



HAL
open science

A New Low Rank Tensor Format for High Dimensional Sparse Regression

Rémi Cloarec, Sebastian Rodriguez, Lila Achour, Xavier Kestelyn, Francisco Chinesta

► **To cite this version:**

Rémi Cloarec, Sebastian Rodriguez, Lila Achour, Xavier Kestelyn, Francisco Chinesta. A New Low Rank Tensor Format for High Dimensional Sparse Regression. *Mortech 2025, 7th International Workshop on Model Order Reduction Techniques*, Nov 2025, Zaragoza (Universidad de Zaragoza), Spain. <hal-05546159>

HAL Id: hal-05546159

<https://hal.science/hal-05546159v1>

Submitted on 10 Mar 2026

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

A New Low Rank Tensor Format for High Dimensional Sparse Regression

R. Cloarec^{*1}, S. Rodriguez¹, L. Achour¹, X. Kestelyn², F. Chinesta¹

¹RTE Chair, PIMM Lab, Arts et Métiers Institute of Technology, 151 Boulevard de l'Hôpital, 75013 Paris, France

²RTE Chair, ULR2697-L2EP, Centrale Lille, Junia ISEN Lille, Arts et Métiers Institute of Technology, University of Lille, 59000 Lille, France

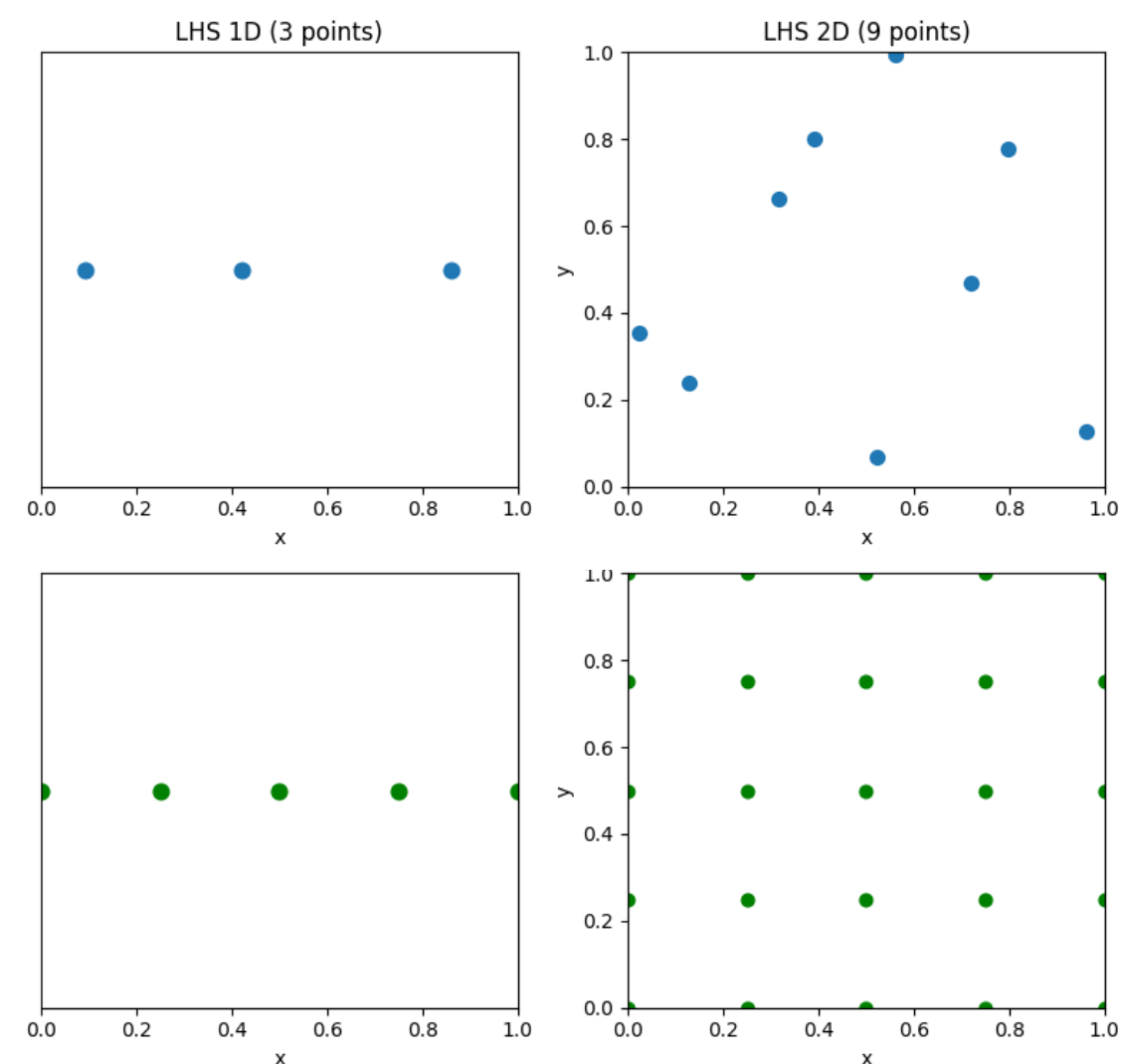
*Contact: remi.cloarec@ensam.eu



Keywords: multilinear algebra, tensor format, high dimensional regression

Context:

We work with surrogates that require an *explicit formulation in terms of input parameters*. No matter how much we reduce the dimensionality of the problem, that formulation constitutes a computational bottleneck.



In general, we work with a limited amount of data, therefore we constrain the structure of the equations along a *sampling axis*.

The equation we build at each sampled point preserves the full structure, unreachable by sampling in high dimensions.

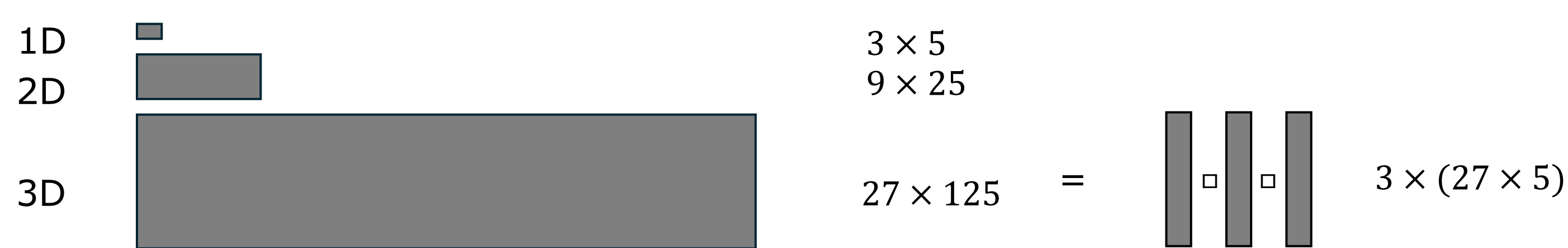
We leverage the face-splitting product to cover combinatorial structures in a factorized manner. For instance, with polynomial models, we can build from separated one-dimensional structures:

$$[1, x_1, x_1^2] \otimes [1, x_2, x_2^2] = [1, x_1, x_2, x_1^2, x_2^2, x_1 x_2, x_1^2 x_2^2].$$

The linear system to solve is then highly structured $Ax = b$

$$\begin{bmatrix} \dots & \dots & \dots \\ \dots & \dots & \dots \\ \dots & \dots & \dots \end{bmatrix} = \begin{bmatrix} \dots \\ \dots \\ \dots \end{bmatrix}$$

That allows us to prevent the explosion in size of A by using a factorization:



We generalize the *face-splitting product* in as a tensor operation: m-tensor product \otimes

Tensor regression in m-tensor format

We seek an approximation $\hat{f}(x) = \varphi(x) : C$ to a N dimensional function f evaluated at m arbitrarily sampled points $\{x_1, x_2, x_3, \dots, x_m\}$ such that \hat{f} minimizes $\|f - \hat{f}\|_2^2$.

1°) Chose the 1D bases, for example

$$\psi(x) = [1, x, x^2, \dots] \text{ or } \psi(x) = [1, \cos(x), \cos^2(x), \dots].$$

2°) From points of evaluation $\{x_1, x_2, x_3, \dots, x_m\} \subset \mathbb{R}^N$ build $m \times p_i$ matrices

$$\Psi_{1 \leq i \leq N} = \begin{bmatrix} \psi([x_1]_i) \\ \psi([x_2]_i) \\ \vdots \\ \psi([x_m]_i) \end{bmatrix}$$

This formulation implicitly gives the *m-tensor* we work with (*row-wise tensor product*)

$$\Phi = \bigotimes_{i=1}^N \Psi_i \in \mathbb{R}^{m \times p_1 \times p_2 \times \dots \times p_N}$$

3°) Solve for coefficients (N-way contracted products between m-tensors)

$$C = \Phi^T (\Phi \Phi^T)^{-1} y.$$

We actually use

$$z = (\Phi \Phi^T)^{-1} y.$$

In practice we don't access C but build the approximation $\hat{f}(x)$ as

$$\hat{f}(x) = \langle \varphi(x) \Phi^T, z \rangle \text{ where } \varphi(x) = \bigotimes_{i=1}^N \psi([x]_i).$$

Inference of that approximation has computational complexity $\mathcal{O}(m(1 + Np))$.

The suggested regression approach only requires the inversion of a $m \times m$ matrix $\Phi \Phi^T$ which, by construction is symmetric positive definite. The method can be regularized by selecting a subset of rows of the m-tensor Φ to speed up the regression at a user-selected cost in terms of precision. We call this approach the Almost Linearly Independent regression.

Numerical example: 2 to 300 dimensions

We demonstrate how m-tensor regression scales with dimension on the example of the N -dimensional Rosenbrock function on $[-5, 10]^{\otimes N}$:

$$f(x) = \sum_{i=1}^{N-1} 100([x]_{i+1} - [x]_i^2)^2 - ([x]_i - 1)^2.$$

We build a regression from $[5, 10, 20, 30, 40] \times N$ points evaluated on a LHS and validate the degree 4 polynomial model on unseen points (5 times more than for the training). We observe low relative errors (**fig.1.1-1.2-1.5**) and a scaling that allows to reach dimensions unattainable by common approaches.

The regression computation stage suffers a *quadratic* dependence on N due to linear scaling in sampling size and complexity. The ALI decorrelates the size of the online operator from the sampling size. This results in a linear scaling in online call of the approximation (**fig.1.7**). The cost in terms of precision (**fig.1.5**) is controlled by the user. Displayed result require more tuning of hyperparameters at low dimensions. ALI regression shows massive time/computational gains (**fig.1.3 vs fig.1.6**).

M-Tensor Least Squares regression

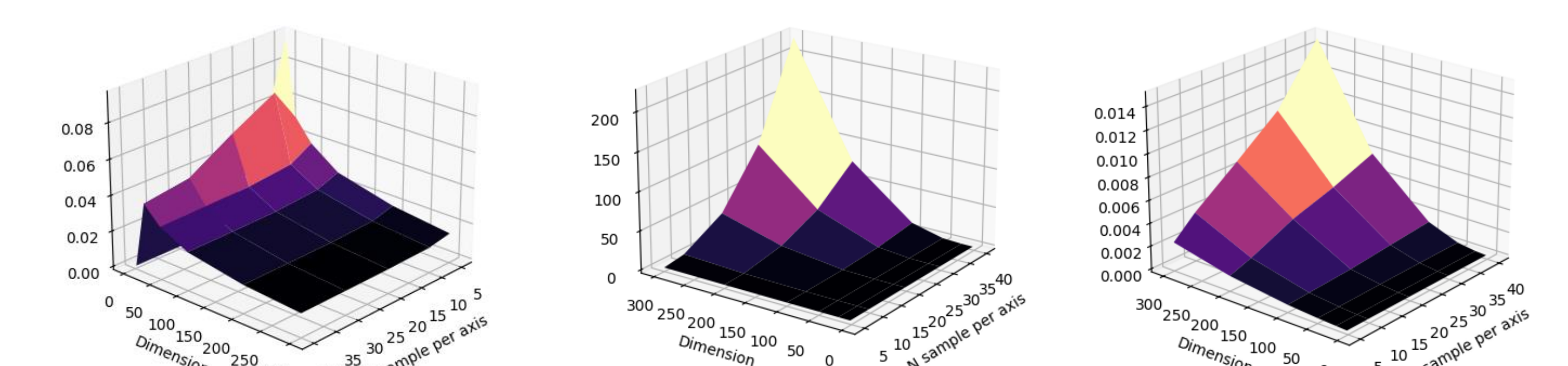


Fig.1.2: Relative error of m-tensor least squares model on unseen data for varying dimension and size of training samples

Fig.1.3: Computation time (s) of z for varying dimension and size of training samples

Fig.1.4: Unit inference time (s) of m-tensor least squares model for varying dimension and size of training samples

M-Tensor Almost Linearly Independent (ALI) Regression

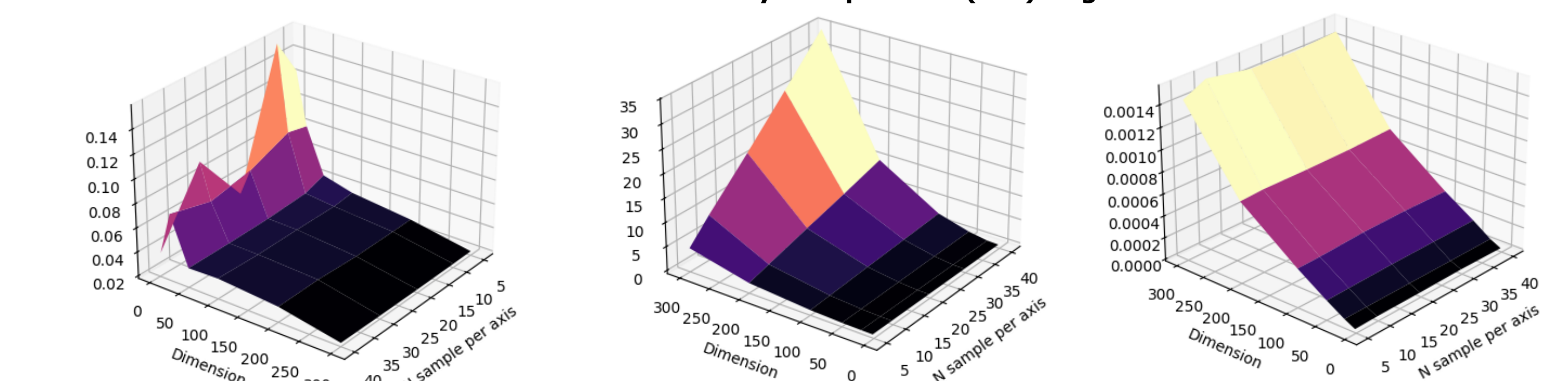


Fig.1.5: Relative error of m-tensor ALI model on unseen data for varying dimension and size of training samples

Fig.1.6: Computation time (s) of ALI z for varying dimension and size of training samples

Fig.1.7: Unit inference time (s) of m-tensor ALI model for varying dimension and size of training samples

We acknowledge fundings from the Arts et Métiers RTE Chair.

Reliable approach for dynamical systems:

We demonstrate the robustness of the approach on the example of Lorenz attractor (trajectory in **fig.2.1**). From a set of 500 timesteps to train our model, the system is learnt with different regularization approaches over 20 000 timesteps (**fig.2.2**). This result generalizes for 50 randomized initial conditions (**fig.2.3**).

We aim at generalizing those results for higher dimensional dynamical systems and provide a ℓ_1 or another proxy to ℓ_0 norm regularization based on a Khatri-Rao leverage score sampling of the mode-1 fibers of the m-tensor.

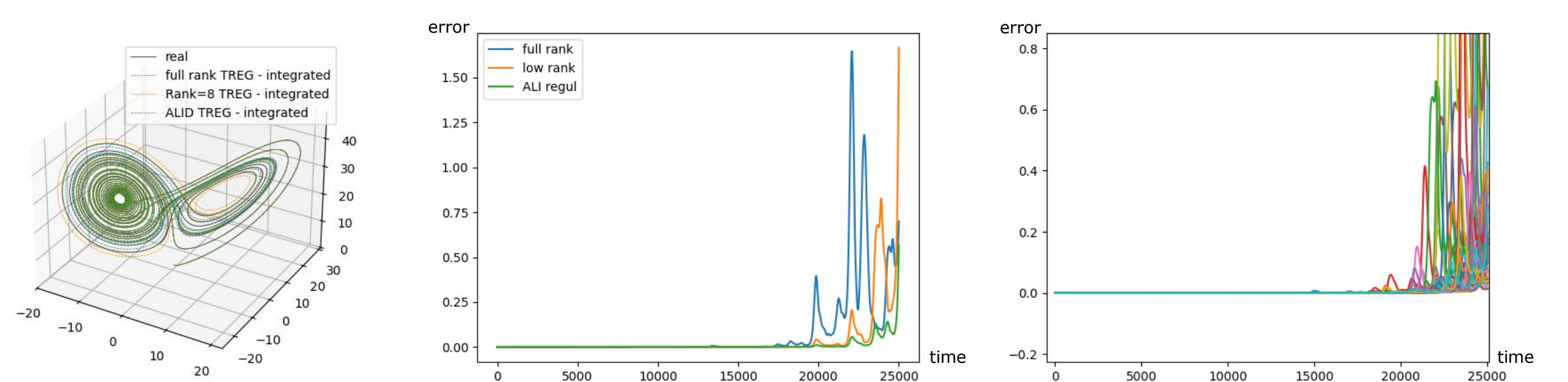


Fig.2.1: Lorenz attractor trajectory and learnt models

Fig.2.2: Relative error of model learnt with 500 timesteps

Fig.2.3: Relative error of model's trajectories for 50 randomized initial conditions

References:

T.G. Kolda et al. "Tensor Decompositions and Applications". In: SIAM Review 51.3 (2009), pp. 455–500. doi: 10.1137/07070111X.
 R. Ibáñez et al. "A Multidimensional Data-Driven Sparse Identification Technique: The Sparse Proper Generalized Decomposition". In: Complexity 2018 (2018), pp. 1–11. doi: 10.1155/2018/5608286.
 A. Ammar et al. "On the Convergence of a Greedy Rank-One Update Algorithm for a Class of Linear Systems". In: Arch Computat Methods Eng (2010), 17, pp. 473–486. doi: 10.1007/s11831-010-9048-z
 P.J. Baddoo et al. "Kernel learning for robust dynamic mode decomposition: linear and nonlinear disambiguation optimization". In: Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences 478.2260 (2022), p. 20210830. doi: 10.1098/rspa.2021.0830.
 Y. Engel et al. "The kernel recursive least-squares algorithm". In: IEEE Transactions on Signal Processing 52.8 (2004), pp. 2275–2285. doi: 10.1109/TSP.2004.830985.
 P. Gelb et al. "Feature space approximation for kernel-based supervised learning". In: Knowledge-Based Systems 221 (2021), p. 106935. issn: 0950-7051. doi: https://doi.org/10.1016/j.knsys.2021.106935.
 Stefan Klus et al. "Tensor-Based Algorithms for Image Classification". In: Algorithms 12.11 (2019). issn: 1999-4893. doi: 10.3390/a12110240.
 B. Schölkopf et al. Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond. The MIT Press, Dec. 2001. isbn: 9780262256933. doi: 10.7551/mitpress/4175.001.0001.
 V.I. Styusar. "A family of face products of matrices and its properties." In: Cybern Syst Anal (1999), pp. 379–384. doi: 10.1007/BF02733426.

We acknowledge fundings from the Arts et Métiers RTE Chair.



MORTECH 2025
 7th International Workshop on Model
 Order Reduction Techniques
 26-28 November 2025 | Zaragoza, Spain

