A kernel extension of the Ensemble Transform Kalman Filter

Sophie Mauran[†], Sandrine Mouysset[‡], Ehouarn Simon[†], Laurent Bertino[*]

[†] Université de Toulouse, INP, IRIT, France
[‡] Université de Toulouse, UT3, IRIT, France
[*] NERSC, Bergen, Norvège
ehouarn.simon@toulouse-inp.fr

9th International Symposium on Data Assimilation
18 October 2023

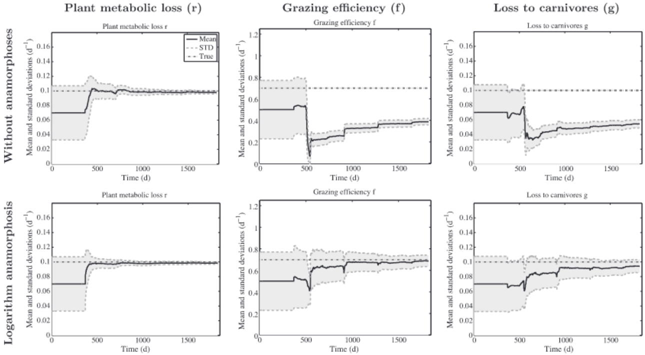
Outline

Outline

Nonlinear transform of the variables?

Gaussian assumptions

- Interpretation in the Bayesian framework: Gaussian assumptions for the EnKF, 4DVar methods...
- Example of combined state-parameter estimation in a 1D ocean ecosystem model



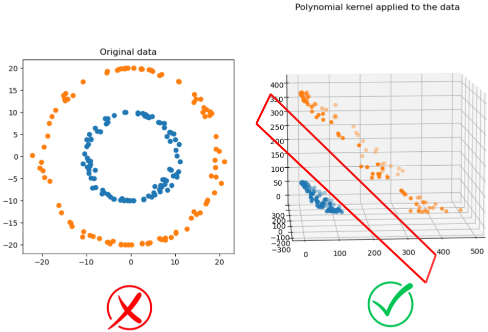
  - Erroneous corrections in parameters with the EnKF.
  - Better results with a nonlinear transformation (logarithm) of the variables.

Kernel strategies in machine learning

Principles

- Mapping data to a higher dimensional space to obtain linear properties.

A SVM example



- Features space: $\phi : \mathcal{X} \rightarrow H$ an application from a set $\mathcal{X}$ to a Hilbert space H.
- Kernel function : $\kappa(x, y) =< \phi(x), \phi(y) >_H$
    - Choice of a function k that implicitly computes the inner product of features.

Outline

ETKF formulation - Bishop et al. (2001)

A linear least square problem

$$\min_{\mathrm{w}\in\mathbb{R}^N} \mathcal{J}(\mathrm{w}) = \frac{N-1}{2}||\mathrm{w}||_2^2 + \frac{1}{2}||\tilde{\mathrm{d}} - \tilde{\mathrm{H}}_{\mathrm{f}}\mathrm{w}||_2^2$$

with N the ensemble size and $\mathrm{x} = \bar{\mathrm{x}}^f + \mathrm{X}_f\mathrm{w}$ the state vector.

Notations

- $\tilde{\mathrm{d}} = \mathrm{R}^{-1/2}(\mathrm{y} - \mathrm{H}\bar{\mathrm{x}}^f) \in \mathbb{R}^p$ the scaled innovation.
- $\tilde{\mathrm{H}}_f = \mathrm{R}^{-1/2}\mathrm{H}\mathrm{X}^{\mathrm{f}}$ the scaled anomaly in observation space.
- $\bar{\mathrm{x}}$ the ensemble mean and $\mathrm{X}_f \in \mathbb{R}^{m\times N}$ the anomaly matrix.

ETKF: an optimization problem on a functional space

Reproducing Kernel Hilbert Space (RKHS)

- Let $\mathcal{H} = \{f : \mathcal{X} \to \mathbb{R}\}$, with the inner product $<,>_{\mathcal{H}}$, be a Hilbert space. $\mathcal{H}$ is a RKHS if

$$\forall x \in \mathcal{X}, \exists! \kappa_x \in \mathcal{H} \text{ s.t. } \forall f \in \mathcal{H}, f(x) =< \kappa_x, f >_{\mathcal{H}}$$

- Reproducing kernel κ:

$$\kappa : \begin{cases} \mathcal{X} \times \mathcal{X} & \to & \mathbb{R} \\ (x, y) & \mapsto & < \kappa_x, \kappa_y >_{\mathcal{H}} \end{cases}$$

- Linear kernel: $\mathcal{X} = \mathbb{R}^n$ and $\kappa(\mathrm{x}, \mathrm{y}) = \mathrm{x}^T \mathrm{y} \quad \forall (\mathrm{x}, \mathrm{y}) \in (\mathbb{R}^n)^2$. Then $\mathcal{H}$ is the dual space of $\mathbb{R}^n$.
- Moore-Aronszajn's theorem: If κ is a symmetric positive definite kernel, then

$$\exists \mathcal{H} \text{ RKHS }, \exists \phi : \mathcal{X} \to \mathcal{H} \text{ s.t. } \kappa(x, y) =< \phi(x), \phi(y) >_{\mathcal{H}} \quad \forall (x, y) \in \mathcal{X}^2.$$

ETKF: an optimization problem on a functional space

An optimization problem on a RKHS

$$\min_{f\in\mathcal{H}_\kappa} \widetilde{\mathcal{J}}(f) = \frac{N-1}{2}||f||^2_{\mathcal{H}_\kappa} + \frac{1}{2}\sum_{i=1}^{p}(f(\tilde{\mathrm{h}}_i) - \tilde{d}_i)^2$$

with $\mathcal{H}_\kappa$ a RKHS and $\tilde{\mathrm{h}}_i$ the line i of $\tilde{\mathrm{H}}_f$.

- The choice of the kernel κ defines the RKHS $\mathcal{H}_\kappa$.
    - Possibly an infinite dimensional vector space.
- Linear kernel: formulation equivalent to the ETKF.
- Implicit transform of the data through the feature map ϕ.

A linear least square problem in finite dimension

- Representation theorem: the minimizers belong to a finite dimensional subspace of $\mathcal{H}_\kappa$.
- Dimension $\sim$ amount of "data":

$$f(.) = \sum_{i=1}^{m} \alpha_i \kappa(., x_i),$$

with $(x_i)_{i=1:m} \in \mathcal{X}^m$.

ETKF reformulation with kernel methods II

A linear least square problem in finite dimension

$$\min_{\alpha \in \mathbb{R}^{p+n}} \bar{\mathcal{J}}(\alpha) = \frac{N-1}{2}\alpha^\top \mathbf{K}\alpha + \frac{1}{2}\|\tilde{\mathrm{d}} - \Pi\mathbf{K}\alpha\|_2^2$$

Notations

- $\Pi = \begin{bmatrix} 0_{nn} & 0_{np} \\ 0_{pn} & I_p \end{bmatrix} \in \mathbb{R}^{(n+p)\times(n+p)}$ projection on observation space
- $\mathbf{K} = \begin{bmatrix} K_X & K_{XH} \\ K_{XH}{}^\top & K_H \end{bmatrix} \in \mathbb{R}^{(n+p)\times(n+p)}$ (this is no Kalman Gain!):

- $K_X = (\kappa(a_i^f, a_j^f))_{1\le i,j\le n} \in \mathbb{R}^{n\times n}$
- $K_{HX} = (\kappa(a_i^f, \widetilde{h_j}))_{1\le i\le n, 1\le j\le p} \in \mathbb{R}^{n\times p}$
- $K_H = (\kappa(\widetilde{h_i}, \widetilde{h_j}))_{1\le i,j\le p} \in \mathbb{R}^{p\times p}$

- $\widetilde{H} = R^{-1/2} H X^f = \begin{bmatrix} \widetilde{h_1}^\top \\ \vdots \\ \widetilde{h_p}^\top \end{bmatrix}$
- $(a_i^f)_{1\le i\le n}$ the lines of X^f

Analysis step

Update of the ensemble mean $\bar{x}^a$

- Solution of the least square problem:

$$\alpha^* = \begin{bmatrix} \alpha^*_{X} \\ \alpha^*_{H} \end{bmatrix} = \begin{bmatrix} 0 \\ [(N-1)I_{n+p} + \Pi K]^{-1}\tilde{d} \end{bmatrix}$$

- Ensemble mean:

$$\bar{x}^a = \bar{x}^f + K_{XH}[(N-1)I_p + K_H]^{-1}\tilde{d}.$$

Symmetric square root of the covariance matrix P^a

- α as random variable:

$$x^a = x^f + \Pi_X K\alpha \quad => \quad P^a = \Pi_X K P^\alpha K \Pi_X,$$

with $\Pi_X \in \mathbb{R}^{n,n+p}$ that selects the components associated with the state vector.

- Approximation of P^α:

$$P^\alpha \sim [\nabla^2 \bar{\mathcal{J}}(\alpha^*)]^{-1} = [(N-1)K + K\Pi^T \Pi K]^{-1}$$

Outline

Experiments with a non-linear kernel

Experimental framework

- Lorenz-63 system with the DAPPER package.
- Observations: first two variables.
- RMSE averaged over 10 different seeds.
- Different inflations ($\{1.0, 1.04, 1.1\}$) and ensemble sizes ($\{3, 6, 10, 12, 15\}$).

Choice of the kernel

- The hyperbolic tangent kernel (Fang et al., 2021):

$$\forall (\mathrm{x}, \mathrm{y}) \in \mathbb{D}_c^N \times \mathbb{D}_c^N, \quad \kappa(\mathrm{x}, \mathrm{y}) = \phi(\mathrm{x})^\top \phi(\mathrm{y})$$

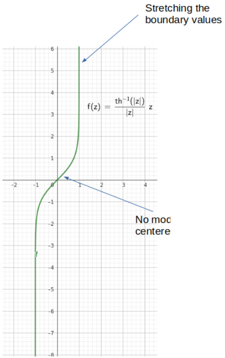
where $\mathbb{D}_c^N = \{\mathrm{z} \in \mathbb{R}^N : c||\mathrm{z}|| < 1\}$ the Poincaré ball and

$$\forall c > 0, \;\; \forall \mathrm{z} \in \mathbb{D}_c^N, \quad \phi(\mathrm{z}) = tanh^{-1}(\sqrt{c}||\mathrm{z}||)\frac{\mathrm{z}}{\sqrt{c}||\mathrm{z}||}$$

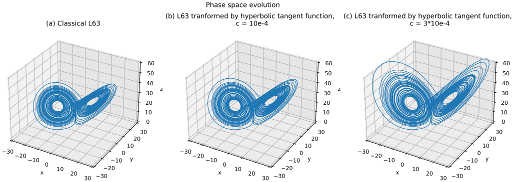
- Choice: $c = 1.e{-}4$.

Effect of hyperbolic tangent kernel on L63

Hyperbolic tangent function



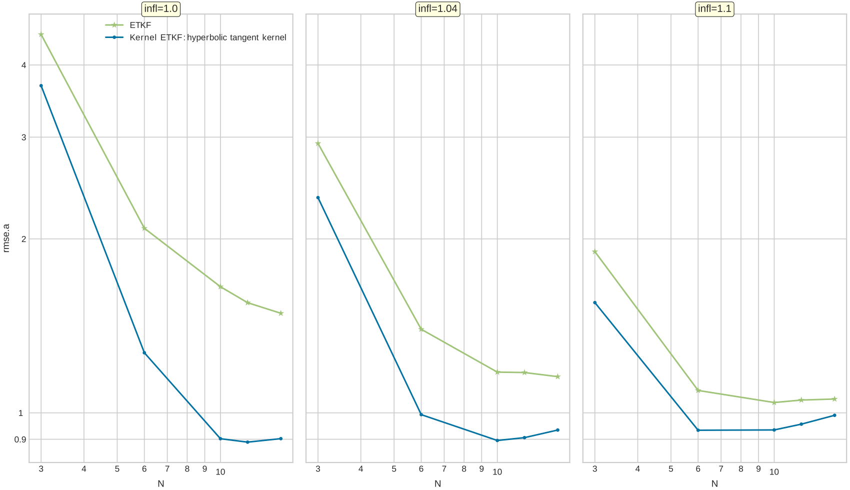
Impact on the representation of the dynamics



- Stretching of the wings in the phase space with high sensitivity to hyperparameter c.

Average RMSE for different ensemble sizes

ETKF (green) vs Kernel ETKF (blue)



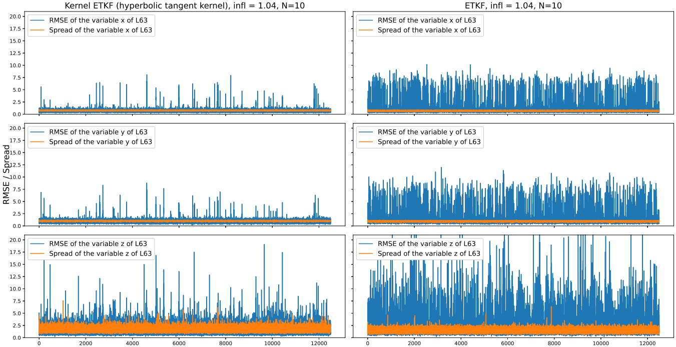
- Better results for the Kernel-ETKF.
- Smaller ensemble size compared to the ETKF?

Focus on one experimental framework: N=10, infl = 1.04

RMSE and standard deviation (STD) averaged over 10 experiments

	ETKF		Kernel ETKF	
L63 variable	RMSE	STD	RMSE	STD
x	1.1 ± 1.0	0.68 ± 0.062	0.69 ± 0.28	0.75 ± 0.057
y	1.4 ± 1.1	0.93 ± 0.082	0.93 ± 0.32	1.0 ± 0.069
z	2.2 ± 1.7	1.4 ± 0.37	1.5 ± 0.82	1.7 ± 0.43

Single experiment: time evolution of RMSE and STD



Outline

Conclusion

- Generalisation of the ETKF problem by introducing kernels.
- Experiments:
  - Similar performances for the linear kernel ETKF and classical ETKF (as expected).
  - Interest of using other kernels in the presence of nonlinearities.

Perspectives

- Adaptive localisation of the Kernel ETKF.
- Dealing with the size of $\mathrm{K} \in \mathbb{R}^{(n+p)\times(n+p)}$: low-rank approximation methods (Cai et al., 2022).

Thank you !

Bibliography

- **D. Cai, J. Nagy, Y. Xi**: Fast deterministic approximation of symmetric indefinite kernel matrices with high dimensional datasets, *SIAM Journal on Matrix Analysis and Applications*, 43:2, 1003-1028, 2022.
- **P. Fang, M. Harandi, L. Petersson**: Kernel methods in Hyperbolic spaces, *ICCV 2021 - IEEE/CVF International Conference on Computer Vision (ICCV 2021)*, 10665-10674, 2021.
- **A. Farchi, M. Bocquet**: On the Efficiency of Covariance Localisation of the Ensemble Kalman Filter Using Augmented Ensembles, *Frontiers in Applied Mathematics and Statistics*, 5, 2019.
- **C.H. Bishop, B.J. Etherton, S.J. Majumdar**: Adaptive Sampling with the Ensemble Transform Kalman Filter. Part I: Theoretical Aspects, *Monthly Weather Review*, 129, 420-436, 2001.
- **S. Mauran, S. Mouysset, E. Simon, L. Bertino**: A kernel extension of the Ensemble Transform Kalman Filter, *ICCS 2023 - International Conference on Computational Science (ICCS 2023)*, 438-452, 2023.
- **E. Simon, L. Bertino**: Gaussian anamorphosis extension of the DEnKF for combined state parameter estimation: Application to a 1D ocean ecosystem model, *Journal of Marine Systems*, 89, 1-18, 2012.

KETKF analysis

Algorithm 1 Kernel ETKF analysis

$\tilde{\mathrm{H}} \leftarrow \mathrm{R}^{-1/2}\mathrm{H}\mathrm{X}^f$, $\tilde{\mathrm{d}} \leftarrow \mathrm{R}^{-1/2}(\mathrm{y} - \mathrm{H}\bar{\mathrm{x}}^f)$

Compute the kernel matrix K.

$\alpha_H^* = [(N-1)\mathrm{I}_p + \mathrm{K}_H]^{-1}\tilde{\mathrm{d}}$ % *Solve a linear system of a SPD matrix*

$\bar{\mathrm{x}}^a = \bar{\mathrm{x}}^f + \mathrm{K}_{XH}\alpha_H^*$

Compute Σ, U the singular values and vectors of P_X^a and truncate Σ to its rank r_Σ and compute its square root: $\tilde{\Sigma}^{1/2}$

${\mathrm{P}_X^a}^{1/2} \leftarrow \tilde{\mathrm{U}}\tilde{\Sigma}^{1/2}$ with $\tilde{\mathrm{U}}$, the first r_Σ columns of U

for i = 1... N-r_Σ **do**

  rotate ${\mathrm{P}_X^a}^{1/2}$ following Farchi and Bocquet (2019)

end for

$\mathrm{E} = \bar{\mathrm{x}}^a + \sqrt{N-1}{\mathrm{P}_X^a}^{1/2}$