A Deep Learning approach to Reduced Order Modeling of Parameter dependent Partial Differential Equations

Paolo Zunino

MOX, Department of Mathematics,

Politecnico di Milano

In collaboration with Nicola-Rares Franco, Stefania Fresca, Andrea Manzoni

POLITECNICO MILANO 1863

March 9 2023

Friedrich-Alexander-Universität Erlangen-Nürnberg

This work is supported by the project

ERA PERMED COFUND JTC 2018 ID 244 RADPRECISE

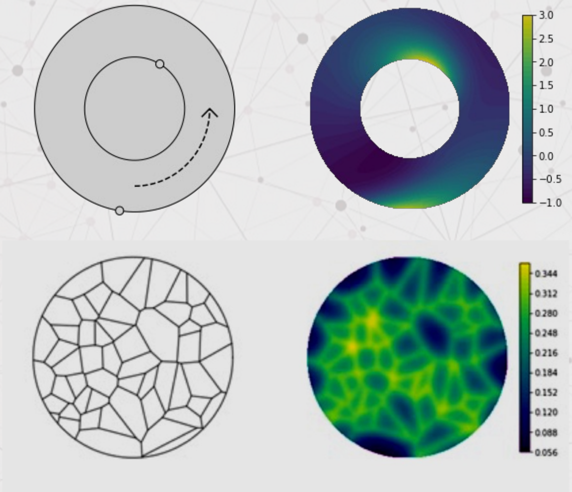
General setting ans objectives

General Objective:

develop a model order reduction strategy for parametrized PDES

- **non-intrusive –** full decoupling of offline and online steps
- **data-driven**
- based on a **supervised learning**

Parameters → **Solutions**



Physical params.

μ

Geometrical params.

We are given a parameter dependent **steady** PDE model, e.g.

Abstract model

$$\begin{cases} A_{\boldsymbol{\mu}} u_{\boldsymbol{\mu}} = f_{\boldsymbol{\mu}} & \text{on } \Omega \\ B_{\boldsymbol{\mu}} u_{\boldsymbol{\mu}} = g_{\boldsymbol{\mu}} & \text{in } \partial\Omega \end{cases}$$

Full Order Model

$$\begin{cases} A^h_{\boldsymbol{\mu}} u^h_{\boldsymbol{\mu}} = f^h_{\boldsymbol{\mu}} & \text{in } \Omega \\ B^h_{\boldsymbol{\mu}} u^h_{\boldsymbol{\mu}} = g^h_{\boldsymbol{\mu}} & \text{on } \partial\Omega \end{cases}$$

Specific objective: build a reduced order model (ROM) based on a **Deep Neural Network (DNN) Φ** to approximate the **parameter-to-solution map**

$$\mathbb{R}^p \supset \boldsymbol{\Theta} \ni \boldsymbol{\mu} \to u^h_{\boldsymbol{\mu}} \in V_h \text{ or } \mathbf{u}^h_{\boldsymbol{\mu}} \in \mathbb{R}^{N_h}$$

or equivalently the **solution manifold** $\mathcal{S} = \{u^h_{\boldsymbol{\mu}}\}_{\boldsymbol{\mu}\in\Theta}$

The **approximation** (minimization) is based on the functional:

$$\int_\Theta \mathcal{L}(\Theta, \Phi) d\mathbb{P}_\Theta \quad \Rightarrow \quad \int_\Theta \frac{\|\mathbf{u}^h_{\boldsymbol{\mu}} - \Phi(\boldsymbol{\mu})\|^2}{\|\mathbf{u}^h_{\boldsymbol{\mu}}\|^2} d\mu$$

The DL-ROM approach: main ideas

- *Franco, N.R, Manzoni, A. and Zunino, P., A Deep Learning approach to Reduced Order Modelling of Parameter Dependent Partial Differential Equations, arXiv:2103.06183 [math.NA], to appear on Math. Comp. 2022.*
- *Fresca, S., Manzoni, A. and Dede, L., A comprehensive deep learning-based approach to reduced order modeling of nonlinear time-dependent parametrized PDEs, Journal of Scientific Computing, 87(2), 1-36 (2021).*
- *Fresca, S., and Manzoni, A., POD-DL-ROM: enhancing deep learning-based reduced or- der models for nonlinear parametrized PDEs by proper orthogonal decomposition, Computer Methods in Applied Mechanics and Engineering, 388: 114181 (2022).*

Ultimate goal. Build a DNN model Φ such that $\Phi(\boldsymbol{\mu}) \approx \mathbf{u}_{\boldsymbol{\mu}}^h$.

Obs. Linear methods exploit a representation of the form

$$\mathbf{V} \in \mathbb{R}^{N_h \times n} \implies \mathbf{u}_{\boldsymbol{\mu}}^h \approx \mathbf{V}\mathbf{u}_{\boldsymbol{\mu}}^n$$

Idea. Compress the solution manifold using an AutoEncoder (AE)

$$\mathcal{S} := \{\mathbf{u}_{\boldsymbol{\mu}}^h\}_{\boldsymbol{\mu}\in\Theta}$$

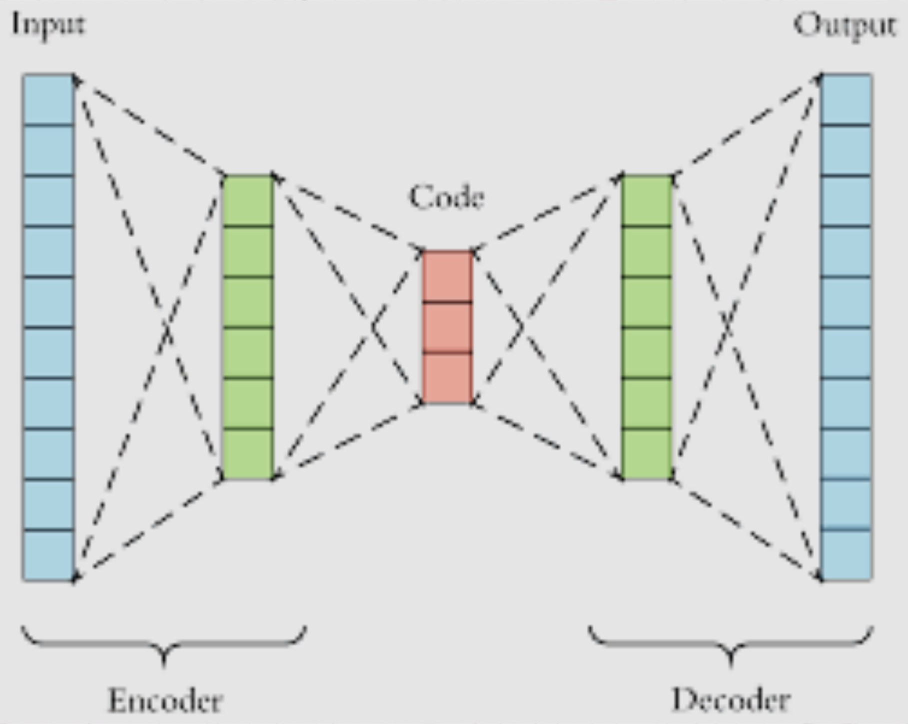
$$\Psi' : \mathcal{S} \to \mathbb{R}^n$$

$$\Psi : \mathbb{R}^n \to \mathbb{R}^{N_h}$$

$$\mathbf{u}_{\boldsymbol{\mu}}^h \approx \Psi\left(\Psi'\left(\mathbf{u}_{\boldsymbol{\mu}}^h\right)\right)$$

This is a low dimensional representation of the solution manifold…

how about the parameter to solution map?



The DL-ROM approch: training and summary

Discrete model: N_h d.o.f

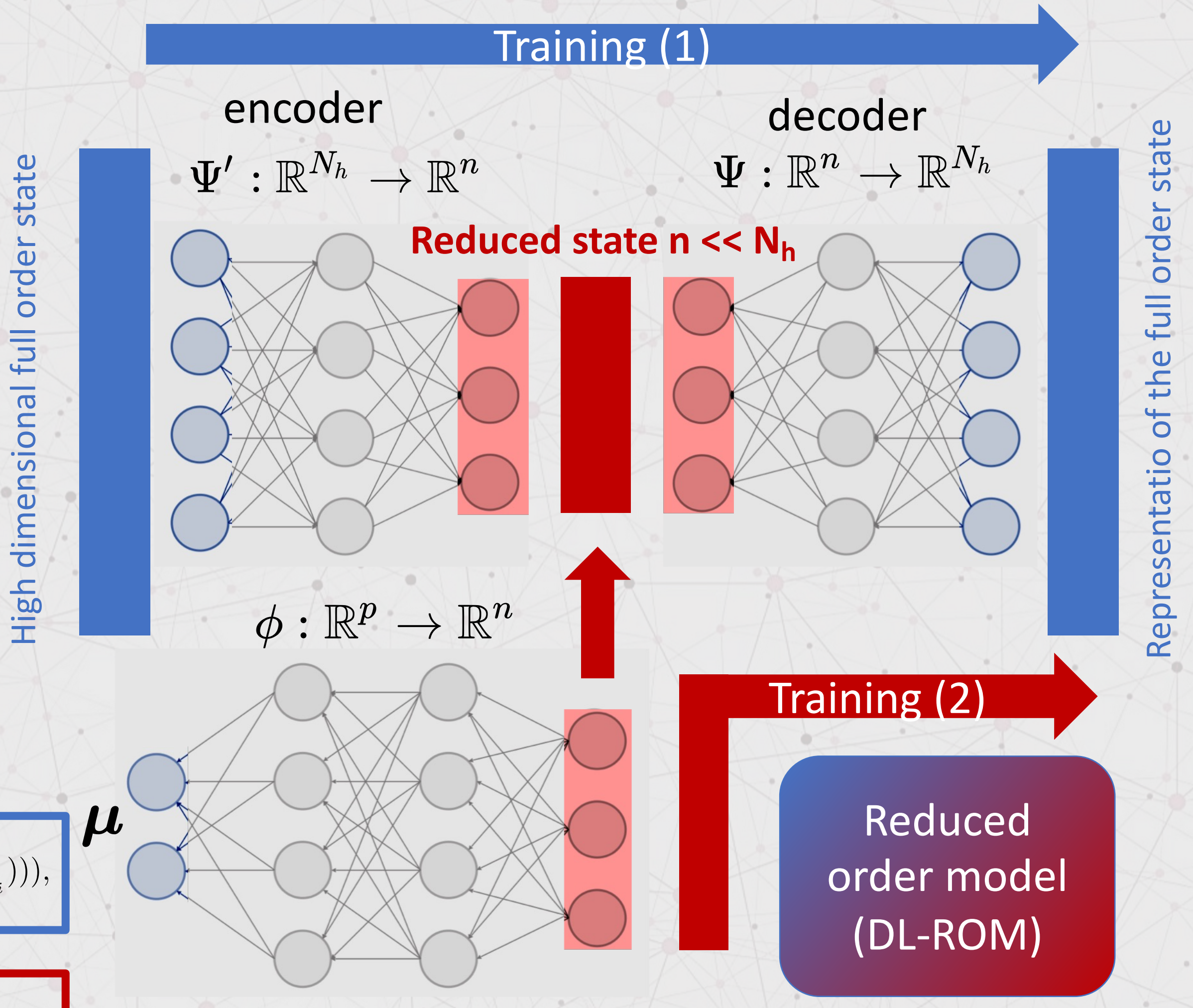
Full order model (FOM)

$$\boldsymbol{\mathcal{A}}^{\mu}\boldsymbol{u}_h^{\mu} = \boldsymbol{\mathcal{F}}_h^{\mu}$$

The loss functions:

$$\mathrm{Loss}(\Psi', \Psi) = \frac{1}{N_{\mathrm{train}}}\sum_{i=1}^{N_{\mathrm{train}}} \mathcal{L}(\mathbf{u}^h_{\boldsymbol{\mu}_i},\ \Psi(\Psi'(\mathbf{u}^h_{\boldsymbol{\mu}_i}))),$$

$$\mathrm{Loss}(\phi) = \frac{1}{N_{\mathrm{train}}}\sum_{i=1}^{N_{\mathrm{train}}} \mathcal{L}(\mathbf{u}^n_{\boldsymbol{\mu}_i},\ \phi(\boldsymbol{\mu}_i))$$



Parameters to reduced state map

$$\Phi := \Psi \circ \phi : \mathbb{R}^p \to \mathbb{R}^{N_h}$$

DL-ROM

- Non intrusive ROM
- Nonlinear trial manifold learning
- Extended to time dependent problems

S. Fresca, L. Dede', A. Manzoni. A Journal of Scientific Computing, 87(2):1-36, 2021.

S. Fresca, A. Manzoni. Computer Methods in Applied Mechanics and Engineering, 388 114181, 2022.

Franco, N.R, Manzoni, A. and Zunino, P., Math. Comp. 2023, (340), pp. 483 - 524

Franco, N.R, Manzoni, A. and Zunino, P., arXiv:2203.11648, 2022.

The DL-ROM approach: some questions

Some important questions:

Q1) How to choose the latent dimension?

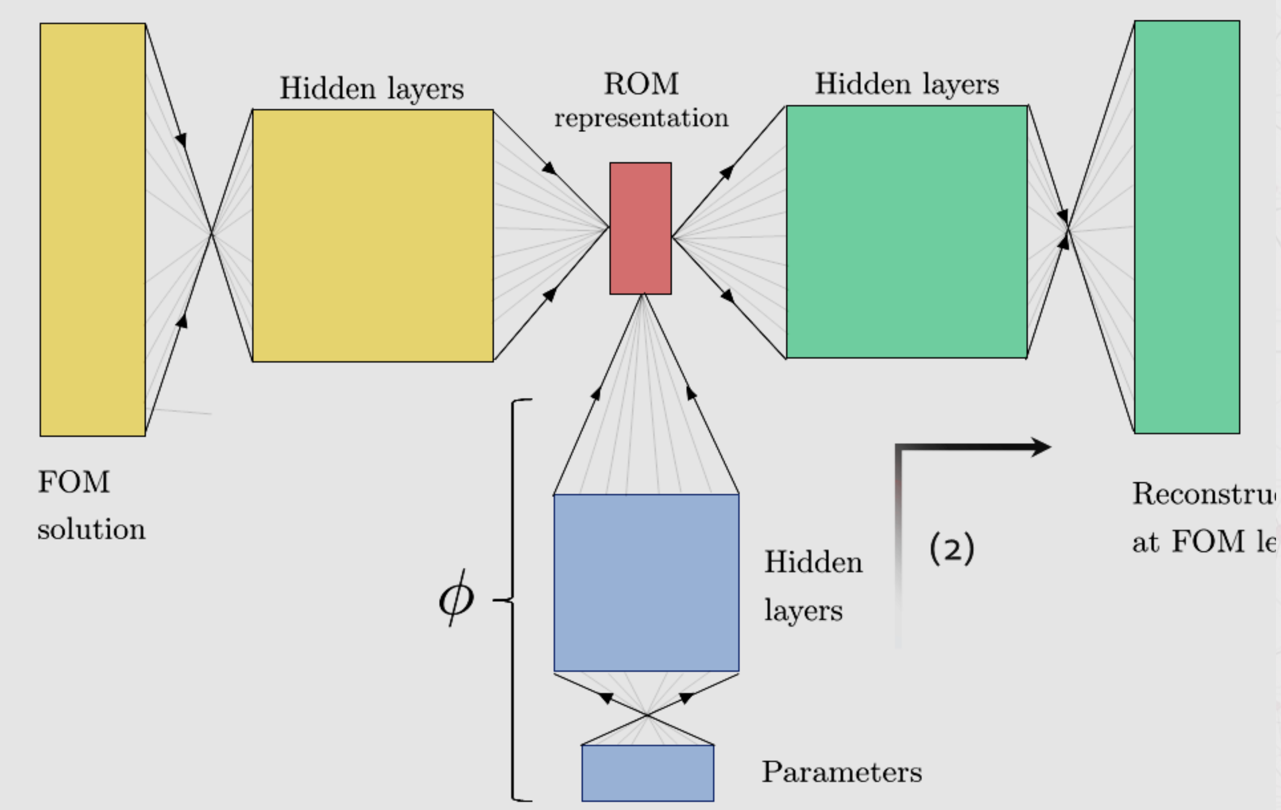
Q2) Error analysis?

Q3) What kind of DNN architectures?

Other topics: general models, general domains, time dependent problems etc.

Q1: Properties of the latent space

What is the relation between the minimal latent dimension and the topological dimension of the solution manifold?



General abstract result:

Franco N.R., Manzoni A., Zunino P., A Deep Learning approach to Reduced Order Modelling of Parameter Dependent Partial Differential Equations, (2023) Mathematics of Computation, 92 (340), pp. 483 - 524

Theorem 1 (informal)

Assume the parameter space is compact;
If S has topological dimension p;
If the parameter-to-solution map is:

a) Lipschitz continuous, **then** $n_{\min}(\mathcal{S}) \leq 2p+1$ (from Menger-Nobeling Th.)

b) Continuous and injective near at least a point, **then** $n_{\min}(\mathcal{S}) \geq p$

c) Continuous and injective, **then** $n_{\min}(\mathcal{S}) = p$

Q1: Properties of the latent space

Application to PDEs

find $u \in W^{1,q}(\Omega) : u_{|\partial\Omega} = \boldsymbol{g}_{\boldsymbol{\mu}}$ and

$$\int_{\Omega} \boldsymbol{\sigma}_{\boldsymbol{\mu}} \nabla u \cdot \nabla w + \int_{\Omega} (\boldsymbol{b}_{\boldsymbol{\mu}} \cdot \nabla u)\, w = \int_{\Omega} f_{\boldsymbol{\mu}} w \quad \forall w \in W_0^{1,q'}(\Omega)$$

Application to PDEs with stochastic coefficients:

$$\begin{cases} \nabla \cdot \left(e^{\sigma(\boldsymbol{\mu})} \nabla u_{\boldsymbol{\mu}}\right) & = f_{\boldsymbol{\mu}} \quad \text{in } \Omega \\ u_{\boldsymbol{\mu}} & = 0 \quad \text{on } \Gamma_D \end{cases}$$

σ is a random field given by a **truncated** KL expansion:

$$\sigma(\boldsymbol{x}; \boldsymbol{\mu}) = m + \sum_{i=1}^{p} \sqrt{\lambda_i} \mu_i \varphi_i(\boldsymbol{x})$$

Theorem 2 (informal)

a) **if** the dependence of $\sigma_\mu, b_\mu, f_\mu, g_\mu$ on $\boldsymbol{\mu} \in \mathbb{R}^p$ is Lipschitz continuous, **then** $n_{min}(S) \leq 2p + 1$

b) **if** $\sigma_\mu, b_\mu, f_\mu, g_\mu$ depend continuously on $\boldsymbol{\mu}$ and the solution map $\boldsymbol{\mu} \to u_{\boldsymbol{\mu}}$ is one-to-one, **then** $n_{min}(S) = p$

Theorem 3 (informal)

For all $n \geq 2p + 1$, if $sup_i \|\varphi_i\|_{L^\infty} < \infty$, then

for some C, independent of n, p

$$\inf_{\substack{\Psi' \in \mathcal{C}(L^2(\Omega), \mathbb{R}^n) \\ \Psi \in \mathcal{C}(\mathbb{R}^n, L^2(\Omega))}} \mathbb{E} \| u_\mu - \Psi(\Psi'(u_{\mu^{(p)}})) \|_{L^2(\Omega)} \leq C \sqrt{\sum_{i>p} \lambda_i},$$

Very efficient (optimal) dimensional reduction

Franco, N.R, Manzoni, A. and Zunino, P., Math. Comp. 2023

Q2: general concepts about the error analysis of DNN

The error of the Deep Learning approximation can be represented as follows:

$$\mathcal{E}(f_{\mathcal{D}}) - \mathcal{E}^* =$$

Generalization error $\quad = \mathcal{E}(f_D) - \widehat{\mathcal{E}}_D(f_D)$

Optimization error $\quad + \widehat{\mathcal{E}}_D(f_D) - \widehat{\mathcal{E}}_D(f_F^*)$

Generalization error $\quad + \widehat{\mathcal{E}}_D(f_F^*) - \mathcal{E}(f_F^*)$

Approximation error $\quad + \mathcal{E}(f_F^*) - \mathcal{E}^*$

The functional to be minimized is (also called MRE):

$$\mathcal{E}(f) = \int_Z L(z, f) dP_Z$$

The global minimum of the error is:

$$\mathcal{E}^* = inf_{f \in M(X,Y)} \mathcal{E}(f)$$

The optimal function in the hypothesis set F is:

$$f_F^* = argmin_{f \in F} \mathcal{E}(f)$$

The empirical MRE that is minimized on the training set D is:

$$\hat{\mathcal{E}}_D(f) = \frac{1}{m} \sum_{i=1}^{m} L(f, z^i)$$

The output of the learning algorithm is the function f_D where D are the training data

J. Berner, P. Grohs, G. Kutyniok, P. Petersen, The Modern Mathematics of Deep Learning, arXiv:2105.04026v1, 2021

Q2: Error analysis of DL-ROM approximation

What is the size of the DNN that satisfies a given tolerance on E_A?

$$\Phi = \Phi(\ldots, W^{(l)}, b^{(l)}, \ldots)$$

$$\text{size}(\Phi) = \sum_{l=1}^{L} (\|W^{(l)}\|_{NNZ} + \|b^{(l)}\|_{NNZ})$$

Worst case analysis $\mathrm{E(f)} = \int_Z L(z,f)dP_Z = sup_{z\in Z} L(z,f)$

The approximation error is: $E_A = sup_{\mu\in\Theta} \left\| u_\mu^h - \Psi^*(\phi^*(\mu)) \right\|$

Approximation error (E_A) $\leq$ reconstruction error (E_R) + parametric error (R_P)

Theorem 4 (informal)

Assume that the map $\boldsymbol{\mu} \rightarrow \boldsymbol{u}_{\boldsymbol{\mu}}^{\boldsymbol{h}}$ is Lipshitz continuous (with constant L)

Assume that the infimum of E_A is attained for Ψ^* and Ψ'^* that are s-times different

by the constants C_1, C_2 respectively.

Assume that, for a tolerance ε there exists $m \in \mathbb{N}$ s.t. $d_m(\mathcal{S}) < \varepsilon$.

Then, for a constant $c(\Theta, L, C_1, C_2, p, n, s)$, there exists

- a ReLU decoder Ψ having at most $\boldsymbol{cm^{1*n/(s-1)}\varepsilon^{-n/(s-1)}\log(m/\varepsilon) + mN_h}$ **active weights** and $\mathbf{c\log(m/\varepsilon)}$ **layers**,
- a ReLU reduced map ϕ having at most $\boldsymbol{c\varepsilon^{-p}\log(1/\varepsilon)}$ **active weights** and $\mathbf{c\log(1/\varepsilon)}$ **layers**,

such that the approximation error satisfies $E_A \leq 2\varepsilon$

Franco, N.R, Manzoni, A. and Zunino, P., Math. Comp. 2023

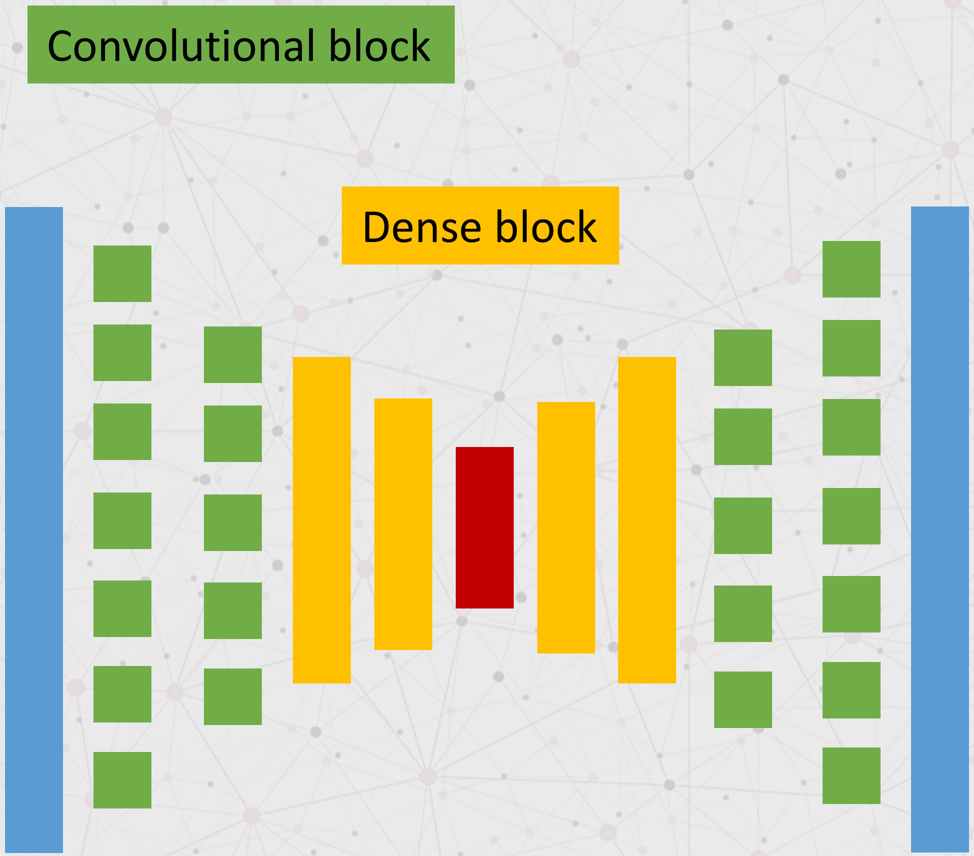
The expressivity of DNN:
NN depth and size increase poly-logarithmically with respect to the target approximation accuracy

$$\text{size}(\Phi) \approx |\log(\varepsilon)|$$

Marcati, C., Opschoor, J.A.A., Petersen, P.C. et al. Exponential ReLU Neural Network Approximation Rates for Point and Edge Singularities. Found Comput Math (2022).

Q3: DNN Architecure Design – Main guidelines

How should we define the autoencoder architecture?
What is the balance between convolutional and dense blocks?



$$\Psi' : \mathbb{R}^{N_h} \to \mathbb{R}^n \qquad \Psi : \mathbb{R}^n \to \mathbb{R}^{N_h}$$

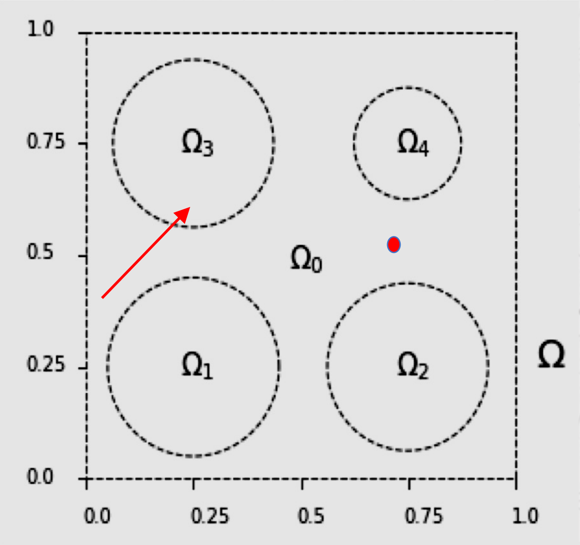
Franco, N.R., Fresca, S., Manzoni, A., Zunino, P. Approximation bounds for convolutional neural networks in operator learning (2023) Neural Networks, 161, pp. 129-141.

FOM features:	**Specific Architecture:**
Mesh resolution	Depth of convolutional block
Desired Accuracy	Depth of dense block
Accuracy and regularity of the manifold	Width of dense block
Regularity of the solution space	Number of channels in the convolutional block

- $C\varepsilon^{-2/(2s-1)}$ **input** and **output channels**

Numerical results

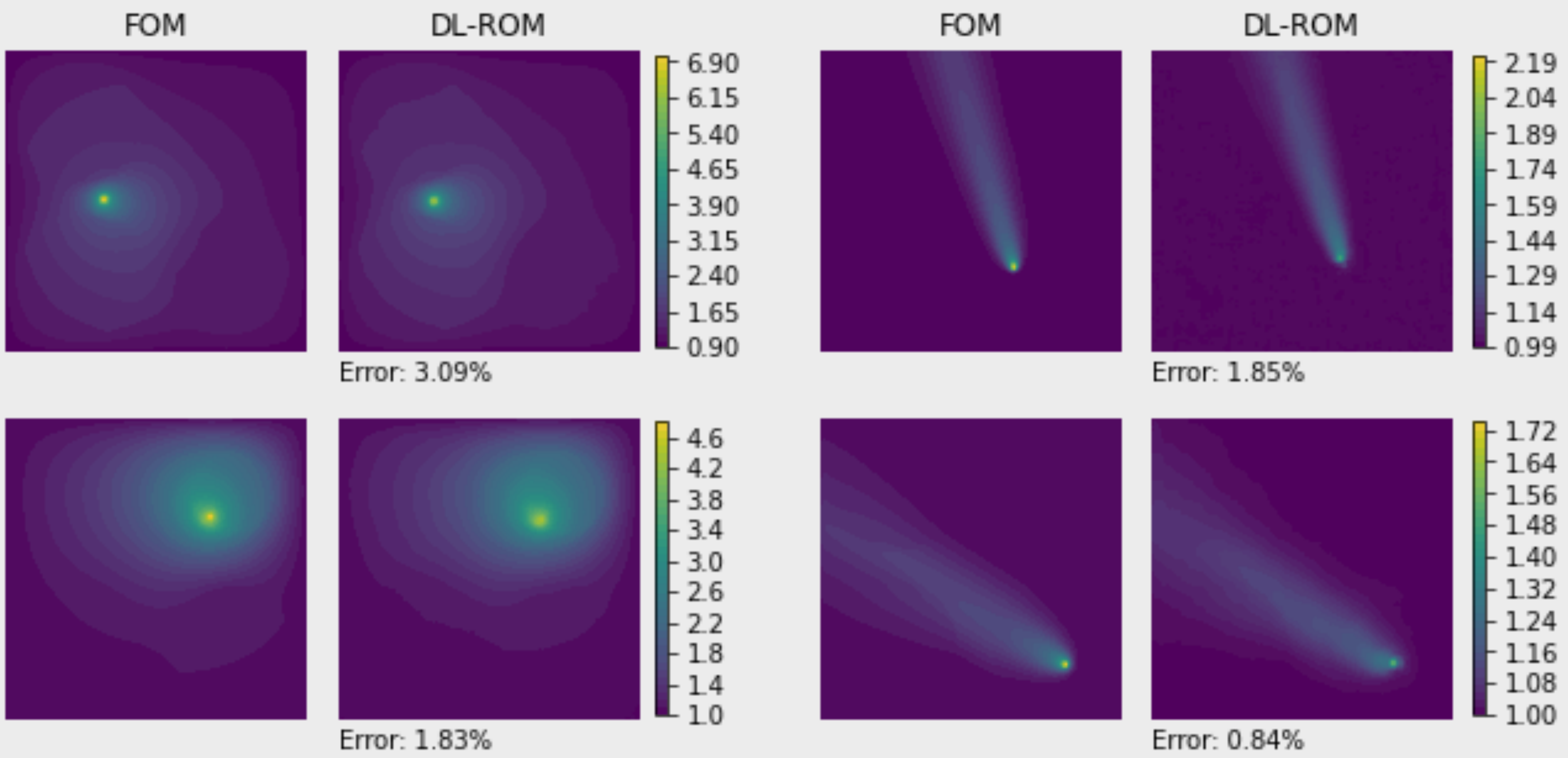
Advection-diffusion equation.



The parameters describe

- The conductivity coefficient
  (piecewise constant, cookie-like)
- Singular force location
- Transport direction

Here, $p = 7$.

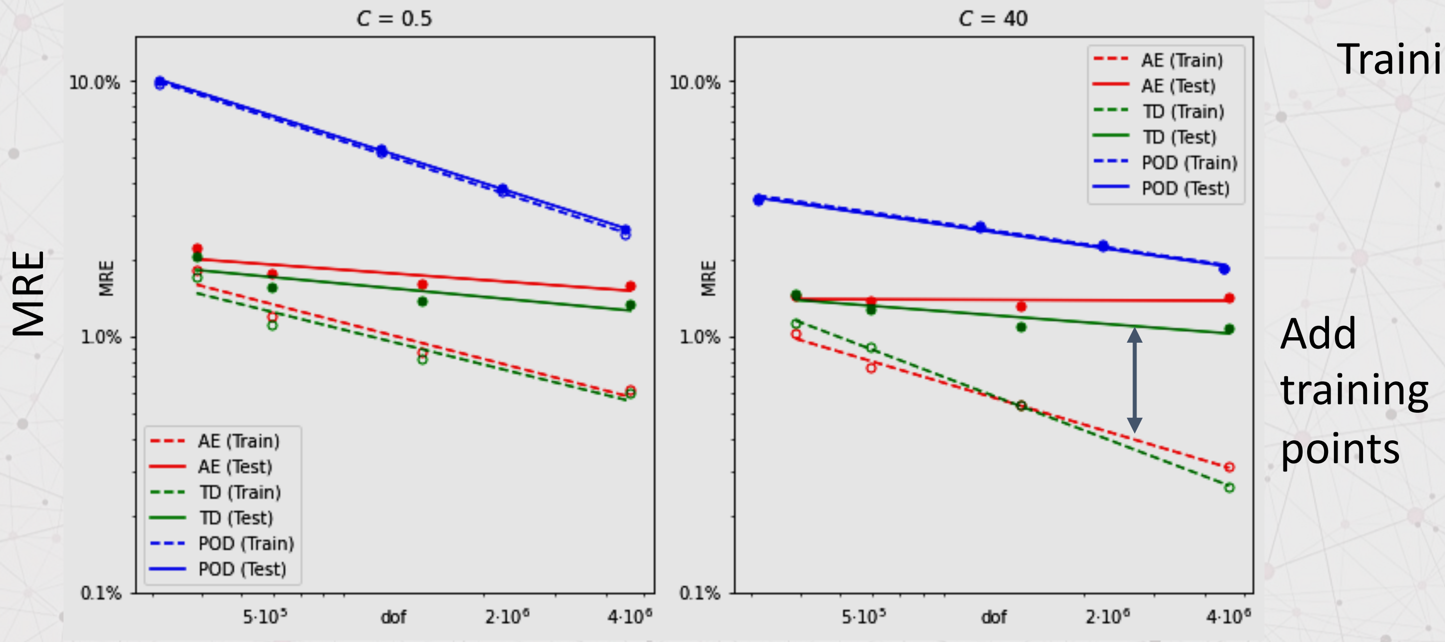


Good results for both the dissipative and transport-dominated case.

Numerical results

Advection-diffusion equation - Overcoming the Kolmogorov *n*-width barrier

Coherently with Theorem 3, we can improve the AE accuracy by enriching the decoder, by increasing the active weights **without changing the latent dimension $n_{min} = p = 7$.**



Active weights

Active weights

The compression rate doesn't change $n_{min} = p = 7$

Training on 10K data points 3-4- points per each dimension.

Online cost: negligible

Offline cost (1): moderate for training

Offline cost (2): high for snapshot collection.

Deep Learning enhanced ROM for complex problems

Deep learning ehnanced reduced order models

- **Satisfy a desired accuracy;**
- **Control the computational effort;**
- **Streamlining the workflow for complex problems;**
- **Increase the interoperability of computational models.**

Many-query scenario
- Sensitivity analysis
- Uncertainty quantification

Model order reduction
- Real time computations

Heterogeneous model coupling
- Multiphysics
- Multiscale-upscaling

Physics based models
- **Fluid mechanics**
- **Solid mechanics**
- …

Model learning
- Discovery of constitutive laws

Dr. Nicola-Rares Franco

MOX-DMAT-PoliMi

Prof. Andrea Manzoni

MOX-DMAT-PoliMi

Thank you!

Dr. Stefania Fresca

MOX-DMAT-PoliMi