POLITECNICO MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

Executive Summary of the Thesis

Guiding adaptivity in projection-based surrogates via model diagnostics

Laurea Magistrale in Mathematical Engineering - Ingegneria Matematica

Author: Samuele Caccavelli

Advisor: Prof. Andrea Manzoni

Co-advisor: Dr. Nicola Rares Franco

Academic year: 2025-2026

1. Introduction

Many engineering applications, from structural design to uncertainty quantification, require repeated numerical simulations of complex physical systems modeled with parametric Partial Differential Equations (PDEs). Within this context, relying exclusively on classical numerical solvers —hereafter referred to as Full Order Models (FOMs)— can become computationally prohibitive. Indeed, since each simulation often necessitates substantial computational resources, thoroughly exploring the solution space across different configurations of the parameter vector $\boldsymbol{\theta}$ can result in a highly challenging task. Reduced Order Modeling [4, 5] addresses this challenge by constructing efficient surrogate models, called Reduced Order Models (ROMs), that approximate the FOM solutions at a fraction of the computational cost.

One fundamental approach for ROMs is dimensionality reduction, where FOM solutions are used to create a low-rank latent space capable of representing the data with a minimal loss of accuracy. Such an approach ranges from linear methods, such as the Principal Orthogonal Decomposition (POD) [5], to fully nonlinear ones, such as deep autoencoders.

For linear models, the efficacy in the production of such latent spaces is strictly linked to the Kolmogorov n-width [3]. This quantity measures how well a set can be approximated by a linear subspace of dimension n. If the Kolmogorov n-width is fast-decaying in n, a linear approach is well-suited, otherwise a nonlinear approach should be considered. There also exist hybrid scenarios, where the set of parametric solutions has a slow-decaying Kolmogorov n-width, but if we fix some of the parameters, it becomes fast-decaying. For example, consider a parametric solution of the form

$$u(x,\boldsymbol{\theta}) = \theta_3 e^{-100(x-\theta_1)^2} + \theta_4 e^{-100(x-\theta_2)^2},$$

with the components of the parameter vector subdivided as $\boldsymbol{\mu} = [\theta_1, \theta_2]$ and $\boldsymbol{\nu} = [\theta_3, \theta_4]$. While the set of such solutions is slow decaying, for each fixed $\boldsymbol{\mu}$ the Kolmogorov n-width goes to zero after $n = 2$.

In the ROM literature, the peculiar structure of these problems has led to the development of *adaptive-basis methods*. Rather than relying on a single global low-rank basis, these methods allow the latent space to evolve as a function of a subset of the system's parameters, here denoted by $\boldsymbol{\mu}$. Among many such techniques, we will focus on deep neural network-based ones, and in particular, the Deep Orthogonal Decomposition (DOD) presented in [3], that aims at learning a map between $\boldsymbol{\mu}$ and a corresponding adaptive basis $\mathbf{V}_{\boldsymbol{\mu}}$ using the powerful learning capabilities of deep neural networks. Despite their power, the practical deployment of adaptive-basis approaches is hindered by the absence of rigorous *a priori* guidelines for two key architectural choices: the optimal dimension of the adaptive basis, and the partitioning of the full parameter vector into an adaptive component $\boldsymbol{\mu}$ and a static component $\boldsymbol{\nu}$.

This thesis addresses precisely these gaps by developing a data-driven diagnostic pipeline that provides quantitative, pre-training guidance for both decisions. Moreover, leveraging these newly developed instruments, this work presents the first application of the DOD to time-dependent problems.

2. Methodology

2.1. Weighted-POD

The core engine of the proposed pipeline is the weighted Principal Orthogonal Decomposition (w-POD). The latter is not used to build an online surrogate, but rather serves as an auxiliary tool for the computational of several quantities needed for our analysis. Given a reference parameter vector $\hat{\boldsymbol{\theta}}$, the w-POD constructs an adaptive space $\mathbf{V}_{\hat{\boldsymbol{\theta}}}$ by minimizing

$$\frac{1}{N}\sum_{i=1}^{N}\|\omega(\hat{\boldsymbol{\theta}},\boldsymbol{\theta}_i)\boldsymbol{u}_{\boldsymbol{\theta}_i}-\mathbf{V}_{\hat{\boldsymbol{\theta}}}\mathbf{V}_{\hat{\boldsymbol{\theta}}}^T\omega(\hat{\boldsymbol{\theta}},\boldsymbol{\theta}_i)\boldsymbol{u}_{\boldsymbol{\theta}_i}\|^2,$$

where $\{\boldsymbol{u}_{\boldsymbol{\theta}_i},\boldsymbol{\theta}_i\}_{i=1}^N$ represent a collection of N snapshots obtained from a FOM, and a Gaussian kernel $\omega(\hat{\boldsymbol{\theta}},\boldsymbol{\theta}_i) = \exp\left(-\|\hat{\boldsymbol{\theta}}-\boldsymbol{\theta_i}\|^2/\lambda^2\right)$ controls the locality of the weighting. This yields a locally-adapted basis via SVD, without requiring any neural network training phase. The w-POD thus emulates the flexible behavior of adaptive-basis methods and, thanks to its closed-form solution, guarantees the optimality of the found basis, making it ideal for pre-training analysis. At the same time, w-POD remains unsuited for online inference due to its associated computational costs, reason for which we only use it in the offline phase of our proposed pipeline.

2.2. Adaptivity Scores via the Grassmann Manifold

To measure how much an adaptive-basis actually varies as a function of each parameter, the outputs of such methods are treated as elements of the *Grassmann manifold* $\mathcal{G}_n(\mathcal{A})$ [1], the space of all n-dimensional subspaces of a given ambient space $\mathcal{A}$. By noting that each subspace $\mathcal{V}_{\boldsymbol{\mu}} = \mathrm{span}(\mathbf{V}_{\boldsymbol{\mu}}) \subseteq \mathcal{A}$ is uniquely identified by its orthogonal projector $\mathbf{P}_{\boldsymbol{\mu}} = \mathbf{V}_{\boldsymbol{\mu}}\mathbf{V}_{\boldsymbol{\mu}}^T$, we can define an appropriate distance between the produced basis as

$$d_{\mathcal{G}}(\mathbf{V}_{\boldsymbol{\mu}},\mathbf{V}_{\boldsymbol{\mu}'}) := \|\mathbf{P}_{\boldsymbol{\mu}}-\mathbf{P}_{\boldsymbol{\mu}'}\|_{\mathrm{F}},$$

with $\|\cdot\|_{\mathrm{F}}$ the Frobenius norm, so that the structure of the Grassmann manifold is preserved. Leveraging this distance, two adaptivity scores are then defined to rank parameters by their influence on the obtained basis:

- a derivative-based score, K_j^{tot}, measuring the average maximum rate of change of the basis along the j-th parametric direction, obtained via Monte Carlo sampling combined with a numerical optimization routine;
- a variance-based score, D_j, adapted from first-order Sobol' sensitivity indices, which compares the expected conditional dispersion (when the j-th component of $\boldsymbol{\mu}$ is fixed) to the total dispersion across the parameter domain, again approximated via Monte Carlo sampling.

High scores on either metric identify parameters that should be assigned to the adaptive set $\boldsymbol{\mu}$; low scores instead suggest they belong to the static set $\boldsymbol{\nu}$.

2.3. Rank Selection via Singular Value Analysis

Once the parameter partition is established, the pipeline determines the appropriate dimension of latent space. By examining the spectrum of singular values produced by the w-POD across the parameter domain, two diagnostic quantities are proposed: *(i)* the *gap* δ, defined as the relative difference between successive singular values, and *(ii)* the *range* of the singular values, useful to measure the consistency of performance across different parametric configurations. A rank with a compact range and a large gap is preferred, as it implies, respectively, both a consistent representation and favorable learning conditions [2, Corollary 4.2] for the downstream adaptive model.

3. Proposed Pipeline

By combining the two adaptivity scores, and the rank analysis, we propose a design pipeline for arbitrary adaptive-basis methods. The complete diagnostic workflow proceeds in five steps. It is assumed that FOM snapshots have already been collected.

1. Initialize the w-POD with a heuristic rank $n = \dim(\boldsymbol{\theta}) + 1$, motivated by a first-order Taylor expansion, and sweep over different bandwidth values λ.
2. Compute the two adaptivity scores for the full parameter vector $\boldsymbol{\theta}$.
3. Partition $\boldsymbol{\theta}$ into $\boldsymbol{\mu}$ (high scores) and $\boldsymbol{\nu}$ (low scores).
4. Perform the singular value analysis to optimize the rank of the final basis. This step, can be carried out using a different bandwidth, e.g. $\lambda_{\text{rank}} = 10^{-1}\lambda_{\text{scores}}$.
5. Hand off the optimized architectural choices to the actual adaptive-basis model for training.

The pipeline is training-free, requiring only standard SVD and score computations on the snapshot dataset.

4. Numerical Results

4.1. Analytical Test Case

The pipeline was first validated on the aforementioned toy problem, consisting of a synthetic signal composed of two Gaussian pulses, where positions $[\theta_1, \theta_2]$ and amplitudes $[\theta_3, \theta_4]$ represent the four parameters. The variance-based score correctly identified θ_1 and θ_2 as the dominant adaptive parameters, reporting essentially null scores for θ_3 and θ_4. The singular value analysis correctly suggests $n = 2$ as the basis dimension, consistent with the known two-component structure of the signal.

4.2. Navier-Stokes Benchmark: Flow Past a Cylinder

The pipeline was then applied to the unsteady incompressible Navier-Stokes equations on a channel domain with a cylindrical obstacle. For this test case, the dataset was restructured as collection of steady snapshots, with time being reinterpreted as a problem parameter. Then, the parameter vector consisted of $\boldsymbol{\theta} = [1/\rho, \eta, t]$, that is, the inverse density, the kinematic viscosity, and the time variable, respectively. The scores, as seen from Table 1, overwhelmingly identified time as the primary adaptive parameter, while the physical parameters received significantly lower scores. The pipeline graphs in Fig. 1 recommended $n = 4$ as the optimal basis dimension, maintaining a compact range and avoiding gap values that were too low.

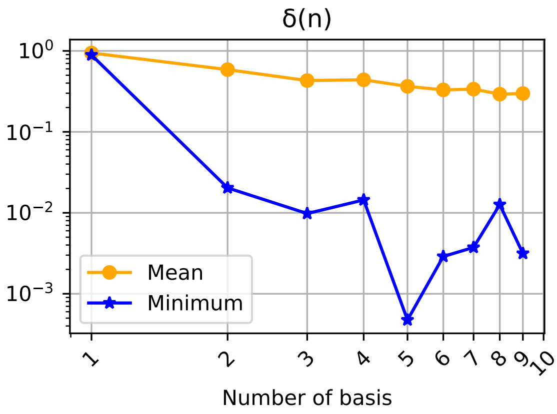
	$\boldsymbol{K_j^{\text{tot}}}$	$\boldsymbol{D_j}$
$\boldsymbol{1/\rho}$	4.136×10^{-1}	0.000
$\boldsymbol{\eta}$	6.036×10^{-2}	0.000
$\boldsymbol{t}$	1.856×10^{1}	9.311×10^{-1}

Table 1: Adaptivity scores for the **Navier-Stokes test case**.

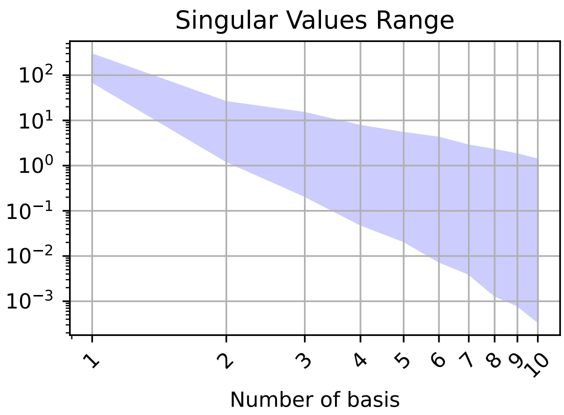
To further validate these findings, an extensive experimental campaign was conducted using the DOD. Training multiple DOD instances with different parameter splits and basis sizes confirmed all pipeline predictions: models using time as the sole adaptive parameter significantly outperformed those ignoring it. Furthermore, we were also able to study in this phase how to best link the original DOD framework with time-dependent problems, showing that a Fourier-enriched time-handling architectures consistently outperformed standard ones. The observed performance stagnation between $n = 5$ and $n = 6$, as predicted by the spectral gap analysis, was clearly visible in experimental error curves.

4.3. Linear Elasticity: the Footing Problem

The full surrogate modeling pipeline was finally applied to a time-dependent linear elasticity problem modeling the deformation of a layered domain under a load on the top surface. The parameter vector consisted of $\boldsymbol{\theta} = [\delta, \gamma, m, t]$, that is, layer thickness, load angle, load mass, and the time variable, respectively. The pipeline, as shown in Table 2, identified time and δ as the two dominant adaptive parameters, with $n = 4$ recommended as the optimal rank.



(a) Gap graphs: mean and average values of δ across different parameter configurations.



(b) Range graph.

Figure 1: Singular values graphs for the **Navier-Stokes test case**. The graphs suggest $n = 4$, as that results in a large gap and a quite compact range.

	K_j^{tot}	D_j
δ	1.210	4.880×10^{-2}
γ	5.018×10^{-1}	0.000
m	5.572×10^{-1}	0.000
t	5.871	6.004×10^{-1}

Table 2: Adaptivity scores for the **linear elasticity test case**.

As the diagnostic pipeline was shown to produce the correct choices in the previous test cases, the obtained results were now utilized to construct a full surrogate model. The choice was made to utilize the DOD-NN, which links the already tested DOD, used for basis production, to another neural network for the prediction of the corresponding coefficients. The numerical results were promising. A DOD model trained with $\boldsymbol{\mu} = [\delta, t]$ allowed for an accurate reconstruction of the FOM solutions when projected onto the produced latent space, obtaining a mean relative projection error of 2.95%, against a benchmark of 7.93% obtained with a global POD basis. The full DOD-NN surrogate produced solutions visually indistinguishable (Fig. 2) from the FOM reference, scoring a reconstruction error of 4.46%, and marking the first application of the DOD framework to a time-dependent surrogate setting, and greatly lowering the computational cost for the solution of this problem.

5. Conclusions

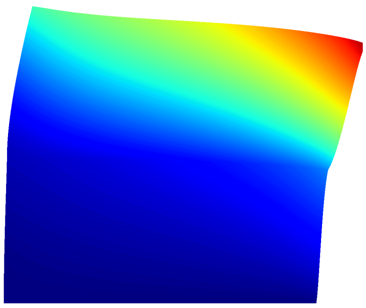
This thesis proposed and validated a complete a priori diagnostic pipeline for adaptive-basis ROM configuration. Its main contributions are:

- two novel adaptivity scores, one derivative-based and one variance-based, grounded in the geometry of the Grassmann manifold, providing a rigorous quantitative ranking of parameters for the partition $\boldsymbol{\theta} \to [\boldsymbol{\mu}, \boldsymbol{\nu}]$;
- a singular value diagnostic framework, enabling pre-training prediction of model performance and optimal rank selection;
- the first application of DOD and DOD-NN to time-dependent problems, significantly broadening the applicability of the two strategies.

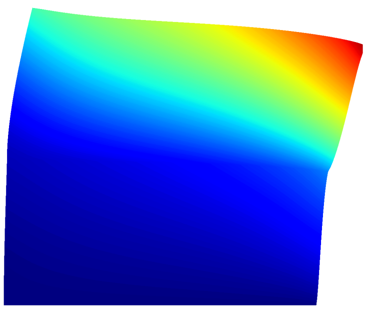
Across all test cases, the pipeline successfully replaced heuristic trial-and-error with systematic, data-driven architectural decisions. Results were consistently corroborated by trained model experiments.
Future work may focus on removing the sensitivity to the bandwidth hyperparameter λ by developing hyperparameter-free weighting strategies.

References

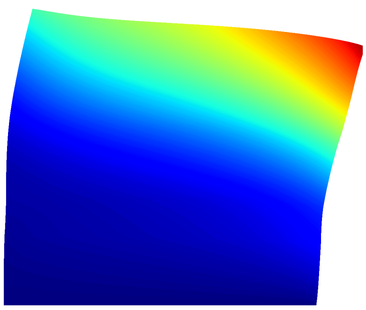
[1] Thomas Bendokat, Ralf Zimmermann, and P.-A. Absil. Correction: A grassmann man-



(a) FOM solution.



(b) Projection onto the DOD space.



(c) DOD-NN surrogate.

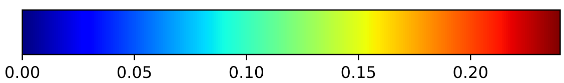


Figure 2: FOM solution vs projection onto the DOD space vs DOD-NN prediction. The deformation is represented by warping the originally square domain. The color, instead, represents the magnitude of the deformation vector.

ifold handbook: basic geometry and computational aspects. *Advances in Computational Mathematics*, 50(2), March 2024.

[2] Nicola Rares Franco. Measurability and continuity of parametric low-rank approximation in hilbert spaces: linear operators and random variables. *Revista Matemática Complutense*, pages 1–42, 2025.

[3] Nicola Rares Franco, Andrea Manzoni, Paolo Zunino, and Jan S. Hesthaven. Deep orthogonal decomposition: a continuously adaptive data-driven approach to model order reduction. *Advances in Computational Mathematics*, accepted, 2026.

[4] Jan S Hesthaven, Gianluigi Rozza, and Benjamin Stamm. *Certified Reduced Basis Methods for Parametrized Partial Differential Equations*. Springer International Publishing, 2016.

[5] Alfio Quarteroni, Andrea Manzoni, and Federico Negri. *Reduced Basis Methods for Partial Differential Equations*. Springer International Publishing, 2016.