Enhancing Deep Learning for Slow-Decaying Problems: An Optimal Transport-based Approach with Sinkhorn Loss and Wasserstein Kernel

Moaad Khamlich [1], Federico Pichi [2], Gianluigi Rozza [1]

[1] Mathlab, Mathematics Area, International School of Advanced Studies (SISSA), Trieste, Italy

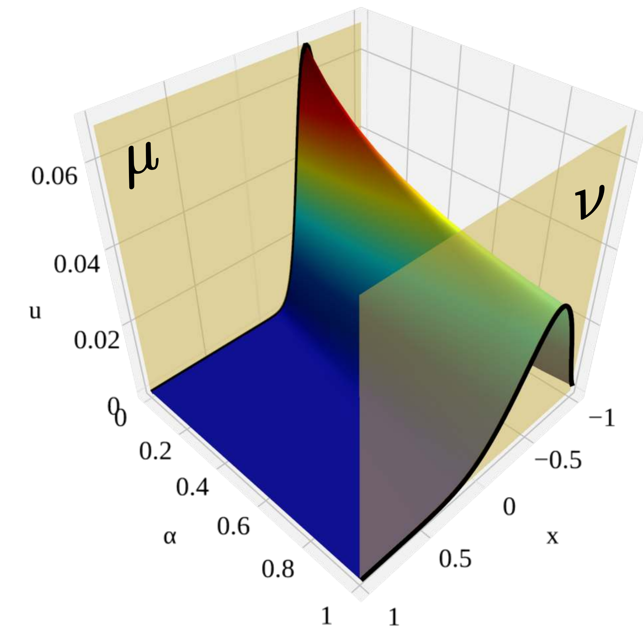
[2] Chair of Computational Mathematics and Simulation Science, École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland

Introduction

In the field of scientific computing, reduced order models (ROMs) play a crucial role in addressing high-dimensional systems. However, traditional ROM techniques often fail to fully capture the inherent geometric properties of the data, such as fundamental structures, relationships, and essential features necessary for accurate modeling.

We propose a novel ROM framework that combines optimal transport theory with neural networks. This framework effectively captures geometric data characteristics, enhancing accuracy and computational efficiency compared to traditional ROM methods. It utilizes Sinkhorn divergence as a loss function for training, resulting in better stability, resistance to overfitting, and faster convergence.

To demonstrate the effectiveness of our approach, we conduct experiments on a challenging set of test cases characterized by a gradual decay of the Kolmogorov n-width. The results demonstrate that our framework outperforms traditional ROM methods in terms of accuracy and computational efficiency.



Wasserstein barycenters between the solution to the Burgers equation at two different time instants.

OT-inspired Deep Learning Framework

We consider the parametrized partial differential equation (PDE) given by:

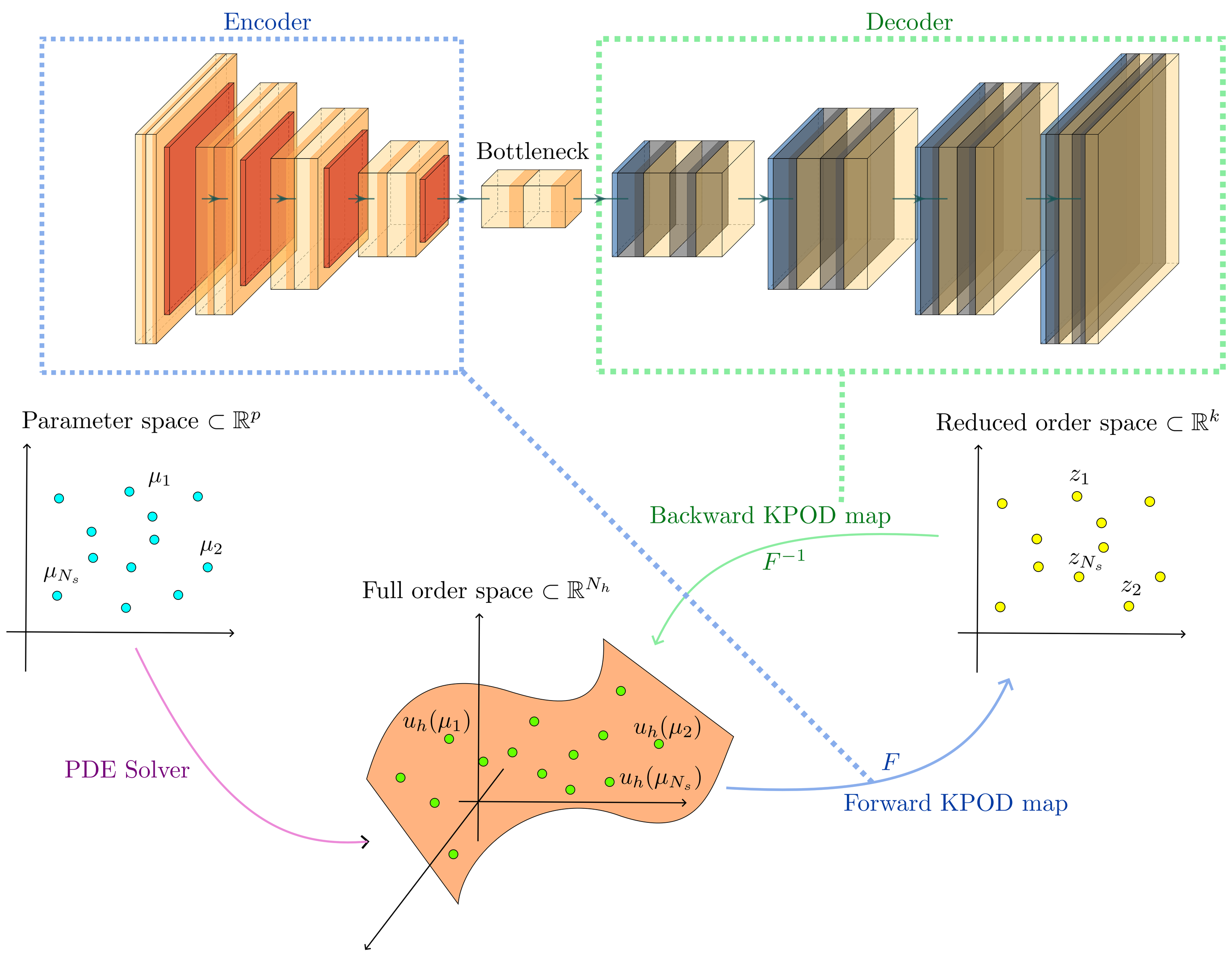
$$\mathcal{L}(u(\boldsymbol{\mu}), \boldsymbol{\mu}) = 0,$$

where $u(\boldsymbol{\mu}) \in \mathcal{V}$ and $\boldsymbol{\mu} \in \mathcal{P} \subset \mathbb{R}^p$ represents the parameter vector.

Approach:

- **Nonlinear Reduction:** Use kPOD with Wasserstein distance-based kernel for complex data features.
- **Deep Learning:** Apply autoencoders with Sinkhorn divergence for efficient dimensionality reduction.
- **Forward and Backward Maps:** Establish forward map F to project solutions into reduced space. Find backward map F^{-1} to reconstruct full-order solutions.
- **Complex Systems:** Explore Convolutional Autoencoders (CAE) for managing high-dimensional states.

Outcome: Integrated approach ensures accurate, efficient, and scalable parametrized PDE modeling with forward and backward mappings.



Schematic illustration of the different steps of the proposed methodology.

Sinkhorn Divergence [1]

The Sinkhorn algorithm efficiently computes the regularized Wasserstein distance (W_{reg}) between probability measures by introducing entropy (ϵ) into the optimization problem. The regularized Wasserstein distance is defined as:

$$W_{\text{reg}}(\mu, \nu) = \min_{\boldsymbol{\pi} \in \Pi_{\text{dis}}(\mu,\nu)} \langle \mathbf{C}, \boldsymbol{\pi} \rangle - \epsilon H(\boldsymbol{\pi});$$

where $H(\boldsymbol{\pi})$ is the Shannon entropy, $\mathbf{C}$ is the cost matrix and $\Pi_{\text{dis}}(\mu, \nu)$ is the set of admissible transport plans between the two probability measures μ and ν. The Sinkhorn distance ($W_\epsilon(\mu, \nu)$) can be debiased to obtain Sinkhorn divergence ($S_\epsilon(\mu, \nu)$):

$$S_\epsilon(\mu, \nu) = W_\epsilon(\nu, \mu) - \frac{1}{2} W_\epsilon(\nu, \nu) - \frac{1}{2} W_\epsilon(\nu, \nu).$$

Sinkhorn divergence ensures $S_\epsilon(\mu, \mu) = 0$, serving as a valid dissimilarity measure. Computationally, Sinkhorn divergence is highly parallelizable and can exploit GPUs, making it efficient for large-scale problems ($\mathcal{O}(n \log n + m \log m)$).

Kernel Proper Orthogonal Decomposition (kPOD) [2]

kPOD is a kernel-based extension of classical POD. It operates in a higher-dimensional feature space, enhancing nonlinear separability. To apply kPOD, using the kernel function $\kappa(\cdot, \cdot)$, compute the transformed Gram matrix $\tilde{\mathbf{G}}$:

$$[\tilde{\mathbf{G}}]_{ij} = \kappa(\boldsymbol{u}_h^i, \boldsymbol{u}_h^j).$$

Define the *forward mapping* F:

$$F : \mathbb{R}^{N_h} \to \mathbb{R}^k, \quad \boldsymbol{u}_h \mapsto \boldsymbol{z} = \tilde{\mathbf{V}}^{\star\mathbf{T}} g(\boldsymbol{u}_h), \quad \tilde{\mathbf{G}} = \tilde{\mathbf{V}} \Lambda \tilde{\mathbf{V}}^{\mathbf{T}},$$

Defining the *backward mapping* requires a suitable target space $\mathcal{V}$ and an approximation criterion:

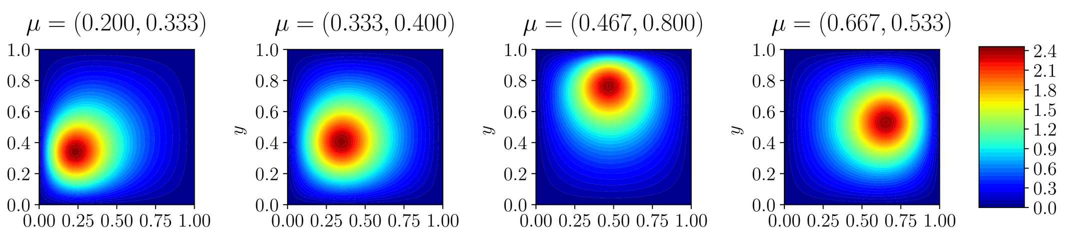
$$\tilde{\boldsymbol{u}}_h = F^{-1}(\boldsymbol{z}) \stackrel{\text{def}}{=} \arg\min_{\boldsymbol{u}_h \in \mathcal{V}} \|F(\boldsymbol{u}_h) - \boldsymbol{z}\|.$$

For the inverse problem, we propose using autoencoders based on Neural Networks, which are effective in solving kPOD's inverse problem.

Numerical Results [3]

Poisson test case

Solution w.r.t. the source location.

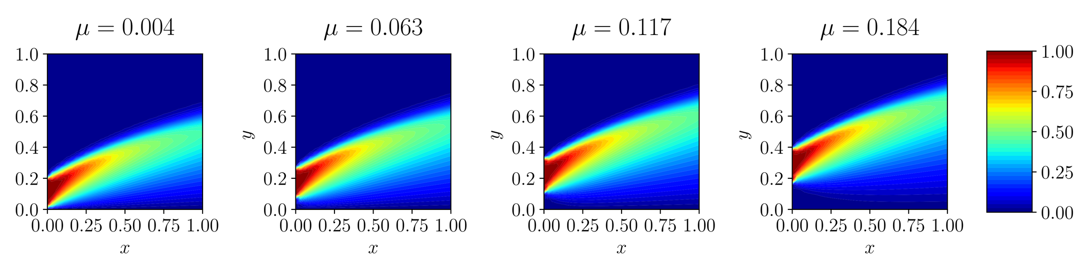


Comparison of mean relative errors for different NNs.

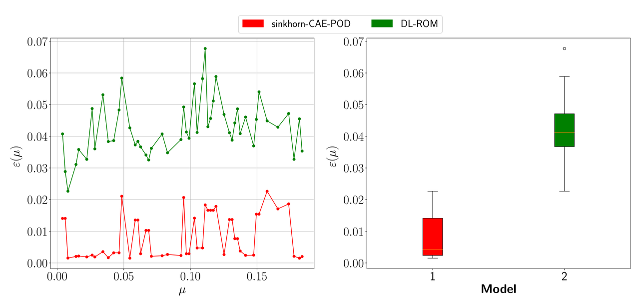
	Architecture / Loss type	kPOD FF	kPOD CAE	POD FF	POD CAE
Autoencoder	Sinkhorn	2.02 %	0.50 %	1.95 %	1.00 %
	MSE	2.22 %	0.70 %	2.08 %	1.20 %
Decoder	Sinkhorn	2.60 %	1.50 %	2.84 %	1.75 %
	MSE	2.22 %	2.42 %	2.80 %	2.61 %

Burgers equation test case

Solution by varying the position of the initial condition .

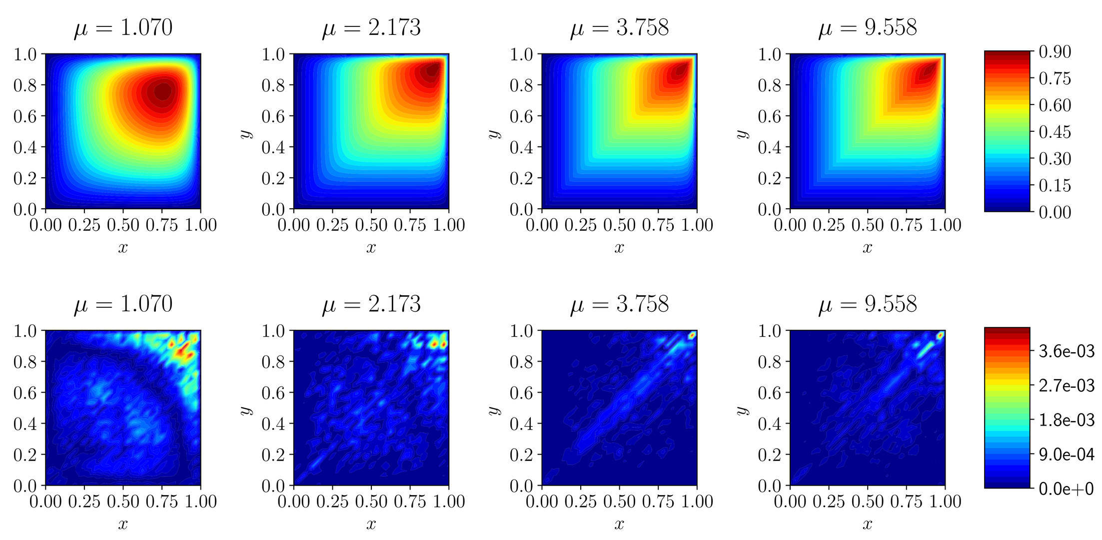


Relative errors for DL-ROM and the kPOD-autoencoder.



Linear advection-dominated test case

Solution (top row) and relative error fields (bottom row) by varying the diffusivity $\alpha(\mu) = 10^{-\mu}$.



References

[1] M. Cuturi. Sinkhorn Distances: Lightspeed Computation of Optimal Transport. In C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 26. Curran Associates, Inc., 2013. URL `https://proceedings.neurips.cc/paper_files/paper/2013/file/af21d0c97db2e27e13572cbf59eb343d-Paper.pdf`.

[2] B. Schölkopf, A. Smola, and K.-R. Müller. Kernel principal component analysis. In *Artificial Neural Networks—ICANN'97: 7th International Conference Lausanne, Switzerland, October 8–10, 1997 Proceeedings*, pages 583–588. Springer, 2005. doi:10.7551/mitpress/1130.003.0026.

[3] M. Khamlich, F. Pichi, and G. Rozza. Optimal transport-inspired deep learning framework for slow-decaying problems: Exploiting sinkhorn loss and wasserstein kernel, 2023, arxiv:2308.13840.