Regularization of the EnKF for non-local observations: application to elliptic observations

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

Abstract

We introduce a low-rank factorization of the EnKF (LREnKF) for non-local observations (e.g., radiances measured by satellites, fluxes through surfaces, or solutions of elliptic PDEs). Classical regularization techniques assume that the ***observations have local state dependence*** *and* ***suppress all correlations at long distances****. For* ***non-local observations****, we cannot separate* ***slowly decaying physical interactions*** *from* ***spurious long-range correlations****. Instead, non-local inverse problems have structure: a low-dimensional projection of the observations strongly informs a low-dimensional subspace of the state space.*

Filtering problem

Consider a nonlinear state-space model for $(\boldsymbol{x}_t, \boldsymbol{y}_t)$:

- Nonlinear and non-local dynamics $\boldsymbol{x}_{t+1} = \boldsymbol{f}(\boldsymbol{x}_t) + \boldsymbol{w}_t \in \mathbb{R}^n$
- Nonlinear and non-local observations $\boldsymbol{y}_t = \boldsymbol{h}(\boldsymbol{x}_t) + \boldsymbol{\epsilon}_t \in \mathbb{R}^d$

where $\boldsymbol{w}_t, \boldsymbol{\epsilon}_t$ are independent Gaussian random variables

Filtering problem

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

Elliptic observation model

Consider a Poisson equation evaluated at d locations $\{\boldsymbol{r}_j\} \in \Omega$:

$$\nabla^2 u_t(\boldsymbol{r}_j) = q(\boldsymbol{r}_j; \boldsymbol{x}_t), \text{ with } u_t(\boldsymbol{r}) = 0 \text{ for } ||\boldsymbol{r}|| \to \infty \qquad (1)$$

where $q(\boldsymbol{r}; \boldsymbol{x}_t)$ is a forcing term that depends nonlinearly on the state $\boldsymbol{x}_t$.

By convolution with the Green's function G of the Laplacian ∇^2,

$$u_t(\boldsymbol{r}_j) = \int_{\boldsymbol{s} \in \Omega} G(\boldsymbol{r}_j - \boldsymbol{s}) q(\boldsymbol{s}; \boldsymbol{x}_t) \mathrm{d}\boldsymbol{s}. \qquad (2)$$

The solution u_t of the elliptic PDE (2) is a **non-local** function of $\boldsymbol{x}_t$.

Inference with elliptic observations

Estimate the state X_t from limited and noisy evaluations of the solution u_t of the elliptic PDE (2):

$[\boldsymbol{y}_t]_j = u_t(\boldsymbol{r}_j) + [\boldsymbol{\epsilon}_t]_j, \; j = 1, \ldots, d.$

Ensemble Kalman filtering

Kalman filter update: $\boldsymbol{x}_t^a = \boldsymbol{x}_t - \underbrace{\Sigma_{\mathsf{X}_t, \mathsf{Y}_t} \Sigma_{\mathsf{Y}_t}^{-1}}_{\text{Kalman gain } \boldsymbol{K}_t} (\boldsymbol{y}_t - \boldsymbol{y}_t^\star)$

The EnKF constructs an estimate $\widehat{\boldsymbol{K}}_t \in \mathbb{R}^{n \times d}$ from M forecast (i.e., prior) samples $\{\boldsymbol{x}_t^{(1)}, \ldots, \boldsymbol{x}_t^{(M)}\}$ with $M \ll n, d$.

Regularization of the EnKF

Estimated gain $\widehat{\boldsymbol{K}}_t$ suffers from
- rank-deficiency, sampling errors,
- spurious long-range state correlations.

Regularization of $\widehat{\boldsymbol{K}}_t$ is essential in high-dimensions.

Limitations of distance localization

Distance localization regularizes $\widehat{\boldsymbol{K}}_t$ by systematically removing all long-range interactions.

Not suited for elliptic observations, as we cannot disentangle:
- the slowly decaying physical interactions (algebraic decay of the Green's function G)
- the spurious long-range correlations (finite ensemble size)

How to regularize $\widehat{\boldsymbol{K}}_t$ with non-local observations?

Low-rank informative structure

- Only part of the state is informed by the observations
- Only part of the observation space is relevant to the states

Perform the inference in the low-dimensional informative subspaces.

Low-rank factorization of the Kalman gain

For a linear-Gaussian observation model:

$$\mathsf{Y} = \boldsymbol{H}\mathsf{X} + \mathcal{E}, \; \mathsf{X} \sim \mathcal{N}(\boldsymbol{0}, \Sigma_{\mathsf{X}}), \; \mathcal{E} \sim \mathcal{N}(\boldsymbol{0}, \Sigma_{\mathcal{E}}).$$

The SVD of the whitened observation matrix is given by $\tilde{\boldsymbol{H}} = \Sigma_{\mathcal{E}}^{-1/2} \boldsymbol{H} \Sigma_{\mathsf{X}}^{1/2} \in \mathbb{R}^{d \times n}$ reads:

$$\tilde{\boldsymbol{H}} = \underbrace{\boldsymbol{U}}_{\text{observation modes}} \underbrace{\boldsymbol{\Lambda}}_{\text{singular values}} \underbrace{\boldsymbol{V}^\top}_{\text{state modes}}.$$

Thus, the Kalman gain $\boldsymbol{K}$ factorizes as:

$$\boldsymbol{K} = \Sigma_{\mathsf{X},\mathsf{Y}} \Sigma_{\mathsf{Y}}^{-1} = \Sigma_{\mathsf{X}}^{1/2} \boldsymbol{V} \boldsymbol{\Lambda} (\boldsymbol{\Lambda}^2 + \boldsymbol{I}_d)^{-1} \boldsymbol{U}^\top \Sigma_{\mathcal{E}}^{-1/2}.$$

Decomposition of the inference process

The innovation term $(\boldsymbol{y} - \boldsymbol{y}^\star)$ is:
1. **Whitened and rotated**
2. **Assimilated in the informative subspace**
3. **Lifted to the original space**

How can we identify the informative directions that generalize the columns of $\boldsymbol{U}$ and $\boldsymbol{V}$ for nonlinear models?

$$\mathsf{Y} = \boldsymbol{h}(\mathsf{X}) + \mathcal{E}, \; \mathsf{X} \sim \pi_{\mathsf{X}} = \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}), \; \mathcal{E} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$$

Idea: Compute the top $r_{\mathsf{X}}, r_{\mathsf{Y}}$ eigenvectors of the state Gramian $\boldsymbol{C}_{\mathsf{X}} \in \mathbb{R}^{n \times n}$ and observation Gramian $\boldsymbol{C}_{\mathsf{Y}} \in \mathbb{R}^{d \times d}$ using gradient information of $\boldsymbol{h}$ (i.e., $\nabla \boldsymbol{h}$):

- $\boldsymbol{C}_{\mathsf{X}} = \int \nabla \boldsymbol{h}(\boldsymbol{x})^\top \nabla \boldsymbol{h}(\boldsymbol{x}) \mathrm{d}\pi_{\mathsf{X}}(\boldsymbol{x})$ [1]
- $\boldsymbol{C}_{\mathsf{Y}} = \int \nabla \boldsymbol{h}(\boldsymbol{x}) \nabla \boldsymbol{h}(\boldsymbol{x})^\top \mathrm{d}\pi_{\mathsf{X}}(\boldsymbol{x})$ [2]

Algorithm for the low-rank EnKF (LREnKF)

1. Compute the first $r_{\mathsf{X}}, r_{\mathsf{Y}}$ eigenvectors of $\widehat{\boldsymbol{C}}_{\mathsf{X}}, \widehat{\boldsymbol{C}}_{\mathsf{Y}}$ with $r_{\mathsf{X}} \ll n$ and $r_{\mathsf{Y}} \ll d$.
2. Project the states and observation samples $\{\boldsymbol{x}_t^{(i)}, \boldsymbol{y}_t^{(i)}\}$ on these eigenvectors.
3. Compute the Kalman gain $\breve{\boldsymbol{K}}_t \in \mathbb{R}^{r_{\mathsf{X}} \times r_{\mathsf{Y}}}$ in this low-dimensional projected subspace.
4. Lift the result to the original space.

For elliptic inverse problems, a few eigenvectors capture the row/column spaces of $\boldsymbol{K}_t$.

Optimal bias-variance trade-off

We estimate a **lower dimensional** Kalman gain $\breve{\boldsymbol{K}}_t$ in the **span of the informative directions** resulting in a **lower variance estimator** than the stochastic EnKF.

Lagrangian data assimilation in inviscid vortex models

Estimate the positions and strengths of N vortices over time from pressure sensors.

State: positions $\{\boldsymbol{r}_1, \ldots, \boldsymbol{r}_N\}$ and strengths $\{\Gamma_1, \ldots, \Gamma_N\}$ of the point vortices.

Dynamical model (Biot-Savart law): Vortices are advected by the local velocity $\boldsymbol{v}$ given by the curl of the streamfunction ψ.

$$\boldsymbol{v} = \nabla \times (\psi \boldsymbol{e}_z), \text{ with } \nabla^2 \psi = -\boldsymbol{\omega} = -\sum_J \Gamma_J \delta(\boldsymbol{r} - \boldsymbol{r}_J)$$

Observation model: Poisson equation for the pressure p

$$\nabla^2(\underbrace{p}_{\text{unknown}} + \frac{1}{2}\rho ||\boldsymbol{v}(\{\boldsymbol{r}_J\}, \{\Gamma_J\})||^2) = \rho \nabla \cdot (\boldsymbol{v}(\{\boldsymbol{r}_J\}, \{\Gamma_J\}) \times \boldsymbol{\omega}(\{\boldsymbol{r}_J\}, \{\Gamma_J\}))$$

Pressure observations nonlinearly encode information of the (entire state (elliptic PDE).

Inviscid toy vortex model

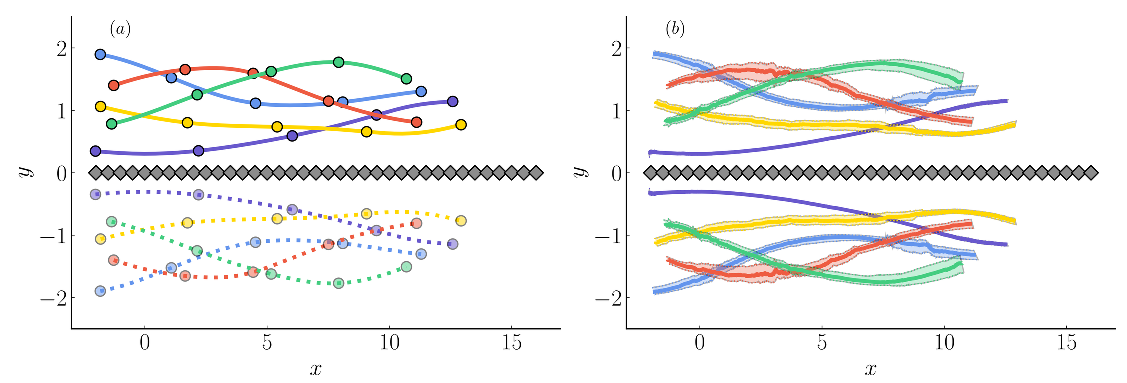


Figure: (a): True trajectories of the point vortices. Pressure sensors are depicted by grey diamonds. (b): Estimated trajectories of the vortices with the LREnKF for $M = 40$ with the ranks r_{X} and r_{Y} set to capture 99% of the cumulative energy spectra.

Adaptive rank selection: Set $r_{\mathsf{X}}, r_{\mathsf{Y}}$ to achieve a threshold $\alpha \in [0, 1]$ for the cumulative energy $E_i = \sum_{j=1}^i \lambda_j^2 / \sum_j \lambda_j^2$ in the eigenvalues λ_j^2 of $\boldsymbol{C}_{\mathsf{X}}$ and $\boldsymbol{C}_{\mathsf{Y}}$

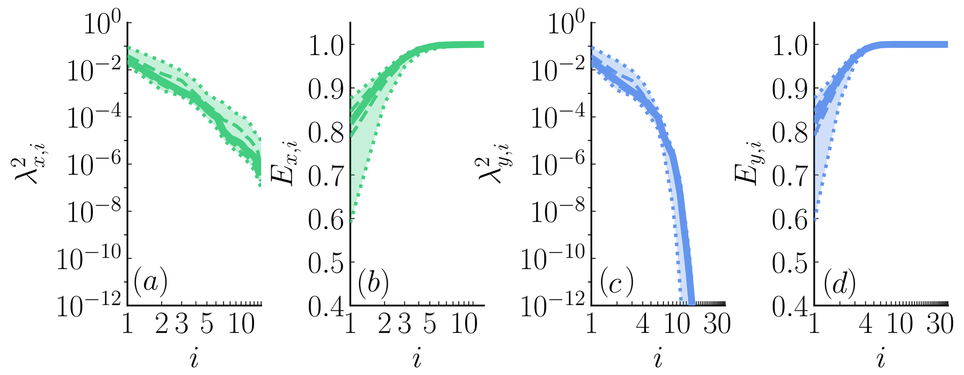


Figure: Panels [(a), (c)]: Median eigenvalues of $\widehat{C}_{\mathsf{X}}$ and $\widehat{C}_{\mathsf{Y}}$. Panels [(b), (d)]: Median normalized cumulative energy E_i of $\widehat{C}_{\mathsf{X}}$ and $\widehat{C}_{\mathsf{Y}}$.

Fast spectral decay of $\widehat{\boldsymbol{C}}_{\mathsf{X}}$ and $\widehat{\boldsymbol{C}}_{\mathsf{Y}}$ confirms the low-rank informative structure.

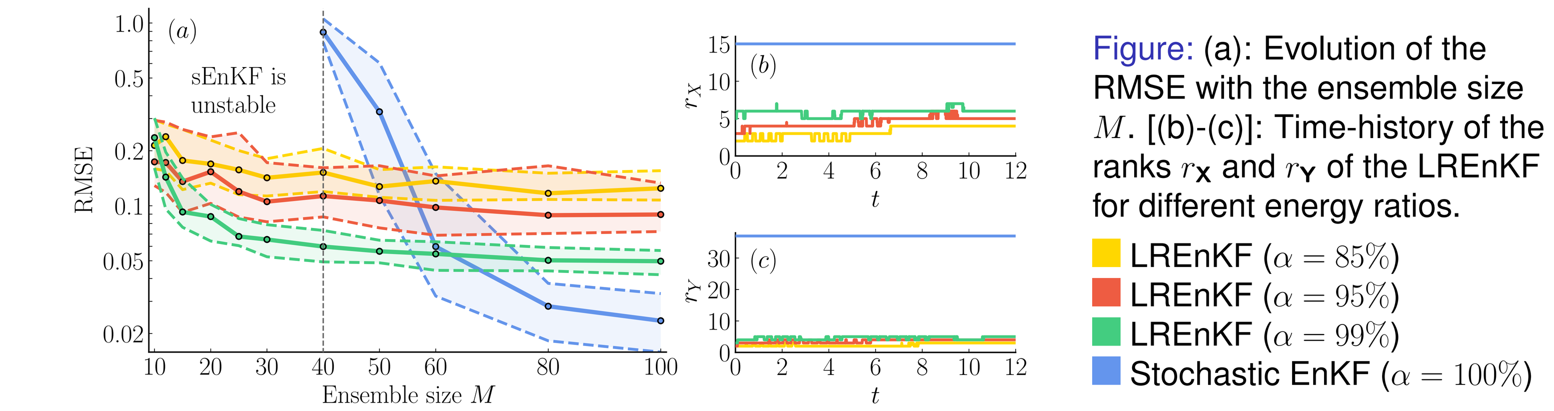
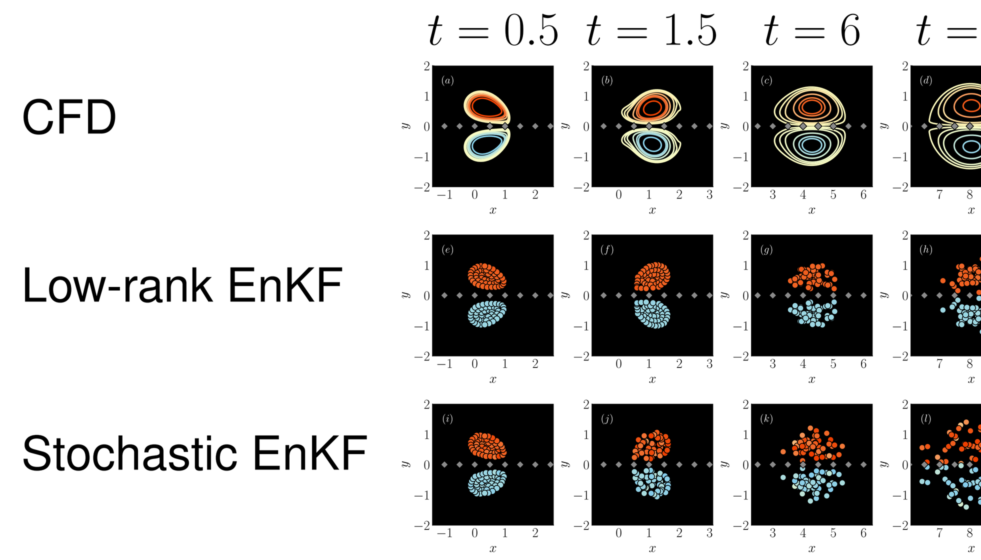
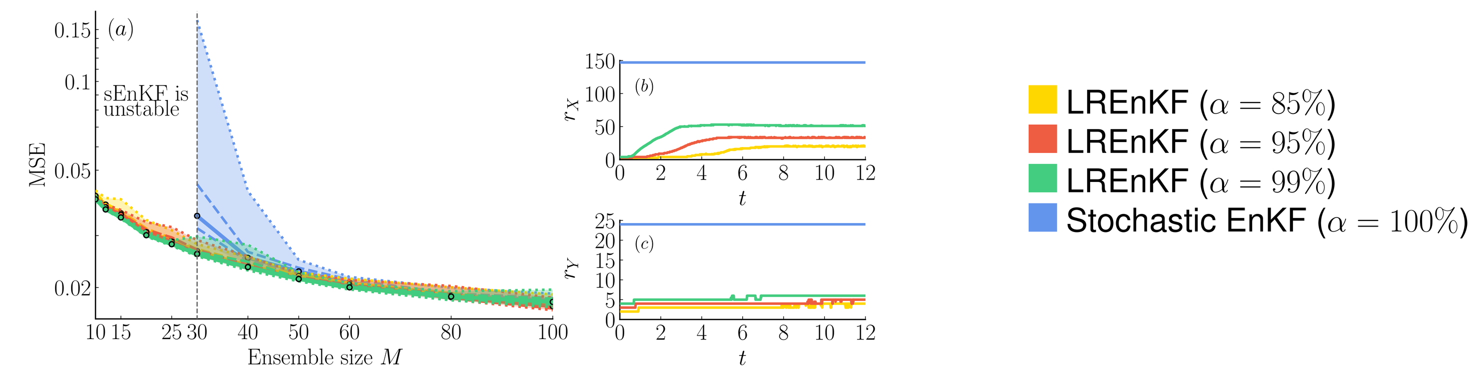


Figure: (a): Evolution of the RMSE with the ensemble size M. [(b)-(c)]: Time-history of the ranks r_{X} and r_{Y} of the LREnKF for different energy ratios.

Counter-rotating vortex patches at $Re = 1000$





Links

Scan me

ArXiv print: https://arxiv.org/abs/2203.05120
Paper in PRSA: https://doi.org/10.1098/rspa.2022.0182
Github repository: https://github.com/mleprovost/lowrankvortex.jl
Correspondence email: mleprovo@mit.edu

References

1. Cui, T. & Zahm, O. Data-free likelihood-informed dimension reduction of Bayesian inverse problems. *Inverse Problems* **37**, 045009 (2021).
2. Le Provost, M., Baptista, R., Marzouk, Y. & Eldredge, J. D. A low-rank ensemble Kalman filter for elliptic observations. *Proceedings of the Royal Society A* **478**, 20220182 (2022).

Acknowledgements

MLP and JE acknowledge support of the AFOSR. MLP acknowledges support of the Dissertation Year Fellowship by the University of California, Los Angeles. RB and YM acknowledge support from the Department of Energy, Office of Advanced Scientific Computing Research, and AEOLUS center.