A kernel extension of the Ensemble Transform Kalman Filter

S. Mauran [1] L. Bertino [2] S. Mouysset [3] E. Simon [1]

[1]Université de Toulouse, INP, IRIT
Toulouse, France

[2]Nansen Environmental and Remote Sensing Center
Bergen, Norway

[3]Université de Toulouse, UT3, IRIT
Toulouse, France

EnKF Workshop, May 2023

Table of Contents

Table of Contents

Background and key issues

- Interpretation in the Bayesian framework: Gaussian assumptions of the EnKF, 4DVar methods...
- Limits:[1]

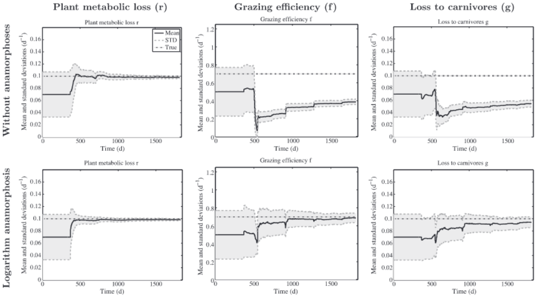

Figure: Estimation of the three parameters (r,f,g) of an 1D ecosystem model. Time evolution of the mean and the mean plus/minus the standard deviation of the ensemble for the three estimated parameters.

- Introduction of kernel methods: **Data linearisation property**

[1] SB12.

Table of Contents

Table of Contents

ETKF formulation and classical resolution I

Assuming the observation operator is linear, the ETKF problem[2] reads:

ETKF formulation

$$\arg\min_{\mathbf{w}\in\mathbb{R}^N} \mathcal{J}(\mathbf{w}) = \frac{N-1}{2}||\mathbf{w}||_2^2 + \frac{1}{2}||\mathbf{y} - \mathbf{H}\bar{\mathbf{x}}^{\mathbf{f}} - \mathbf{H}\mathbf{X}^{\mathbf{f}}\mathbf{w}||_{\mathbf{R}^{-1}}^2 \quad (1)$$

Notations

- Observations model: $y_k = \mathbf{H}x_k + \epsilon_k$, $\mathbf{H} \in \mathbb{R}^{n\times m}$ and ϵ_k observations error, with covariance matrix $\mathbf{R}$
- $\bar{\boldsymbol{x}}$ mean of ensemble states
- $\boldsymbol{X_f} \in \mathbb{R}^{m\times N}$ anomaly matrix of centred states

[2] HKS07.

ETKF formulation and classical resolution II

For the sake of later clarity, we introduce some additional notations:

Notations

- $\widetilde{\mathbf{d}} = \mathbf{R}^{-1/2}(\mathbf{y} - \mathbf{H}\overline{\mathbf{x}}^{\mathbf{f}}) \in \mathbb{R}^p$
- $\widetilde{\mathbf{H}} = \mathbf{R}^{-1/2}\mathbf{H}\mathbf{X}^{\mathbf{f}} = \begin{bmatrix} \widetilde{\mathbf{h_1}}^\top \\ \vdots \\ \widetilde{\mathbf{h_p}}^\top \end{bmatrix} \in \mathbb{R}^{p \times N}.$

ETKF solution (First Order Condition)

$$\mathbf{w}^* = [(N-1)\mathbf{I} + \widetilde{\mathbf{H}}^\top \widetilde{\mathbf{H}}]^{-\mathbf{1}} \widetilde{\mathbf{H}}^\top \widetilde{\mathbf{d}} \tag{2}$$

Table of Contents

ETKF reformulation with kernel methods I

The ETKF problem (1) is equivalent to the extented optimisation problem in the Reproducing Kernel Hilbert Space (RKHS) $\mathcal{H}_\kappa$ of reproducing kernel $k(x, y) = x^\top y$ (linear kernel):

Reformulating the ETKF problem

$$\underset{f \in \mathcal{H}_\kappa}{\arg\min}\, \widetilde{\mathcal{J}}(f) = \frac{N-1}{2} ||f||^2_{\mathcal{H}_\kappa} + \frac{1}{2} \sum_{i=1}^{p} (f(\widetilde{\mathbf{h}_\mathbf{i}}) - \widetilde{d}_i)^2 \tag{3}$$

with $f \in \mathcal{H}_\kappa$ such as: $f : \begin{cases} \mathbb{R}^p & \rightarrow & \mathbb{R} \\ \mathbf{x} & \mapsto & \kappa(\mathbf{x}, \mathbf{w}) \end{cases}$

ETKF reformulation with kernel methods II

Here, we apply the representation theorem and obtain the following formulation of the ETKF:

Reformulating the ETKF

$$\arg\min_{\alpha\in\mathbb{R}^{p+n}} \widetilde{\widetilde{\mathcal{J}}}(\alpha) = \frac{N-1}{2}\alpha^\top \mathbf{K}\alpha + \frac{1}{2}\|\widetilde{\mathbf{d}} - \mathbf{\Pi_H}\mathbf{K}\alpha\|_2^2 \tag{4}$$

Notations

- $\mathbf{\Pi_H} = \begin{bmatrix} \mathbf{0_{nn}} & \mathbf{0_{np}} \\ \mathbf{0_{pn}} & \mathbf{I_p} \end{bmatrix} \in \mathbb{R}^{(n+p)\times(n+p)}$ the projection matrix on the observation space
- $\mathbf{K} = \begin{bmatrix} \mathbf{K_X} & \mathbf{K_{XH}} \\ \mathbf{K_{XH}}^\top & \mathbf{K_H} \end{bmatrix} \in \mathbb{R}^{(n+p)\times(n+p)}$ with
  - $\mathbf{K_X} = (\kappa(\mathbf{a_i^f}, \mathbf{a_j^f}))_{1\le i,j\le n} \in \mathbb{R}^{n\times n}$
  - $\mathbf{K_{HX}} = (\kappa(\mathbf{a_i^f}, \widetilde{\mathbf{h}}_\mathbf{j}))_{1\le i\le n, 1\le j\le p} \in \mathbb{R}^{n\times p}$
  - $\mathbf{K_H} = (\kappa(\widetilde{\mathbf{h}}_\mathbf{i}, \widetilde{\mathbf{h}}_\mathbf{j}))_{1\le i,j\le p} \in \mathbb{R}^{p\times p}$

ETKF reformulation with kernel methods III

Resolution of (4) (First Order Condition)

$$\alpha^* = \begin{bmatrix} \alpha^*_{\mathbf{X}} \\ \alpha^*_{\mathbf{H}} \end{bmatrix} = \begin{bmatrix} \mathbf{0_{n1}} \\ [(N-1)\mathbf{I_{n+p}} + \mathbf{\Pi_H K}]^{-1}\widetilde{\mathbf{d}} \end{bmatrix} \tag{5}$$

Thus, the mean of the ensemble after the analysis will be:

Ensemble mean after analysis

$$\bar{\mathbf{x}}^{\mathbf{a}} = \bar{\mathbf{x}}^{\mathbf{f}} + \mathbf{K_{XH}}[(N-1)\mathbf{I_p} + \mathbf{K_H}]^{-1}\widetilde{\mathbf{d}} \tag{6}$$

Table of Contents

Since we work both with observed and unobserved variables, we followed the strategy implemented in [Eve09] and extended the deterministic algorithm of the EnKF:

EnKF extented to unobserved variables

$$\begin{bmatrix} \mathbf{E^a} \\ \mathbf{HE^a} \end{bmatrix} = \begin{bmatrix} \mathbf{\bar{x}^f} \\ \mathbf{H\bar{x}^f} \end{bmatrix} + \begin{bmatrix} \mathbf{X^f} \\ \mathbf{HX^f} \end{bmatrix} \mathbf{w} + \sqrt{N-1}\mathbf{P^a}^{1/2} \tag{7}$$

with $\mathbf{P^a} = \begin{bmatrix} \mathbf{P^a_X} & \mathbf{P^a_{XH}} \\ \mathbf{P^a_{XH}}^\top & \mathbf{P^a_H} \end{bmatrix} \in \mathbb{R}^{(n+p)\times(n+p)}$ the analysis error covariance matrix.

Ensemble's reconstruction II: Determining $\mathbf{P_X^a}$

- Perspective of random variables: $\alpha \sim \mathcal{N}(\mu_\alpha, \mathbf{P}^\alpha)$
- Expression of ensemble's mean: $\mathbf{x^a} = \bar{\mathbf{x}}^\mathbf{f} + \mathbf{\Pi_X K}\alpha$

Expression of $\mathbf{P_X^a}$

$$\mathbf{P_X^a} = Cov(\mathbf{x_i^a}, \mathbf{x_j^a}) = \mathbf{\Pi_X K P^\alpha K \Pi_X} \quad (8)$$

We can besides approximate[3] $\mathbf{P}^\alpha$:

Approximation of $\mathbf{P}^\alpha$

$$\mathbf{P}^\alpha \approx [\nabla^\mathbf{2} \tilde{\tilde{\mathcal{J}}}]^{-1} \quad (9)$$

with

$$\tilde{\tilde{\mathcal{J}}}(\alpha) = \frac{N-1}{2}\alpha^\top \mathbf{K}\alpha + \frac{1}{2}\|\tilde{\mathbf{d}} - \mathbf{\Pi_H K}\alpha\|_2^2$$

[3] Aur03.

Table of Contents

Table of Contents

Experimental setup

- L63 model
- experiments performed with the DAPPER package
- 2 sets of experiments:
  - Comparison between classical ETKF and linear Kernel ETKF:

$$\forall(\mathbf{x},\mathbf{y}) \in \mathbb{R}^N \times \mathbb{R}^N, \quad \kappa(\mathbf{x},\mathbf{y}) = \mathbf{x}^\top \mathbf{y} \tag{10}$$

  - Comparison between classical ETKF and non linear Kernel ETKF (hyperbolic tangent kernel[4] with $c = 10^{-4}$):

$$\forall(\mathbf{x},\mathbf{y}) \in \mathbb{D}_c^N \times \mathbb{D}_c^N, \quad \kappa(\mathbf{x},\mathbf{y}) = \phi(\mathbf{x})^\top \phi(\mathbf{y}) \tag{11}$$

    where $\mathbb{D}_c^N$ is the Poincaré ball:

$$\mathbb{D}_c^N = \{z \in \mathbb{R}^N : c||z|| < 1\} \tag{12}$$

    and

$$\forall c > 0, \;\; \forall \mathbf{z} \in \mathbb{D}_c^N, \quad \phi(\mathbf{z}) = tanh^{-1}(\sqrt{c}||z||)\frac{z}{\sqrt{c}||z||} \tag{13}$$

- For all experiments, RMSE averaged over 10 different seeds

[4] FHP21.

Table of Contents

Specific setup

- We observe only the first two variables
- The observations are generated every $\delta to = 0.02$
- 5×10^5 observations vectors generated for each experiment with a burn-in period of $5 \times 10^3 \times \delta to$
- We compare different inflation factors: $infl \in \{1.0, 1.04, 1.1\}$ For each inflation factor, the evaluated ensemble sizes were $N \in \{3, 6, 9, 10, 12, 15\}$.

Results

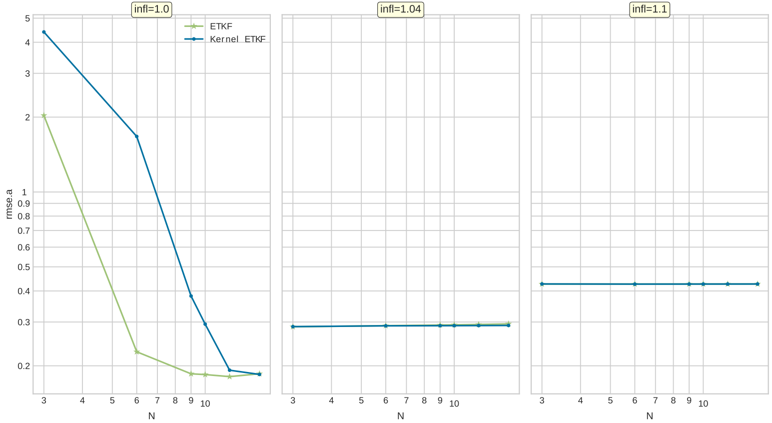

Figure: Average RMSE obtained by ETKF (in green) and linear Kernel ETKF (in blue) assimilation methods when applied to the Lorenz 63 model and observing only the first two variables. The average is computed upon 10 different seeds generating observations, initial state... On each subfigure, a different inflation factor is applied to each method: Left: no inflation (infl = 1.0), Middle: infl = 1.04, Right: infl = 1.1. In each subfigure, different ensemble sizes N were tested, in each case $N \in \{3, 6, 9, 10, 12, 15\}$.

Table of Contents

Effect of hyperbolic tangent kernel on L63 I

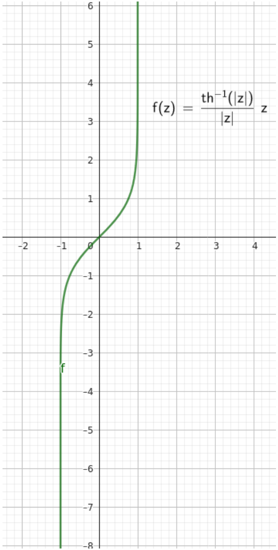

Figure: Hyperbolic tangent fonction

Effect of hyperbolic tangent kernel on L63 II

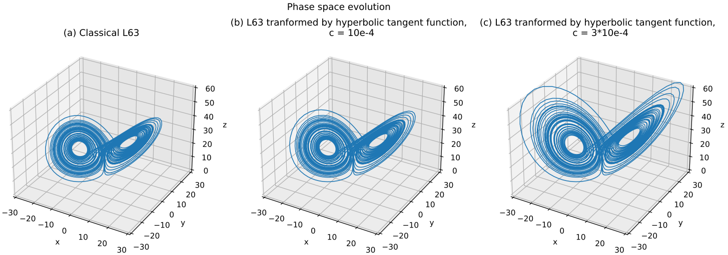

Figure: Phase space evolution of the Lorenz 63 model: (a) classical L63; (b) L63 tranformed by hyperbolic tangent fonction (13) with $c = 10^{-4}$; (c) L63 tranformed by hyperbolic tangent fonction (13) with $c = 3 \times 10^{-4}$ to accentuate the visual effect

Effect of hyperbolic tangent kernel on L63 III

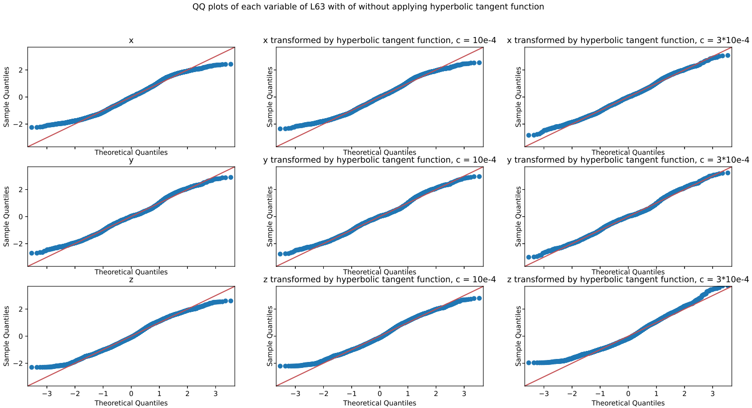

Figure: QQ plots of each variable of the L63 relative to the normal distribution, Left: classical L63; Middle: L63 tranformed by hyperbolic tangent fonction (13) with $c = 10^{-4}$; Right: L63 tranformed by hyperbolic tangent fonction (13) with $c = 3 \times 10^{-4}$ to accentuate the visual effect

Specific setup

- 2 experiments, one observing all variables, one observing only 2 variables.
- **Non linearity reinforcement:** $\delta to = 0.50$ when all variables are observed; $\delta to = 0.25$ in the second experiment.
- 2×10^4 observation vectors and a burn-in period of $2 \times 10^2 \times \delta to$ when all variables are observed; 5×10^4 observation vectors and a burn-in period of $5 \times 10^2 \times \delta to$ when only 2 variables are observed.
- We compare different inflation factors: $infl \in \{1.0, 1.04, 1.1\}$ For each inflation factor, the evaluated ensemble sizes were $N \in \{3, 6, 10, 12, 15\}$.

Results when all variables are observed

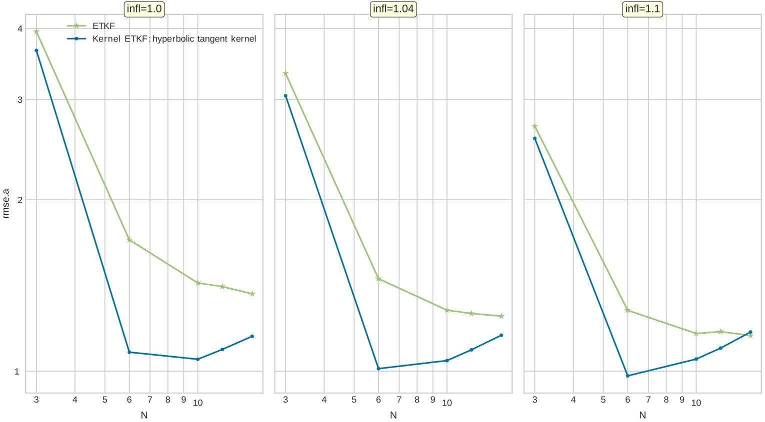

Figure: Average RMSE obtained by ETKF (in green) and Kernel ETKF applied to hyperbolic tangent kernel with $c = 10^{-4}$ (in blue) when applied to the Lorenz 63 model and observing all variables. The average is computed upon 10 different seeds generating observations, initial state... Left: $infl = 1.0$. Middle: $infl = 1.04$. Right: $infl = 1.1$. For each subfigure, different ensemble sizes are tested: $N \in \{3, 6, 10, 12, 15\}$.

Results when the 2 first variables are observed I

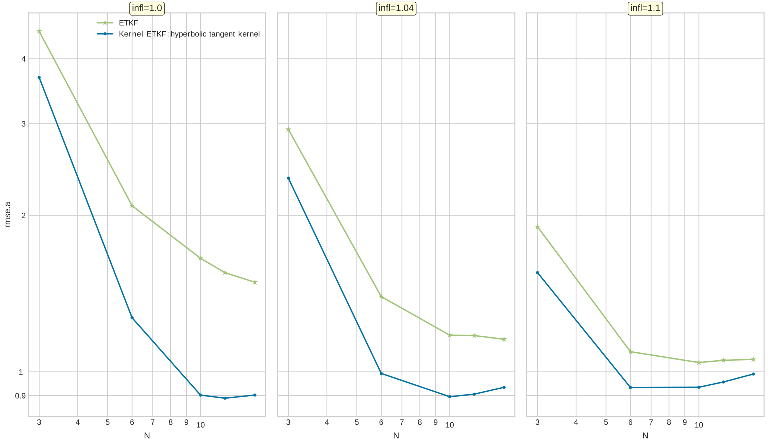

Figure: Average RMSE obtained by ETKF (in green) and Kernel ETKF applied to hyperbolic tangent kernel with $c = 10^{-4}$ (in blue) when applied to the Lorenz 63 model and observing the first two variables. The average is computed upon 10 different seeds generating observations, initial state... Left: $infl = 1.0$. Middle: $infl = 1.04$. Right: $infl = 1.1$. In each one, different ensemble sizes are tested: $N \in \{3, 6, 10, 12, 15\}$.

Results when the 2 first variables are observed II

	ETKF		Kernel ETKF	
L63 variable	RMSE	Spread	RMSE	Spread
x	1.1 ± 1.0	0.68 ± 0.062	$\mathbf{0.69 \pm 0.28}$	0.75 ± 0.057
y	1.4 ± 1.1	0.93 ± 0.082	$\mathbf{0.93 \pm 0.32}$	1.0 ± 0.069
z	2.2 ± 1.7	1.4 ± 0.37	$\mathbf{1.5 \pm 0.82}$	1.7 ± 0.43

Table: RMSE and Spread of each variables for the hyperbolic tangent Kernel ETKF and the classical ETKF when applied to the Lorenz 63 model and observing the first two variables variables in the case where $N = 10$, $infl = 1.04$.

Some validation

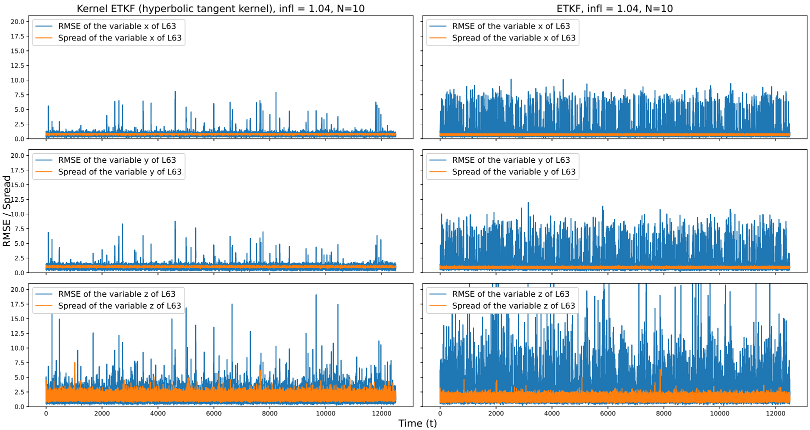

Figure: Ensemble spread (in orange) and RMSE (in blue) obtained by hyperbolic tangent Kernel ETKF (left panel) and classical ETKF (right panel) assimilation methods when applied to the Lorenz 63 model and observing the first two variables variables in the case where $N = 10$, $infl = 1.04$. The ensemble spread and RMSE are displayed for each three variables of the L63 model individually

Table of Contents

Conclusion & Perspectives

- Generalisation of the ETKF problem by introducing kernels
- Explicit algorithm of the KETKF
- Experiments:
  - Similar performances for the linear kernel ETKF and classical ETKF (as expected)
  - Interest of using other kernels in the presence of strong nonlinearties with the results of the hyberbolic tangent KETKF on small ensemble sizes
- **Proceeding paper for MLDADS (ICCS) 2023**
- **Perspectives:** Integration of localisation to the KETKF

Institut de Recherche en Informatique de Toulouse
CNRS - INP - UT3 - UT1 - UT2J

Table of Contents

Bibliography I

[Aur03] Didier Auroux. *Hessienne et covariances d'erreur*. 2003. URL: https://math.unice.fr/~auroux/Work/These/html/node40.html. (accessed: 29.08.2021).

[Eve09] Geir Evensen. *Data Assimilation : The Ensemble Kalman Filter*. Springer, 2009.

[FB19] Alban Farchi and Marc Bocquet. "On the Efficiency of Covariance Localisation of the Ensemble Kalman Filter Using Augmented Ensembles". In: *Frontiers in Applied Mathematics and Statistics* 5 (2019). ISSN: 2297-4687. DOI: 10.3389/fams.2019.00003. URL: https://www.frontiersin.org/articles/10.3389/fams.2019.00003.

[FHP21] Pengfei Fang, Mehrtash Harandi, and Lars Petersson. "Kernel Methods in Hyperbolic Spaces". In: (Oct. 2021), pp. 10665–10674.

Bibliography II

[HKS07] Brian R. Hunt, Eric J. Kostelich, and Istvan Szunyogh. "Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter". In: *Physica D: Nonlinear Phenomena* 230.1 (2007). Data Assimilation, pp. 112–126. ISSN: 0167-2789. DOI: https://doi.org/10.1016/j.physd.2006.11.008. URL: https://www.sciencedirect.com/science/article/pii/S0167278906004647.

Bibliography III

[SB12] Ehouarn Simon and Laurent Bertino. "Gaussian anamorphosis extension of the DEnKF for combined state parameter estimation: Application to a 1D ocean ecosystem model". In: *Journal of Marine Systems* 89.1 (2012), pp. 1–18. ISSN: 0924-7963. DOI: https://doi.org/10.1016/j.jmarsys.2011.07.007. URL: https://www.sciencedirect.com/science/article/pii/S0924796311001576.

Table of Contents

KETKF analysis

Algorithm Kernel ETKF analysis

$\widetilde{\mathbf{H}} \leftarrow \mathbf{R}^{-1/2}\mathbf{H}\mathbf{X}^{\mathbf{f}}$

$\widetilde{\mathbf{d}} \leftarrow \mathbf{R}^{-1/2}(\mathbf{y} - \mathbf{H}\bar{\mathbf{x}}^{\mathbf{f}})$

Compute $\mathbf{K}$ ▷ depends on the chosen kernel

$\alpha^*_{\mathbf{H}} = [(N-1)\mathbf{I_p} + \mathbf{K_H}]^{-1}\widetilde{\mathbf{d}}$ ▷ Solve a linear system of a SPD matrix

$\bar{\mathbf{x}}^{\mathbf{a}} = \bar{\mathbf{x}}^{\mathbf{f}} + \mathbf{K_{XH}}^{\top}\alpha^*_{\mathbf{H}}$

Compute $\mathbf{\Sigma}$, $\mathbf{U}$ the singular values and vectors of $\mathbf{P^a_X}$ ▷ refer to **Algorithm 2**

Truncate $\mathbf{\Sigma}$ to its rank r_Σ and compute its square root: $\widetilde{\mathbf{\Sigma}}^{1/2}$

${\mathbf{P^a_X}}^{1/2} \leftarrow \widetilde{\mathbf{U}}\widetilde{\mathbf{\Sigma}}^{1/2}$ ▷ with $\widetilde{\mathbf{U}}$, the first r_Σ columns of $\mathbf{U}$

for i = 1... N-r_Σ **do**

  rotate $\mathbf{P^a_X}^{1/2}$ following the rotation step of **Algorithm 3**

end for

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

Computation of $\mathbf{P_X^a}$

Algorithm Computation of $\mathbf{P_X^a}$

if K is not invertible **then**

Compute $\nabla^2 \widetilde{\widetilde{\mathcal{J}}} \leftarrow [(N-1)\mathbf{K} + \mathbf{K}\mathbf{\Pi_H}\mathbf{\Pi_H}\mathbf{K}]$

Compute the SVD of $\nabla^2 \widetilde{\widetilde{\mathcal{J}}} = \mathbf{U}_{\mathcal{J}}\mathbf{\Sigma}_{\mathcal{J}}\mathbf{V}_{\mathcal{J}}^\top$

Compute $\widetilde{\mathbf{U}_{\mathcal{J}}}$, $\widetilde{\mathbf{\Sigma}_{\mathcal{J}}}$ and $\widetilde{\mathbf{V}_{\mathcal{J}}}^\top$ the respective matrix of $\mathbf{U}_{\mathcal{J}}, \mathbf{\Sigma}_{\mathcal{J}}, \mathbf{V}_{\mathcal{J}}^\top$ truncated at the rank of $\mathbf{K}$

$\mathbf{\Sigma} \leftarrow \widetilde{\mathbf{\Sigma}_{\mathcal{J}}}^{-1}$

Compute $\mathbf{U} \leftarrow \mathbf{\Pi_X}\mathbf{K}\widetilde{\mathbf{U}_{\mathcal{J}}}$ $\triangleright$ $\mathbf{P_X^a}^{1/2} = \mathbf{U}\mathbf{\Sigma}^{1/2}$

else

Compute $\mathbf{P_X^a} \leftarrow \mathbf{K_X} - \mathbf{K_{HX}}\mathbf{U_H}\,diag(\frac{1}{(N-1)+\lambda_i})\mathbf{U_H}^\top\mathbf{K_{HX}}^\top$ $\triangleright$ with $[\lambda_i]_{1\leq i\leq n}$ the eigenvalues of $\mathbf{K_H}$

Compute the SVD of $\mathbf{P_X^a} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$

end if

Rotation step

Algorithm Rotation step of $\mathbf{P}_{\mathbf{X}}^{\mathbf{a}^{1/2}}$, directly derived from Annex A of [FB19]

Require: $1 \leq i \leq N - r_\Sigma$

$\epsilon \leftarrow 1.0$

Compute $c \leftarrow r_\Sigma + i$

Compute $\theta \leftarrow \frac{\sqrt{c}}{\sqrt{c-\epsilon}}$

Compute $\mathbf{Q}_\epsilon \leftarrow \frac{-\theta}{c} \times \begin{bmatrix} \frac{\epsilon}{\sqrt{c}} & \cdots & \cdots & \cdots & \cdots & \frac{\epsilon}{\sqrt{c}} \\ \vdots & 1-\frac{\theta}{c} & \frac{-\theta}{c} & \cdots & \cdots & \frac{-\theta}{c} \\ \vdots & \frac{-\theta}{c} & 1-\frac{\theta}{c} & \frac{-\theta}{c} & \cdots & \frac{-\theta}{c} \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 1-\frac{\theta}{c} & \frac{-\theta}{c} \\ \frac{\epsilon}{\sqrt{c}} & \frac{-\theta}{c} & \cdots & \cdots & \frac{-\theta}{c} & 1-\frac{\theta}{c} \end{bmatrix} \in \mathbb{R}^{c \times c}$

$\mathbf{W} \leftarrow \begin{bmatrix} \mathbf{0_n} & \mathbf{P}_{\mathbf{X}}^{\mathbf{a}^{1/2}} \end{bmatrix} \in \mathbb{R}^{n \times c}$

Compute $\mathbf{P}_{\mathbf{X}}^{\mathbf{a}^{1/2}} \leftarrow \mathbf{W}\mathbf{Q}_\epsilon$