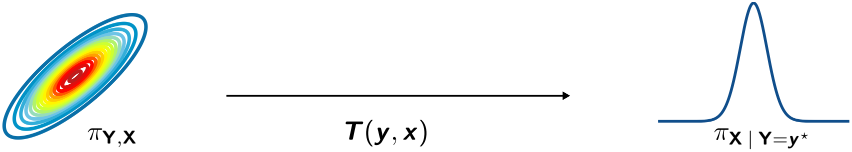
Construction of the analysis map [Spantini et al., 2022]



Construction of the analysis map [Spantini et al., 2022]

Consider the KR rearrangement $\boldsymbol{S}$ s.t. $\boldsymbol{S}_\sharp \pi_{\mathbf{Y},\mathbf{X}} = \eta$

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

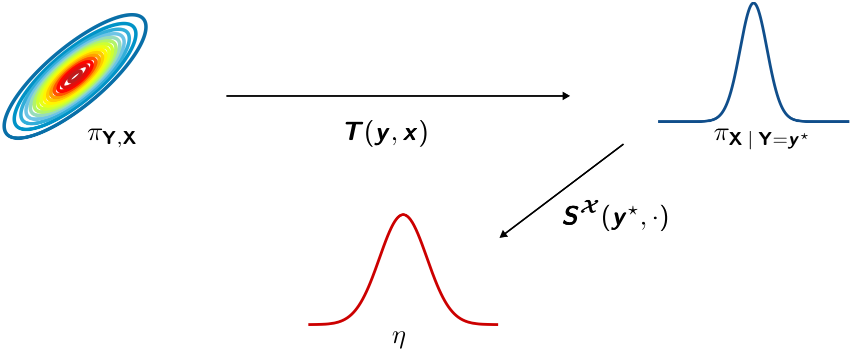


Construction of the analysis map [Spantini et al., 2022]

Consider the KR rearrangement $\boldsymbol{S}$ s.t. $\boldsymbol{S}_\sharp \pi_{\mathbf{Y},\mathbf{X}} = \eta$

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

- The map $\xi \mapsto \boldsymbol{S}^{\mathcal{X}}(\boldsymbol{y}^\star, \xi)$ pushes forward $\pi_{\mathbf{X} \mid \mathbf{Y}}(\cdot \mid \boldsymbol{y}^\star)$ to η

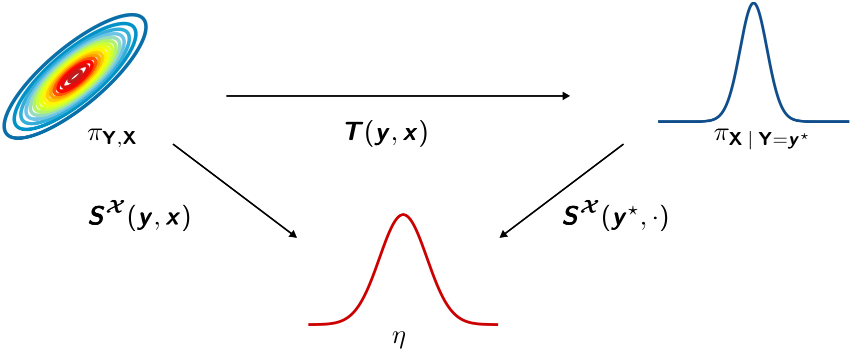


Construction of the analysis map [Spantini et al., 2022]

Consider the KR rearrangement $\boldsymbol{S}$ s.t. $\boldsymbol{S}_\sharp \pi_{\mathbf{Y},\mathbf{X}} = \eta$

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

- The map $\xi \mapsto \boldsymbol{S}^{\mathcal{X}}(\boldsymbol{y}^\star, \xi)$ pushes forward $\pi_{\mathbf{X} \mid \mathbf{Y}}(\cdot \mid \boldsymbol{y}^\star)$ to η
- $\boldsymbol{S}^{\mathcal{X}}(\mathbf{Y}, \mathbf{X}) \sim \eta$

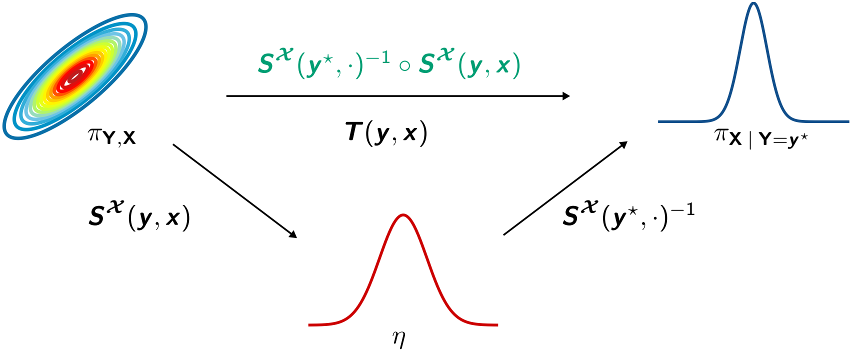


Construction of the analysis map [Spantini et al., 2022]

Consider the KR rearrangement $\boldsymbol{S}$ s.t. $\boldsymbol{S}_\sharp \pi_{\mathbf{Y},\mathbf{X}} = \eta$

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

- The map $\xi \mapsto \boldsymbol{S}^{\mathcal{X}}(\boldsymbol{y}^\star, \xi)$ pushes forward $\pi_{\mathbf{X} \mid \mathbf{Y}}(\cdot \mid \boldsymbol{y}^\star)$ to η
- $\boldsymbol{S}^{\mathcal{X}}(\mathbf{Y}, \mathbf{X}) \sim \eta$



Derivation of the analysis map of the Kalman filter [Spantini et al., 2022]

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

We recover the analysis map of the Kalman filter when

1. $\boldsymbol{S}^{\boldsymbol{\mathcal{X}}}$ is linear
2. The reference density is the standard normal distribution $\eta = \mathcal{N}\left(\mathbf{0}_{d+n}, \boldsymbol{I}_{d+n}\right)$

Ensemble filters differ in the choice of

- the reference density
- the class of functions to represent $\boldsymbol{S}^{\boldsymbol{\mathcal{X}}}$
- the estimation of $\boldsymbol{S}^{\boldsymbol{\mathcal{X}}}$ from samples

Limitations of light-tailed filter for heavy-tailed distributions

Many ensemble filters like the EnKF assume **at least** that the **tails of the forecast distribution are Gaussian**.

These filters don't provide consistent inference for heavy-tailed filtering problems.

Our contribution: Introduce a new ensemble filter called **ensemble robust filter (EnRF)** based on the following assumptions:

- We restrict $\boldsymbol{S}^{\boldsymbol{\mathcal{X}}}$ to be linear
- We choose a reference distribution whose **tail-heaviness can be adapted to the data**.

t–distributions

t–distributions are a family of distributions parameterized by a mean $\boldsymbol{\mu}_{\mathbf{X}} \in \mathbb{R}^n$, a scale matrix $\boldsymbol{C}_{\mathbf{X}} \in \mathbb{R}^{n \times n}$, and a degree of freedom $\nu_{\mathbf{X}} \in [1, \infty[$.

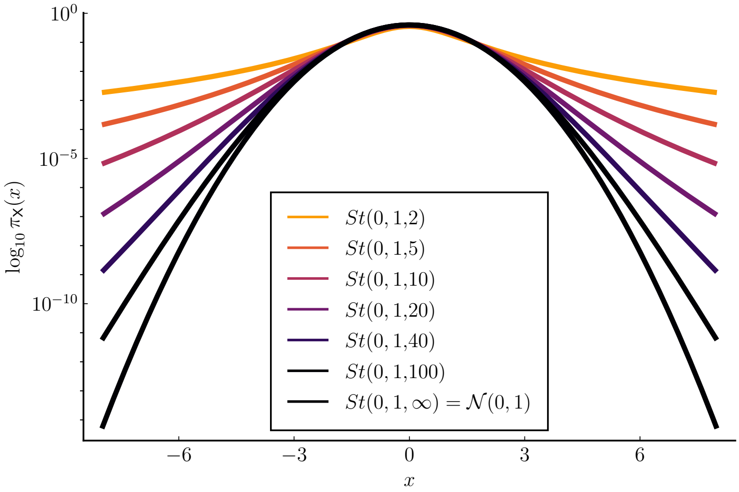
The degree of freedom $\nu_{\mathbf{X}}$ characterizes the tail-heaviness:

- For $\nu_{\mathbf{X}} = 1$, we recover the Cauchy distribution
- For $\nu_{\mathbf{X}} = \infty$, we recover the Gaussian distribution

Relation between the parameters and the moments of a t–distribution:

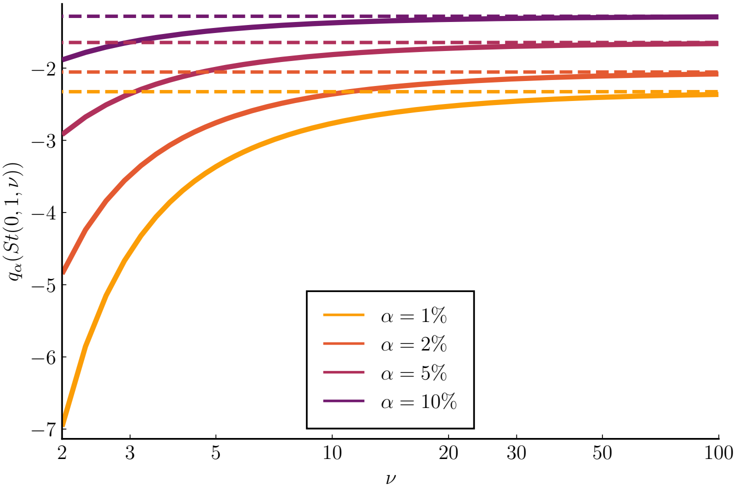
$$\mathrm{E}_{\pi_{\mathbf{X}}}[\boldsymbol{x}] = \boldsymbol{\mu}_{\mathbf{X}}, \text{ for } \nu_{\mathbf{X}} > 1$$
$$\mathrm{E}_{\pi_{\mathbf{X}}}\left[(\boldsymbol{x} - \boldsymbol{\mu}_{\mathbf{X}})(\boldsymbol{x} - \boldsymbol{\mu}_{\mathbf{X}})^\top\right] = \frac{\nu_{\mathbf{X}}}{\nu_{\mathbf{X}} - 2} \boldsymbol{C}_{\mathbf{X}} \text{ for } \nu_{\mathbf{X}} > 2.$$

Probability density function of t–distributions



PDF of the univariate standard t–distribution $St\,(0, 1, \nu)$ for $\nu = 2, 5, 10, 20, 100$ and the univariate standard Gaussian distribution $\mathcal{N}\,(0, 1) = St\,(0, 1, \infty)$ (black).

Quantiles of the t–distributions



Evolution of α-quantile q_α for $\alpha = 1\%, 2\%, 5\%, 10\%$ with the degree of freedom ν of the univariate standard t–distribution. Dashed lines corresponds to the α-quantiles for the standard Gaussian distribution $\mathcal{N}(0,1) = St(0,1,\infty)$.

Useful properties of t–distributions:

1. t–distributions are closed under affine transformations:

   If $\mathbf{X} \sim St\left(\boldsymbol{\mu}_{\mathbf{X}}, \boldsymbol{C}_{\mathbf{X}}, \nu_{\mathbf{X}}\right)$, then $\mathbf{Z} = \boldsymbol{A}\mathbf{X} + \boldsymbol{b} \sim St\left(\boldsymbol{A}\boldsymbol{\mu}_{\mathbf{X}} + \boldsymbol{b}, \boldsymbol{A}\boldsymbol{C}_{\mathbf{X}}\boldsymbol{A}^{\top}, \nu_{\mathbf{X}}\right)$

2. Conditional and marginal distributions are known in closed form.

   Consider $\begin{bmatrix}\mathbf{Y}\\ \mathbf{X}\end{bmatrix} \sim St\left(\begin{bmatrix}\boldsymbol{\mu}_{\mathbf{X}}\\ \boldsymbol{\mu}_{\mathbf{Y}}\end{bmatrix}, \begin{bmatrix}\boldsymbol{C}_{\mathbf{Y}} & \boldsymbol{C}_{\mathbf{X},\mathbf{Y}}^{\top}\\ \boldsymbol{C}_{\mathbf{X},\mathbf{Y}} & \boldsymbol{C}_{\mathbf{X}}\end{bmatrix}, \nu\right)$

- **Marginal**: $\mathbf{X} \sim St\left(\boldsymbol{\mu}_{\mathbf{X}}, \boldsymbol{C}_{\mathbf{X}}, \nu_{\mathbf{X}}\right)$
- **Conditional**: $\pi_{\mathbf{X}\,|\,\mathbf{Y}=\boldsymbol{y}} \sim St\left(\boldsymbol{\mu}_{\mathbf{X}\,|\,\mathbf{Y}=\boldsymbol{y}}, \boldsymbol{C}_{\mathbf{X}\,|\,\mathbf{Y}=\boldsymbol{y}}, \nu_{\mathbf{X}\,|\,\mathbf{Y}=\boldsymbol{y}}\right)$ with

$$\boldsymbol{\mu}_{\mathbf{X}\,|\,\boldsymbol{y}} = \boldsymbol{\mu}_{\mathbf{X}} + \boldsymbol{C}_{\mathbf{X},\mathbf{Y}}\boldsymbol{C}_{\mathbf{Y}}^{-1}(\boldsymbol{y} - \boldsymbol{\mu}_{\mathbf{Y}})$$

$$\boldsymbol{C}_{\mathbf{X}\,|\,\mathbf{Y}=\boldsymbol{y}} = \underbrace{\frac{\nu + (\boldsymbol{y} - \boldsymbol{\mu}_{\mathbf{Y}})^{\top}\boldsymbol{C}_{\mathbf{Y}}^{-1}(\boldsymbol{y} - \boldsymbol{\mu}_{\mathbf{Y}})}{\nu + d}}_{\alpha_{\mathbf{Y}}(\boldsymbol{y})>0}\underbrace{\left(\boldsymbol{C}_{\mathbf{X}} - \boldsymbol{C}_{\mathbf{X},\mathbf{Y}}\boldsymbol{C}_{\mathbf{Y}}^{-1}\boldsymbol{C}_{\mathbf{X},\mathbf{Y}}^{\top}\right)}_{\text{Schur complement } \boldsymbol{C}_{\mathbf{X}\setminus\mathbf{Y}}}$$

$$\nu_{\mathbf{X}\,|\,\mathbf{Y}=\boldsymbol{y}} = \nu + d$$

Analysis map $\boldsymbol{T}_\nu$ for t–distributions

Let $\boldsymbol{S}_\nu$ the KR that pushes forward the joint t–distribution $\pi_{\mathbf{Y},\mathbf{X}}$ with dof ν to a "judicious" t–distribution η_ν with same dof, i.e. $\boldsymbol{S}_{\nu\sharp}\pi_{\mathbf{Y},\mathbf{X}} = \eta_\nu$.

Key: $\boldsymbol{S}_\nu$ can be computed in closed form (new result).

We obtain the analysis map $\boldsymbol{T}_\nu : \mathbb{R}^d \times \mathbb{R}^n \to \mathbb{R}^n$ by partial inversion of $\boldsymbol{S}_\nu^{\boldsymbol{\mathcal{X}}}$:

$$
\begin{aligned}
\boldsymbol{T}_\nu(\boldsymbol{y}, \boldsymbol{x}) &= \boldsymbol{S}_\nu^{\boldsymbol{\mathcal{X}}}(\boldsymbol{y}^\star, \cdot)^{-1} \circ \boldsymbol{S}_\nu^{\boldsymbol{\mathcal{X}}}(\boldsymbol{y}, \boldsymbol{x}) \\
&= \boldsymbol{\mu}_{\mathbf{X}} + \boldsymbol{C}_{\mathbf{X},\mathbf{Y}}\boldsymbol{C}_{\mathbf{Y}}^{-1}(\boldsymbol{y}^\star - \boldsymbol{\mu}_{\mathbf{Y}}) + \sqrt{\frac{\alpha_{\mathbf{Y}}\left(\boldsymbol{y}^\star\right)}{\alpha_{\mathbf{Y}}\left(\boldsymbol{y}\right)}}\left[(\boldsymbol{x} - \boldsymbol{\mu}_{\mathbf{X}}) - \boldsymbol{C}_{\mathbf{X},\mathbf{Y}}\boldsymbol{C}_{\mathbf{Y}}^{-1}(\boldsymbol{y} - \boldsymbol{\mu}_{\mathbf{Y}})\right]
\end{aligned}
$$

Note that $\boldsymbol{C}_{\mathbf{X},\mathbf{Y}}\boldsymbol{C}_{\mathbf{Y}}^{-1} = \boldsymbol{\Sigma}_{\mathbf{X},\mathbf{Y}}\boldsymbol{\Sigma}_{\mathbf{Y}}^{-1}$.

Connection with the Kalman filter

Let's perform an asymptotic expansion of $\boldsymbol{T}_\nu$ for large ν:

$$\boldsymbol{T}_\nu = \underbrace{\boldsymbol{T}_\infty}_{\boldsymbol{T}_{KF}} + \text{higher order terms in } O\left(\frac{1}{\nu}\right)$$

- The zeroth order term $\boldsymbol{T}_\infty$ is exactly the analysis map of the Kalman filter $\boldsymbol{T}_{\text{KF}}$.
- The higher order terms correct the analysis map $\boldsymbol{T}_{\text{KF}}$ to account for the finite degree of freedom.

Takeaway

The analysis map $\boldsymbol{T}_\nu$ generalizes the analysis map of the Kalman filter for t–distributions with finite degree of freedom!

Sensitivity to outlying synthetic observations

Context: Observation models often rely on simplified physics and suffer from mis-specifications of the observation operator and observation noise.

Let $(\boldsymbol{y}^{(j)}, \boldsymbol{x}^{(j)})$ be a joint forecast sample with an outlying synthetic observation $\boldsymbol{y}^{(j)}$ generated by the likelihood model $\pi_{\mathbf{Y} \mid \mathbf{X}=\boldsymbol{x}^{(j)}}$, such that $\delta_{\mathbf{Y}}\big(\boldsymbol{y}^{(j)}\big) = \big(\boldsymbol{y}^{(j)} - \boldsymbol{\mu}_{\mathbf{Y}}\big)^{\top} \boldsymbol{C}_{\mathbf{Y}}{}^{-1} \big(\boldsymbol{y}^{(j)} - \boldsymbol{\mu}_{\mathbf{Y}}\big) \to \infty$. Then, the analysis map $\boldsymbol{T}_{\nu}$ reduces to

$$\boldsymbol{T}_{\nu}(\boldsymbol{y}, \boldsymbol{x}) = \boldsymbol{\mu}_{\mathbf{X}} + \boldsymbol{C}_{\mathbf{X},\mathbf{Y}} \boldsymbol{C}_{\mathbf{Y}}^{-1}(\boldsymbol{y}^{\star} - \boldsymbol{\mu}_{\mathbf{Y}}).$$

Takeaways:

- The outlying observation $\boldsymbol{y}^{(j)}$ vanishes from the analysis map.
- The prior sample $\boldsymbol{x}^{(j)}$ is mapped to the posterior mean/median/mode.

What is the pushforward of $\pi_{\mathbf{X}}$ by the maps $\boldsymbol{T}_\nu$ and $\boldsymbol{T}_{\mathrm{KF}}$?

$\boldsymbol{T}_\nu$ leads to exact inference: $\boldsymbol{T}_{\nu\sharp}\pi_{\mathbf{X}} = \pi_{\mathbf{X} \mid \boldsymbol{y}^\star} = St\left(\boldsymbol{\mu}_{\mathbf{X} \mid \boldsymbol{y}^\star}, \boldsymbol{C}_{\mathbf{X} \mid \boldsymbol{y}^\star}, \nu_{\mathbf{X} \mid \boldsymbol{y}^\star}\right)$:

$$\begin{aligned} \boldsymbol{\mu}_{\mathbf{X} \mid \boldsymbol{y}^\star} &= \boldsymbol{\mu}_{\mathbf{X}} + \boldsymbol{C}_{\mathbf{X},\mathbf{Y}}\boldsymbol{C}_{\mathbf{Y}}^{-1}(\boldsymbol{y}^\star - \boldsymbol{\mu}_{\mathbf{Y}}), \\ \boldsymbol{C}_{\mathbf{X} \mid \boldsymbol{y}^\star} &= \alpha_{\mathbf{Y}}(\boldsymbol{y}^\star)\, \boldsymbol{C}_{\mathbf{X}\backslash\mathbf{Y}}, \\ \nu_{\mathbf{X} \mid \boldsymbol{y}^\star} &= \nu + d, \end{aligned} \tag{1}$$

We interpret $\alpha_{\mathbf{Y}}(\boldsymbol{y}^\star)$ as an **adaptive** and **data-dependent** multiplicative inflation.

$\boldsymbol{T}_{\mathrm{KF}\sharp}\pi_{\mathbf{X}} = St\left(\boldsymbol{\mu}_{\boldsymbol{T}_{\mathrm{KF}\sharp}\pi_{\mathbf{X}}}, \boldsymbol{C}_{\boldsymbol{T}_{\mathrm{KF}\sharp}\pi_{\mathbf{X}}}, \nu_{\boldsymbol{T}_{\mathrm{KF}\sharp}\pi_{\mathbf{X}}}\right)$ with

$$\begin{aligned} \boldsymbol{\mu}_{\boldsymbol{T}_{\mathrm{KF}\sharp}\pi_{\mathbf{X}}} &= \boldsymbol{\mu}_{\mathbf{X}} + \boldsymbol{C}_{\mathbf{X},\mathbf{Y}}\boldsymbol{C}_{\mathbf{Y}}^{-1}(\boldsymbol{y}^\star - \boldsymbol{\mu}_{\mathbf{Y}}), \\ \boldsymbol{C}_{\boldsymbol{T}_{\mathrm{KF}\sharp}\pi_{\mathbf{X}}} &= 1 \times \boldsymbol{C}_{\mathbf{X}\backslash\mathbf{Y}}, \\ \nu_{\boldsymbol{T}_{\mathrm{KF}\sharp}\pi_{\mathbf{X}}} &= \nu + 0. \end{aligned} \tag{2}$$

The Kalman filter is only consistent to estimate the mean value of t–distributions.

Challenges of estimating heavy-tailed distributions from samples

The classical sample mean and covariance estimators are derived from a maximum likelihood approach for Gaussian distributions.

These light-tailed estimators are very sensitive to outliers and introduce additional variance.

We use a "regularized" expectation-maximization algorithm (`EMq`) to estimate the heavy-tailed joint forecast distribution $\pi_{\mathbf{Y},\mathbf{X}}$ from $\{(\boldsymbol{y}^{(i)}, \boldsymbol{x}^{(i)})\}$ [Doğru et al., 2018].

We call **ensemble robust filter (EnRF)** the ensemble filter that estimates $\boldsymbol{T}_\nu$ with the `EMq` from the joint forecast samples $\{(\boldsymbol{y}^{(i)}, \boldsymbol{x}^{(i)})\}$.

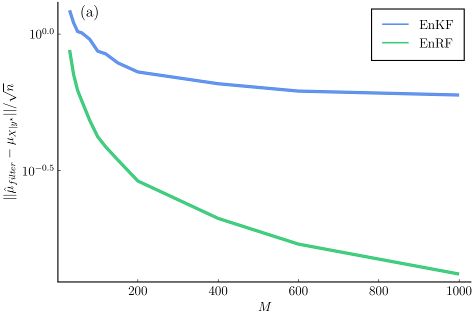
Empirical performance of light-tailed and heavy-tailed estimators

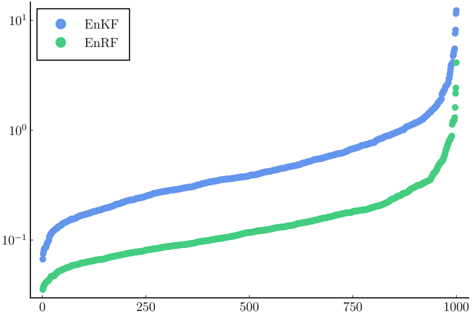
Consider $\begin{bmatrix} \mathbf{Y} \\ \mathbf{X} \end{bmatrix} \sim St\left(\begin{bmatrix} \boldsymbol{\mu}_{\mathbf{X}} \\ \boldsymbol{\mu}_{\mathbf{Y}} \end{bmatrix}, \begin{bmatrix} \boldsymbol{C}_{\mathbf{Y}} & \boldsymbol{C}_{\mathbf{X},\mathbf{Y}}^\top \\ \boldsymbol{C}_{\mathbf{X},\mathbf{Y}} & \boldsymbol{C}_{\mathbf{X}} \end{bmatrix}, \nu\right)$ with $\nu = 2.1$, $\mathbf{X} \in \mathbb{R}^{10}$ and $\mathbf{Y} \in \mathbb{R}^5$.

Experiment:

1. We generate M samples $\{(\boldsymbol{y}^{(i)}, \boldsymbol{x}^{(i)})\} \sim \pi_{\mathbf{Y},\mathbf{X}}$.
2. We apply the analysis map of the EnKF and the EnRF to assimilate a realization $\boldsymbol{y}^\star \sim \pi_{\mathbf{Y}}$
3. We compute the sample mean and covariance for the two posterior ensembles.

Posterior mean estimates



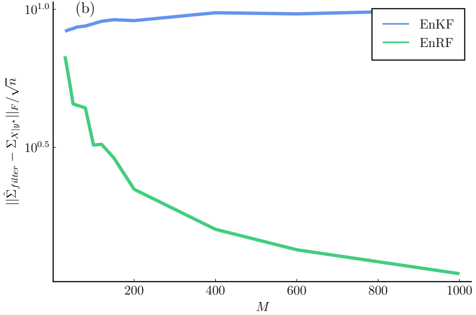


Left panel: Evolution of the error RMSE $= ||\widehat{\boldsymbol{\mu}}_{filter} - \boldsymbol{\mu}_{\mathbf{X} \mid \boldsymbol{y}^\star}||_2/\sqrt{n}$ with the ensemble size M.

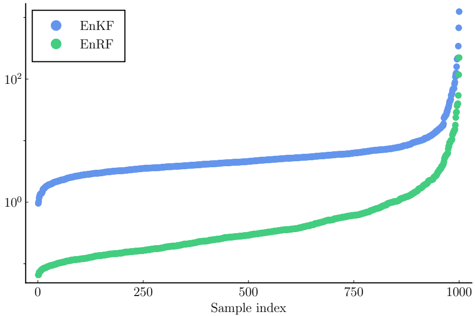
Right panel: Empirical distribution of the RMSE over 1000 realizations of $\boldsymbol{y}^\star$ using $M = 600$ samples.

- Both filters are consistent,
- but the sEnKF has a much slower convergence rate than the EnRF.

Posterior covariance estimates



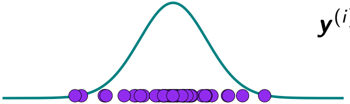
Left panel: Evolution of the error $||\widehat{\mathbf{\Sigma}}_{filter} - \mathbf{\Sigma}_{\mathbf{X} \mid \boldsymbol{y}^\star}||_F/\sqrt{n}$ with the ensemble size M.



Right panel: Empirical distribution of the error $||\widehat{\mathbf{\Sigma}}_{filter} - \mathbf{\Sigma}_{\mathbf{X} \mid \boldsymbol{y}^\star}||_F/\sqrt{n}$ over 1000 realizations of $\boldsymbol{y}^\star$ using $M = 600$.

- The sEnKF is not consistent to estimate covariances of t–distributions.

Analysis step of the EnRF

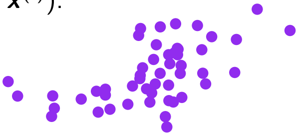


$\boldsymbol{x}^i \sim \pi_{\mathbf{X}}$

Sample $\boldsymbol{y}^{(i)}$ from $\pi_{\mathbf{Y}\,|\,\mathbf{X}}(\cdot \mid \boldsymbol{x}^{(i)})$:

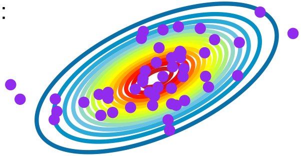
$\boldsymbol{y}^{(i)} = \boldsymbol{h}(\boldsymbol{x}^{(i)}) + \boldsymbol{\epsilon}^{(i)}$

$\{(\boldsymbol{y}^{(i)}, \boldsymbol{x}^{(i)})\} \sim \pi_{\mathbf{Y},\mathbf{X}}$



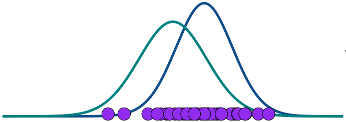
Estimate $\pi_{\mathbf{Y},\mathbf{X}}$ from $\{(\boldsymbol{y}^{(i)}, \boldsymbol{x}^{(i)})\}$ with `EMq`

Estimated distribution



Estimate and apply $\hat{\boldsymbol{T}}_\nu$:

$\boldsymbol{x}_a^{(i)} = \hat{\boldsymbol{T}}_\nu(\boldsymbol{y}^{(i)}, \boldsymbol{x}^{(i)})$



$\boldsymbol{x}_a^{(i)} \sim \pi_{\mathbf{X}\,|\,\mathbf{Y}=\boldsymbol{y}^\star}$

Estimate the tail-heaviness on the fly

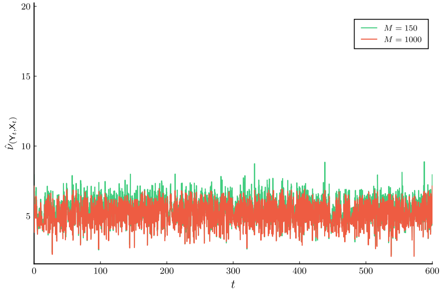
- We can estimate the dof at each assimilation cycle with the `EMq`
- The EnRF can adapt its analysis map to the tail heaviness of the data!

Limitation: Computational cost of the `EMq` when the dof is unknown.

Idea: Present 3 variants of the EnRF that differ in the frequency and the samples used to estimate the dof

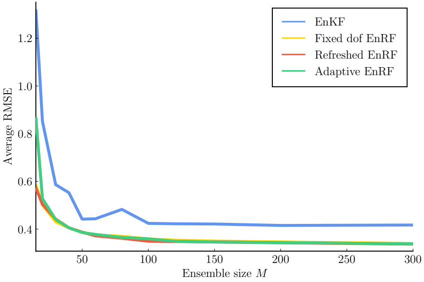
1. The `RefreshEnRF`: We maintain a buffer of past filtering samples, and estimate the dof with this larger ensemble every $\Delta t_{refresh}$.
2. The `FixedEnRF`: The dof is estimated once from a free-run of the state-space model and fixed for the entire assimilation ($\Delta t_{refresh} = \infty$)
3. The `AdaptEnRF`: The dof is estimated at each assimilation cycle ($\Delta t_{refresh} = \Delta t_{obs}$)

Lorenz-63



Empirical degree of freedom $\hat{\nu}_{(\mathbf{Y}_t,\mathbf{X}_t)}$ with $M = 150, 1000$ samples of the joint forecast density $\{(\boldsymbol{y}_t^{(i)}, \boldsymbol{x}_t^{(i)})\} \sim \pi_{(\mathbf{Y}_t,\mathbf{X}_t) \mid \mathbf{Y}_{1:t-1}}$ for the Lorenz-63 problem with t–distributed observation noise.

RMSE results for Lorenz-63



sEnKF FixedEnRF RefreshEnRF AdaptEnRF

Evolution of the RMSE with the ensemble size M for the Lorenz-63 model with t–distributed observation noise with $\nu = 3.0$.

We optimally tune the multiplicative inflation of the sEnKF.

25% reduction of the RMSE with the EnRF!

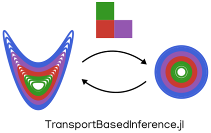
The different EnRFs don't require tuning: Plug and Play!

Conclusion and Outlook

Summary:

- We introduce the EnRF that generalizes the EnKF to heavy-tailed t–distributions
- The EnRF adapts its prior-to-posterior to the tail-heaviness of the data
  $\Longrightarrow$ Adaptive and data-dependent multiplicative inflation
- The EnRF requires no tuning: **Plug and Play!**

Software: The `EnRF` will be soon available in `TransportBasedInference.jl`



Acknowledgements: Support from National Science Foundation is gratefully acknowledged.

Contact: mathieuleprovost1@gmail.com

References I

Baptista, R., Zahm, O., and Marzouk, Y. (2020).
An adaptive transport framework for joint and conditional density estimation.
arXiv preprint arXiv:2009.10303.

Doğru, F. Z., Bulut, Y. M., and Arslan, O. (2018).
Doubly reweighted estimators for the parameters of the multivariate t-distribution.
Communications in Statistics-Theory and Methods, 47(19):4751–4771.

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.

Marzouk, Y., Moselhy, T., Parno, M., and Spantini, A. (2016).
Sampling via measure transport: An introduction.
Handbook of Uncertainty Quantification, pages 1–41.

Spantini, A., Baptista, R., and Marzouk, Y. (2022).
Coupling Techniques for Nonlinear Ensemble Filtering.
SIAM Review, 64(4):921–953.
Publisher: Society for Industrial and Applied Mathematics.