

MOX-Report No. 44/2025

Handling geometrical variability in nonlinear reduced order modeling through Continuous Geometry-Aware DL-ROM

Brivio, S.; Fresca, S.; Manzoni, A.

MOX, Dipartimento di Matematica
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

mox-dmat@polimi.it

<https://mox.polimi.it>

Handling geometrical variability in nonlinear reduced order modeling through Continuous Geometry-Aware DL-ROMs

Simone Brivio^a, Stefania Fresca^a, Andrea Manzoni^a

^aMOX – Department of Mathematics, Politecnico di Milano, Piazza Leonardo da Vinci 32, I-20133, Milano, Italy

Abstract

Deep Learning-based Reduced Order Models (DL-ROMs) constitute a consolidated class of techniques that aim at providing accurate surrogate models for complex physical systems described by Partial Differential Equations (PDEs) by nonlinearly compressing the solution manifold into a handful of latent coordinates. Until now, the development of DL-ROMs mainly focused on physically parameterized problems. Within this work we provide a novel extension of these architectures to problems featuring geometrical variability and parametrized domains, namely, we propose Continuous Geometry-Aware DL-ROMs (CGA-DL-ROMs). Specifically, we emphasize that the space-continuous nature of the proposed architecture matches the necessity to deal with *multi-resolution* datasets, which are quite common in the case of geometrically parametrized problems. Moreover, CGA-DL-ROMs are endowed with a strong inductive bias that makes them aware of geometrical parametrization, thus enhancing both the compression capability and the overall performance of the architecture. Within this work we justify our findings through a suitable theoretical analysis and we experimentally validate our claims by means of a series of numerical tests encompassing physically-and-geometrically parametrized differential problems ranging from the unsteady Navier-Stokes equations for fluid dynamics to advection-diffusion-reaction equations for mathematical biology.

1. Introduction

Nowadays, Partial Differential Equations (PDEs) constitute the most widely used tool to describe complex physical systems from a mathematical viewpoint and, for this reason, they are of substantial importance in science and engineering, thus finding application in diverse areas ranging from life sciences to finance and structural mechanics. Since the analytical solution of PDEs is not in general available in a closed form, in the last two centuries several efforts were directed towards the development of numerical methods that strive to approximate the exact solution by means of Full Order Models (FOMs). The latter involve the discretization of the underlying governing equations on a mesh of step size $h > 0$, thus yielding N_h degrees of freedom (dofs). We emphasize that FOMs feature an outstanding accuracy but require substantial computational resources, for instance when tackling time-dependent problems, nonlinear terms and in the case of a large number of dofs. The latter constitute a major drawback, especially when dealing with *many query* applications involving parametric PDEs – e.g., uncertainty quantification, sensitivity analysis and optimal control – which entail the repeated solution of the numerical problem for different configurations of the parameters’ instance.

ROMs: an historical perspective. In the last decades, the need of *real-time* simulations in *many-query* frameworks paved the way to the development of Reduced Order Models (ROMs), which provide fast and efficient alternatives to FOMs while retaining the most essential features of the underlying physical problem. Specifically, ROMs were originally proposed for problems depending on a set of physical parameters $\boldsymbol{\mu} \in \mathcal{P}$ – and possibly the time variable $t \in \mathcal{T}$ – where \mathcal{P}, \mathcal{T} are compacts finite-dimensional sets. Among the most adopted ROM techniques, we mention classical reduced basis methods (RB) [46, 3, 2, 4], which seek a low dimensional representation of the solution manifold through linear subspaces. However, we mention that the trial manifold entailed by RB methods is global and linear: for this reason, it reproduces inefficiently the local features and the nonlinearities in the solution manifold. Indeed, the speedup of classical RB methods (with respect to the corresponding FOM) is negligible in the case of real-world applications involving nonlinear and non-affine terms. The limitations of classical RB methods paved the way to the development of deep learning-based alternatives. As a matter of fact, in the literature has emerged a wide variety of deep learning

based approaches that include, but are not limited to, Neural Operators [23, 36, 30, 53, 25, 31, 32, 8, 20], DL-ROMs [14, 15, 43, 42] and transformer architectures [18, 28]. Both Neural Operators and DL-ROMs are equipped with a strong theoretical background that backs up their outstanding approximation capabilities [37, 26, 25, 23, 7, 11]. On the other hand, even though transformers and attention-based mechanisms [22, 19, 18] provide state-of-the-art prediction accuracy and grant us the possibility to tackle large datasets, they are less interpretable and not fully understood from a theoretical viewpoint as only some architectures are supplied with a comprehensive theoretical study [9, 45]. Nonetheless, we stress that Neural Operators often consist in complex models featuring a high-dimensional latent space; conversely, DL-ROMs both accurately reproduce the variability of the solution manifold and entail a profound nonlinear dimensionality reduction, ultimately resolving the underlying dynamics into a handful of latent coordinates.

Surrogate models for geometrically parameterized problems. Recently, many authors’ focus shifted towards the development of ROMs for challenging differential problems characterized by geometrical parameters and parametric domain shapes. For instance, classical reduced basis (RB) methods were extended to cope with geometrically parametrized problems [39, 41, 27, 21, 49, 1]. Alternatively, more efficient solutions were proposed in the context of deep learning-based ROMs. A first class of strategies comprise the coupling of proper orthogonal decomposition (POD) and neural networks [50]. However, the discrete nature of the approach limit the range of possible applications to diffeomorphic meshes with the same resolution. Other possible approaches make use of operator learning paradigms in infinite-dimensional function spaces, which are more suitable for parametric domains since they are mesh-agnostic. For example, one may consider GNN approaches in an operator learning framework [38, 12, 44], which are very expressive and capable to reproduce the local features of the solution field with high precision. However, GNNs are data eager, that is, they need a large amount of training data to achieve a suitable accuracy on the test set. Moreover, their graph-based architecture entails a large memory footprint in the case of complex problems comprising high-dimensional data. Among other techniques, domain practitioners can take advantage of Implicit Neural Representations (INR), a class of approaches which is still at its infancy for applications involving PDEs and therefore is not equipped with theoretical results. We also remark that INR-based strategies such as [48] may require inference-time optimization and are therefore expensive to evaluate in a *many-query* context. For the sake of completeness, we note that some authors crafted suitable transformer-based models to deal with geometrically parametrized problems [28, 18]. However, these strategies often involve complex and heavy architectures, and are quite data eager, so that their accuracy is particularly sensitive to the number of available input-output pairs. More remarkably, several approaches successfully adapted Neural Operators to handle geometrical parametrization [29, 33, 52, 34]: the proposed frameworks are extensions of existing architectures but they lack an inductive bias tailored for geometrically parametrized problems. For this reason, even though they provide strikingly accurate approximations, they often feature large neural network architectures, which are prone to a slow training and evaluation. On the other hand, DL-ROMs (and their extensions) usually entail smaller architectures but were originally conceived in the context of physical variability and they were never adapted to problems featuring geometrical parameters.

Main contributions and work outline. The main contribution of our work consists in the extension of the classical DL-ROM architecture to widen its range of applications to parametric differential problems featuring geometrical variability. Specifically,

- We justify the need of devising a space-continuous, infinite-dimensional paradigm to deal with *multi-resolution* datasets encountered in problems featuring geometrical variability.
- We propose Continuous Geometry-Aware DL-ROMs (CGA-DL-ROMs), a comprehensive learning framework to construct DL-ROMs for geometrically parameterized problems, which (i) is based on a space-continuous formulation to account for *multi-resolution* datasets, and (ii) leverages on geometry-aware basis functions that characterize the notable approximation capabilities of the proposed architecture.
- We analyze the proposed framework, thus describing its main properties and assessing its performance against suitable baselines both in light of the theory and by means of involved numerical experiments.

The present work is organized as follows. In §2 we describe the general formulation of problems featuring geometrical variability; on top of that, in §3 we characterize CGA-DL-ROMs for geometrically parameterized problems. Moreover, in §4 we describe the main features of CGA-DL-ROMs, focusing on the analysis of their peculiar geometry-aware basis functions, and we show how to parametrize such basis functions with neural networks. We report in §5 a series of suitable numerical experiments that validate the theoretical claims on benchmarks and assess the generalization capabilities on challenging applications. Finally, we draw some conclusions on the present work, providing also an outline over possible future developments.

2. Geometrically parametrized PDEs: formulation and properties

Within the present work we consider differential problems parametrized by a set of physical parameters, namely $\boldsymbol{\mu} \in \mathcal{P} \subset \mathbb{R}^p$, and a set of geometrical parameters, represented by $\boldsymbol{\xi} \in \mathcal{G} \subset \mathbb{R}^g$ for some $p, g \geq 0$:

$$\begin{cases} \partial_t u + \mathcal{N}[\boldsymbol{\mu}, \boldsymbol{\xi}](u) = 0, & \Omega(\boldsymbol{\xi}) \times (0, T] \\ \mathcal{B}[\boldsymbol{\mu}, \boldsymbol{\xi}](u) = 0, & \partial\Omega(\boldsymbol{\xi}) \times (0, T] \\ u(0, \boldsymbol{\mu}, \boldsymbol{\xi}) - u_0(\boldsymbol{\mu}, \boldsymbol{\xi}) = 0, & \Omega(\boldsymbol{\xi}), \end{cases} \quad (1)$$

where $T > 0$, \mathcal{N} is a generic nonlinear operator, whereas \mathcal{B} enforces the boundary conditions and $u_0(\boldsymbol{\mu}, \boldsymbol{\xi})$ is the initial datum. For the sake of the well-posedness of the problem, from hereon we consider \mathcal{P} and \mathcal{G} as compact subsets of \mathbb{R}^p and \mathbb{R}^g , respectively. Then, we define $\mathcal{T} = [0, T]$ as the time domain. We emphasize that the domain shape is parametrized, that is, $\Omega = \Omega(\boldsymbol{\xi}) \subset \mathbb{R}^d$, for $d > 0$: we refer the reader to Fig. 1 for a visualization of an example of parametric domains. Then, we highlight that the solution field $u = u(\mathbf{x}; t, \boldsymbol{\mu}, \boldsymbol{\xi}) \in \mathbb{R}$ depends on both physical and geometrical parameters as well as the time variable t , for any $\mathbf{x} \in \Omega(\boldsymbol{\xi})$. We also remark that from hereon we assume that the solution field is scalar for the sake of simplicity, but we observe that the results can be generalized to vector problems.

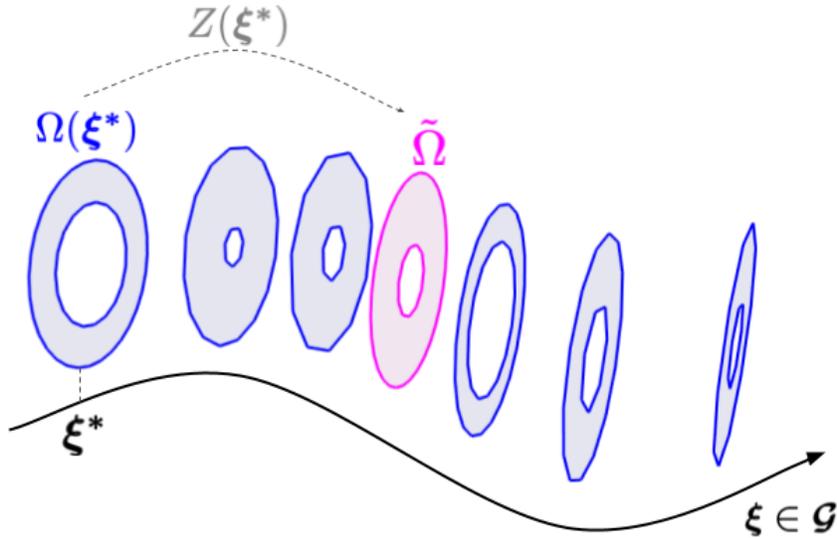


Figure 1. Visualization of the diffeomorphism and a set of parametric domains. The geometrical parameter ξ regulates the radius of the hole.

2.1. Abstract setting

Within this section we aim at proposing a suitable abstract setting for the analysis of the theoretical properties and the well-posedness of problem (1). The main difficulty of the aforementioned analysis originates from the fact that, for any instance $\boldsymbol{\xi} \in \mathcal{G}$ of the geometrical parameter, $u(t, \boldsymbol{\mu}, \boldsymbol{\xi})$ refer to different, parametrized domains $\Omega(\boldsymbol{\xi})$. In this respect, in the following we show that it is possible to perform a change of variables to cast the differential problem entailed by (1) in a pre-defined reference configuration $\tilde{\Omega}$ that

does not depend on the geometrical parameter $\boldsymbol{\xi}$. To do that, we introduce a special mapping that characterizes the deformation from the original configuration $\Omega(\boldsymbol{\xi})$ onto the reference domain $\tilde{\Omega}$ (and viceversa), namely,

Assumption 2.1. *Let $\tilde{\boldsymbol{\xi}} \in \mathcal{G}$ and define $\tilde{\Omega} = \Omega(\tilde{\boldsymbol{\xi}})$ as the reference domain. Then, there exists $(\mathbf{x}, \boldsymbol{\xi}) \mapsto Z(\mathbf{x}; \boldsymbol{\xi})$ which is a C^r -diffeomorphism ($r \geq 1$) as function of $\mathbf{x} \in \mathbb{R}^d$ and Lipschitz as function of $\boldsymbol{\xi} \in \mathcal{G}$, such that:*

$$\Omega(\boldsymbol{\xi}) = \{\mathbf{x} = Z^{-1}(\tilde{\mathbf{x}}; \boldsymbol{\xi}) \in \mathbb{R}^d, \text{ for any } \tilde{\mathbf{x}} \in \tilde{\Omega}\}, \quad \forall \boldsymbol{\xi} \in \mathcal{G}.$$

Thus, we define $\zeta(\mathbf{x}; \boldsymbol{\xi}) := |\nabla_{\mathbf{x}} Z(\mathbf{x}; \boldsymbol{\xi})|$ as the Jacobian of the transformation Z .

In Assumption 2.1 we do not only aim at providing a definition of the mapping Z , but we also characterize its regularity. To this end, we stress that requiring $(\cdot, \boldsymbol{\xi}) \mapsto Z(\cdot, \boldsymbol{\xi})$ to be at least homeomorphic, for any $\boldsymbol{\xi} \in \mathcal{G}$, ensures that the deformation from the original configuration $\Omega(\boldsymbol{\xi})$ and the reference domain $\tilde{\Omega}$ preserves its topological invariants such as connectedness and compactness. In this way, it is straightforward to conclude that all the parametric domains share the same topological properties. However, being homeomorphic is not usually sufficient. Indeed, we must require $Z(\cdot, \boldsymbol{\xi})$ and its inverse to be sufficiently smooth in order to cast (1) in the reference configuration by performing the change of variable $\mathbf{x} = Z^{-1}(\tilde{\mathbf{x}}; \boldsymbol{\xi})$, namely,

$$\begin{cases} \partial_t \tilde{u} + \tilde{\mathcal{N}}[\boldsymbol{\mu}, \boldsymbol{\xi}](\tilde{u}) = 0, & \tilde{\Omega} \times (0, T] \\ \tilde{\mathcal{B}}[\boldsymbol{\mu}, \boldsymbol{\xi}](\tilde{u}) = 0, & \partial\tilde{\Omega} \times (0, T] \\ \tilde{u}(0, \boldsymbol{\mu}, \boldsymbol{\xi}) - \tilde{u}_0(\boldsymbol{\mu}, \boldsymbol{\xi}) = 0, & \tilde{\Omega}, \end{cases} \quad (2)$$

where $\tilde{u}(\tilde{\mathbf{x}}; t, \boldsymbol{\mu}, \boldsymbol{\xi}) = u(Z^{-1}(\tilde{\mathbf{x}}; \boldsymbol{\xi}); t, \boldsymbol{\mu}, \boldsymbol{\xi})$ for any $\mathbf{x} = Z^{-1}(\tilde{\mathbf{x}}; \boldsymbol{\xi}) \in \Omega(\boldsymbol{\xi})$ and $\tilde{\mathcal{N}}$ and $\tilde{\mathcal{B}}$ are the counterparts of the nonlinear operator \mathcal{N} and the linear operator \mathcal{B} on the reference domain, namely $\tilde{\mathcal{N}}[\tilde{\mathbf{x}}; \boldsymbol{\mu}, \boldsymbol{\xi}] = \mathcal{N}[Z^{-1}(\tilde{\mathbf{x}}; \boldsymbol{\xi}); \boldsymbol{\mu}, \boldsymbol{\xi}]$ and $\tilde{\mathcal{B}}[\tilde{\mathbf{x}}; \boldsymbol{\mu}, \boldsymbol{\xi}] = \mathcal{B}[Z^{-1}(\tilde{\mathbf{x}}; \boldsymbol{\xi}); \boldsymbol{\mu}, \boldsymbol{\xi}]$. We stress that the smoothness coefficient $r \geq 1$ is related to the highest derivative degree entailed by \mathcal{N} and \mathcal{B} . It is now evident that problem (2) is no longer geometrically parametrized: indeed, in problem (2) $\boldsymbol{\xi} \in \mathcal{G}$ plays the role of a mere physical parameter. For this reason, it is possible to proceed through classical arguments to establish the well-posedness of problem (2): we refer the interested reader to [46] and to Appendix A for insights on the formal proof. As a final remark, even though within the present section we only considered the strong formulation for the sake of simplicity, we mention that a similar analysis can be carried out by employing the variational formulation.

We are now equipped with suitable theoretical tools to adequately characterize geometrical variability in differential problems: such tools are essential in the following sections to describe the synthetic data generation phase and provide further insights on how the geometrical variability impacts on the dimensionality reduction task.

2.2. Synthetic data generation

Within the present section, we aim at characterizing the training/testing dataset generation in the context of problems featuring geometrical variability. Specifically, since, in general, the exact solution of problem (1) is not available in exact form, we harness a high-fidelity solver to obtain a suitably accurate numerical approximation. Specifically, in the following we distinguish two different strategies to synthetically generate data samples.

We emphasize that if the pair $(Z, \tilde{\Omega})$ is known and fixed a priori, it is possible to systematically generate the realizations of $\boldsymbol{\xi} \mapsto \Omega(\boldsymbol{\xi})$. In the literature several approaches were proposed to suitably define Z ; among them, we mention Free-Form Deformations (FFD) [39, 41, 27], Radial Basis Function (RBF) interpolation [39, 50, 10] or mesh motion techniques [40]. When dealing with such strategies, it is in general possible to generate high-fidelity solutions by deforming (through Z) an adequately rich set of high-fidelity basis functions defined on the reference domain. Formally, being W a suitable Hilbert space, we characterize the *fixed resolution* approach through the following Assumption, namely,

Assumption 2.2. *We suppose that for any $\varepsilon > 0$ there exist N_h and a finite dimensional subspace $W \ni W_h = \text{span}\{\tilde{\phi}_i\}_{i=1}^{N_h}$, where $\psi_i^* \in L^\infty(\tilde{\Omega})$ for $i = 1, \dots, N_h$, such that*

$$\sup_{(t, \boldsymbol{\mu}, \boldsymbol{\xi}) \in \mathcal{T} \times \mathcal{P} \times \mathcal{G}} \|u(t, \boldsymbol{\mu}, \boldsymbol{\xi}) - u_h(t, \boldsymbol{\mu}, \boldsymbol{\xi})\|_* < \varepsilon,$$

where $\|\cdot\|_*$ is a suitable norm and $u_h(\mathbf{x}; t, \boldsymbol{\mu}, \boldsymbol{\xi}) = \sum_{i=1}^{N_h} u_{h,i}(t, \boldsymbol{\mu}, \boldsymbol{\xi}) \tilde{\phi}_i^*(Z(\mathbf{x}, \boldsymbol{\xi}))$ for any $t \in \mathcal{T}$, $\boldsymbol{\mu} \in \mathcal{P}$, $\boldsymbol{\xi} \in \mathcal{G}$ and $\mathbf{x} \in \Omega(\boldsymbol{\xi})$.

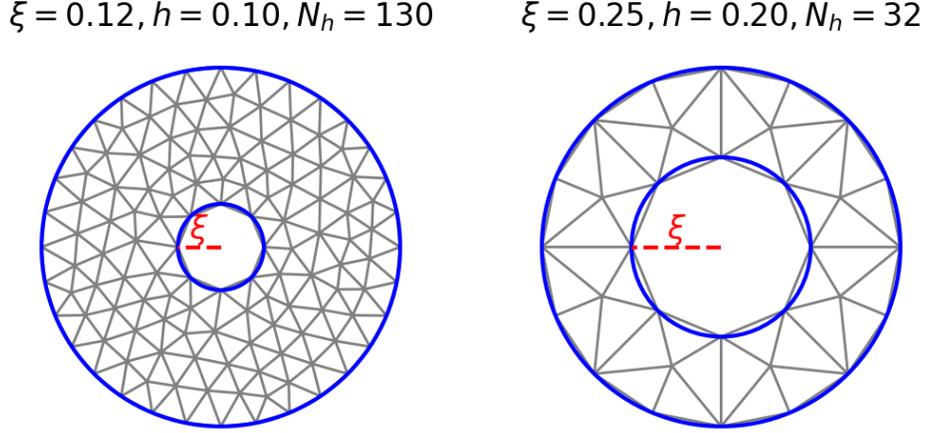


Figure 2. Example of parametric domains with different resolutions (the geometrical parameter ξ is the radius of the hole).

We remark that Assumption 2.2 relies on the assumption that the pair $(Z, \tilde{\Omega})$ must be known explicitly, which is not always the case in real applications. Moreover, it is worth remarking that within this approach the number of dofs N_h is the same for all data samples: for this reason it is quite common in the ROM literature since the majority of techniques rely on a space-discrete formulation. However, while certain domain configurations may require fine discretizations in order to capture possible low-scale effects, for other domain instances a more coarse discretization is sufficient to properly grasp the underlying physics. Thus, in order to limit the dataset size by avoiding the storage of redundant information, we may choose to focus on *multi-resolution* datasets, namely,

Assumption 2.3. We suppose that for any $\varepsilon > 0$ and any $\boldsymbol{\xi} \in \mathcal{G}$ there exist $N_h = N_h(\boldsymbol{\xi})$ and a finite dimensional subspace $W \ni W_h = \text{span}\{\phi_i(\boldsymbol{\xi})\}_{i=1}^{N_h}$, where $\phi_i(\boldsymbol{\xi}) \in L^\infty(\Omega(\boldsymbol{\xi}))$ for $i = 1, \dots, N_h$, such that

$$\sup_{(\boldsymbol{\mu}, t) \in \mathcal{T} \times \mathcal{P}} \|u(t, \boldsymbol{\mu}, \boldsymbol{\xi}) - u_h(t, \boldsymbol{\mu}, \boldsymbol{\xi})\|_* < \varepsilon,$$

where $\|\cdot\|_*$ is a suitable norm and $u_h(\mathbf{x}; t, \boldsymbol{\mu}, \boldsymbol{\xi}) = \sum_{i=1}^{N_h} u_{h,i}(t, \boldsymbol{\mu}, \boldsymbol{\xi}) \phi_i(\mathbf{x}, \boldsymbol{\xi})$ for any $t \in \mathcal{T}$, $\boldsymbol{\mu} \in \mathcal{P}$, $\boldsymbol{\xi} \in \mathcal{G}$ and $\mathbf{x} \in \Omega(\boldsymbol{\xi})$.

We then emphasize that, differently from the *fixed resolution* approach, the *multi-resolution* synthetic data generation does not require the availability of $(Z, \tilde{\Omega})$ in a closed form.

Nevertheless, with either the *fixed resolution* or *multi-resolution* approach we obtain an high-fidelity formulation of the semidiscretized problem that reads

$$\begin{cases} \partial_t \mathbf{u}_h + \mathbf{N}[\boldsymbol{\mu}, \boldsymbol{\xi}](\mathbf{u}_h) = 0, & (0, T] \\ \mathbf{u}(0, \boldsymbol{\mu}, \boldsymbol{\xi}) - \mathbf{u}_0(\boldsymbol{\mu}, \boldsymbol{\xi}) = 0, \end{cases} \quad (3)$$

where $\mathbf{u}_h := \mathbf{u}_h(t, \boldsymbol{\mu}, \boldsymbol{\xi}) \in \mathbb{R}^{N_h}$ is the solution vector, $\mathbf{N} : \mathbb{R}^{N_h} \rightarrow \mathbb{R}^{N_h}$ is a generic nonlinear mapping that encodes the contribution of both \mathcal{N} and \mathcal{B} at a discrete level, whereas $\mathbf{u}_0(\boldsymbol{\mu}, \boldsymbol{\xi})$ is the parameter-dependent initial condition. We emphasize that, subject to the choice of the discretization strategy, N_h possibly depends on $\boldsymbol{\xi}$. Since the discretization is in principle parameter dependent (see Fig. 2), in the following we cannot harness a finite-dimensional formulation: the entire content of this paper is therefore based upon a infinite-dimensional setting, allowing us to handle generic datasets obtained through *multi-resolution*-based synthetic data generation.

2.3. An infinite-dimensional setting for dimensionality reduction

In the following, we first define a suitable infinite-dimensional functional setting and then characterize the solution manifold of the parametric problem defined in Eq. (1). We stress that the following results are relative to the L^2 setting but can be extended to other separable Hilbert spaces. Specifically, for any $\boldsymbol{\xi} \in \mathcal{G}$ we define

$$L^2_\zeta(\Omega(\boldsymbol{\xi})) = \left\{ f : \Omega(\boldsymbol{\xi}) \rightarrow \mathbb{R} \mid \int_{\Omega(\boldsymbol{\xi})} f(\mathbf{x})^2 \zeta(\mathbf{x}, \boldsymbol{\xi}) d\mathbf{x} < +\infty \right\},$$

identifying the scalar product as $(\cdot, \cdot)_{L^2_\zeta(\Omega(\boldsymbol{\xi}))}$ and its induced norm as $\|\cdot\|_{L^2_\zeta(\Omega(\boldsymbol{\xi}))}$. Thus, we aim at delineating a proper solution manifold as a bounded subset of a single separable Hilbert space. However, it is evident that the collection

$$\mathcal{S} = \{u(t, \boldsymbol{\mu}, \boldsymbol{\xi}) \in L^2_\zeta(\Omega(\boldsymbol{\xi})) \mid \boldsymbol{\mu} \in \mathcal{P}, t \in \mathcal{T}, \boldsymbol{\xi} \in \mathcal{G}\}.$$

is an ill-defined solution manifold in a conventional sense since each solution $u(t, \boldsymbol{\mu}, \boldsymbol{\xi})$ is characterized on a different, parameter-dependent Hilbert space, namely $L^2_\zeta(\Omega(\boldsymbol{\xi}))$. On the other hand, we observe that, thanks to $Z(\cdot, \boldsymbol{\xi})$ being a C^r -diffeomorphism ($r \geq 1$), the following change of variable formula holds for any $f \in L^2_\zeta(\Omega(\boldsymbol{\xi}))$,

$$\int_{\Omega(\boldsymbol{\xi})} f(\mathbf{x})^2 \zeta(\mathbf{x}, \boldsymbol{\xi}) d\mathbf{x} = \int_{\tilde{\Omega}} f(Z^{-1}(\tilde{\mathbf{x}}, \boldsymbol{\xi}))^2 d\tilde{\mathbf{x}}. \quad (4)$$

Thus, it is possible to show that, for any $f, g \in L^2_\zeta(\Omega(\boldsymbol{\xi}))$, it is valid that

$$\begin{aligned} (f, g)_{L^2_\zeta(\Omega(\boldsymbol{\xi}))} &= \int_{\Omega(\boldsymbol{\xi})} f(\mathbf{x})g(\mathbf{x})\zeta(\mathbf{x}, \boldsymbol{\xi})d\mathbf{x} \\ &= \int_{\tilde{\Omega}} f(Z^{-1}(\tilde{\mathbf{x}}, \boldsymbol{\xi}))g(Z^{-1}(\tilde{\mathbf{x}}, \boldsymbol{\xi}))d\tilde{\mathbf{x}} = (f \circ Z^{-1}(\boldsymbol{\xi}), g \circ Z^{-1}(\boldsymbol{\xi}))_{L^2(\tilde{\Omega})}, \end{aligned}$$

where $\tilde{f}(\tilde{\mathbf{x}}, \boldsymbol{\xi}) = f(Z^{-1}(\tilde{\mathbf{x}}, \boldsymbol{\xi}))$ and $\tilde{g}(\tilde{\mathbf{x}}, \boldsymbol{\xi}) = g(Z^{-1}(\tilde{\mathbf{x}}, \boldsymbol{\xi}))$, for any $\boldsymbol{\xi} \in \mathcal{G}$ and $\tilde{\mathbf{x}} \in \tilde{\Omega}$. Thus, upon defining the (linear) morphing operators

$$\begin{aligned} \mathcal{Z}_\boldsymbol{\xi} &: L^2(\tilde{\Omega}) \rightarrow L^2_\zeta(\Omega(\boldsymbol{\xi})) & \text{as } \tilde{f} \mapsto \tilde{f} \circ Z(\boldsymbol{\xi}) \\ \mathcal{Z}_\boldsymbol{\xi}^{-1} &: L^2_\zeta(\Omega(\boldsymbol{\xi})) \rightarrow L^2(\tilde{\Omega}) & \text{as } f \mapsto f \circ Z^{-1}(\boldsymbol{\xi}), \end{aligned}$$

the latter equality allows us to properly delineate the solution manifold of the parametric problem defined in Eq. (1) in a conventional manner, namely,

$$\tilde{\mathcal{S}} := \{\mathcal{Z}_\boldsymbol{\xi}^{-1}(u(t, \boldsymbol{\mu}, \boldsymbol{\xi})) \in L^2(\tilde{\Omega}) \mid \boldsymbol{\mu} \in \mathcal{P}, t \in \mathcal{T}, \boldsymbol{\xi} \in \mathcal{G}\} \subset L^2(\tilde{\Omega}).$$

We refer the reader to Fig. 3 for a visualization of the action of the operator $\mathcal{Z}_\boldsymbol{\xi}^{-1}$ and to Appendix B for a more technical analysis of $\mathcal{Z}_\boldsymbol{\xi}$ and $\mathcal{Z}_\boldsymbol{\xi}^{-1}$, which are proven to be linear bounded operators. We also emphasize that $\mathcal{Z}_\boldsymbol{\xi}$ and its inverse enable us to morph back and forth from the original configuration $\Omega(\boldsymbol{\xi})$ onto the reference setting $\tilde{\Omega}$, namely,

$$\mathcal{S} = \{\mathcal{Z}_\boldsymbol{\xi}(\tilde{u}) \mid \tilde{u} \in \tilde{\mathcal{S}}\}, \quad \tilde{\mathcal{S}} = \{\mathcal{Z}_\boldsymbol{\xi}^{-1}(u) \mid u \in \mathcal{S}\}.$$

It is straightforward that the latter observation allows us to set the dimensionality reduction framework in the reference configuration, thus enabling us to address the problem from a classical point of view since $\tilde{\mathcal{S}}$ is a proper subset of $L^2(\tilde{\Omega})$ (see §3 and §4 for more details on the matter).

Thanks to the present functional setting, in the following we aim at providing a general framework to construct DL-ROMs in infinite dimensions specifically suited for problems featuring geometrical variability.

3. CGA-DL-ROMs in an infinite-dimensional setting

Within this section we provide the architecture design of the CGA-DL-ROM architecture. Specifically, we characterize its distinguishing traits, namely (i) the linear dimensionality reduction in an infinite-

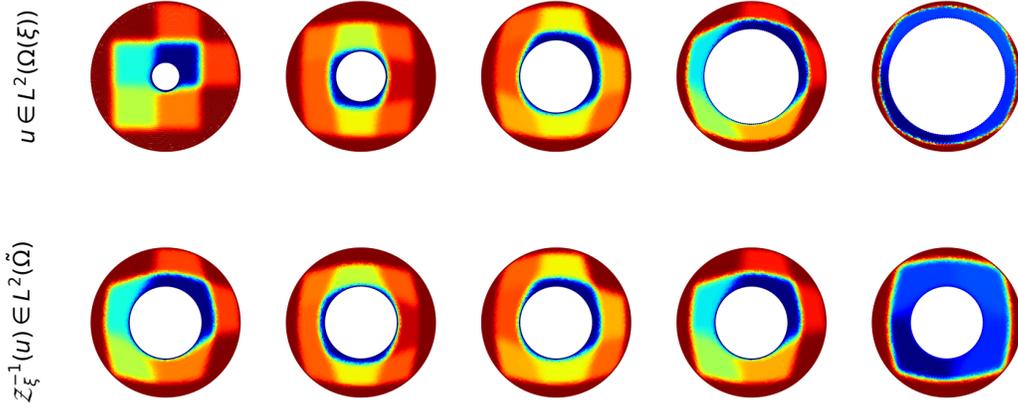


Figure 3. Solution of a physically and geometrically parametrized advection diffusion reaction equation on parametric domains (*first row*) and action of the morphing operator Z_{ξ}^{-1} (*second row*).

dimensional setting, (ii) the nonlinear autoencoder, and (iii) the reduced network. We refer to Fig. 4 for a schematic representation of the full architecture.

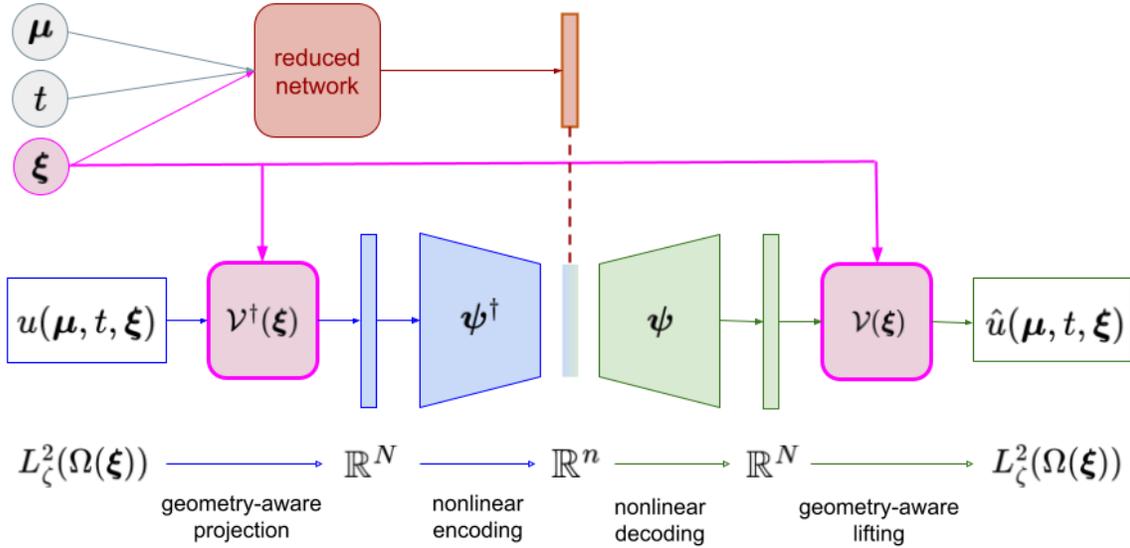


Figure 4. Schematic representation of the CGA-DL-ROM architecture.

CGA projection and lifting. The purpose of the outermost block of the architecture is to provide a finite-dimensional representation of the infinite-dimensional solution field through a linear projection. To do that, we employ the space-continuous projection operator $\mathcal{V}_{\xi}^{\dagger} : L_{\zeta}^2(\Omega(\xi)) \rightarrow \mathbb{R}^N$ and the space-continuous lifting operator $\mathcal{V}_{\xi} : \mathbb{R}^N \rightarrow L_{\zeta}^2(\Omega(\xi))$. Owing to §2.3, it is possible to decompose the action of each one of

the operators \mathcal{V}_ξ^\dagger and \mathcal{V}_ξ into two separate steps, namely,

$$\begin{aligned} \mathcal{V}_\xi^\dagger : L_\zeta^2(\Omega(\boldsymbol{\xi})) &\xrightarrow[\text{morphing}]{\mathcal{Z}_\xi^{-1}} L^2(\tilde{\Omega}) \xrightarrow[\text{projection}]{} \mathbb{R}^N \\ \mathcal{V}_\xi : \mathbb{R}^N &\xrightarrow[\text{lifting}]{} L^2(\tilde{\Omega}) \xrightarrow[\text{morphing}]{\mathcal{Z}_\xi} L_\zeta^2(\Omega(\boldsymbol{\xi})). \end{aligned}$$

Thus, it is evident that, in practice, the effective projection and lifting operations concern only the reference configuration $\tilde{\Omega}$. Thanks to this observation, we define $\{v_n^{CGA}(\boldsymbol{\xi})\}_{n=1}^N \subset L^2(\tilde{\Omega})$, a set global of basis functions that are defined on the reference configuration and that depend on $\boldsymbol{\xi}$, and provide a formal definition of

$$\begin{aligned} \mathcal{V}_\xi^\dagger(f) &:= \{(\mathcal{Z}_\xi^{-1}(f), v_n^{CGA}(\boldsymbol{\xi}))_{L^2(\tilde{\Omega})}\}_{n=1}^N, \quad \forall f \in L_\zeta^2(\Omega(\boldsymbol{\xi})) \\ \mathcal{V}_\xi(\mathbf{a}) &:= \mathcal{Z}_\xi\left(\sum_{n=1}^N a_n v_n^{CGA}(\boldsymbol{\xi})\right), \quad \forall \mathbf{a} = \{a_n\}_{n=1}^N \in \mathbb{R}^N. \end{aligned}$$

Then, owing to the change of variable formula of §2.3 and to the linearity of the morphing operator \mathcal{Z}_ξ , we can derive a more compact, alternative definition of space-continuous projection and lifting operators,

$$\begin{aligned} \mathcal{V}_\xi^\dagger(f) &:= \{(f, \mathcal{Z}_\xi(v_n^{CGA}(\boldsymbol{\xi}))\}_{L_\zeta^2(\Omega(\boldsymbol{\xi}))}\}_{n=1}^N, \quad \forall f \in L_\zeta^2(\Omega(\boldsymbol{\xi})) \\ \mathcal{V}_\xi(\mathbf{a}) &:= \sum_{n=1}^N a_n \mathcal{Z}_\xi(v_n^{CGA}(\boldsymbol{\xi})), \quad \forall \mathbf{a} = \{a_n\}_{n=1}^N \in \mathbb{R}^N. \end{aligned}$$

We emphasize that the compressive capabilities of the couple $(\mathcal{V}_\xi, \mathcal{V}_\xi^\dagger)$ strongly depends on how we characterize the dependence on the geometrical parameters of the global basis functions $\{v_n^{CGA}(\boldsymbol{\xi})\}_{n=1}^N$: we discuss about the latter aspect in §4.1. Nonetheless, since the collection $\{\mathcal{Z}_\xi(v_n^{CGA}(\boldsymbol{\xi}))\}_{n=1}^N$ is not in general available in a closed form, we remark that within this work we model it through neural networks of weights and biases θ^v and we refer the interested reader to §4.2 for a more comprehensive outlook on the matter.

Nonlinear autoencoder. The purpose of the autoencoder architecture is to further compress the representation entailed by $(\mathcal{V}_\xi^\dagger, \mathcal{V}_\xi)$ into a set of $l > 0$ latent coordinates. Thus, we configure the autoencoder architecture as the couple (ψ^\dagger, ψ) , where $\psi^\dagger : \mathbb{R}^N \rightarrow \mathbb{R}^l$ and $\psi : \mathbb{R}^l \rightarrow \mathbb{R}^N$ are such that $\psi^\dagger \circ \psi \approx Id$. Normally, we parametrize the autoencoder (ψ^\dagger, ψ) with (deep) neural networks and we indicate with θ^{AE} the entailed collection of weights and biases. We mention that in the literature, both dense and convolutional architectures have proven to be effective for this purpose [14, 15].

Reduced network. We design the reduced network as a feed-forward neural network architecture $\phi[\theta^\phi] : \mathbb{R}^{p+g+1} \rightarrow \mathbb{R}^l$, which is parametrized by its weights and biases θ^ϕ and strives to approximate the reduced parametric map $(t, \boldsymbol{\mu}, \boldsymbol{\xi}) \mapsto \psi^\dagger \circ \mathcal{V}_\xi^\dagger(u(t, \boldsymbol{\mu}, \boldsymbol{\xi}))$.

Thus, the resulting CGA-DL-ROM architecture is parametrized by $(\theta^v, \theta^{AE}, \theta^\phi)$, which are the weights and biases of the geometry-aware basis functions, the autoencoder, and the reduced network, respectively. We emphasize that the optimization phase of CGA-DL-ROMs consists in seeking the optimal set of neural network weights (and biases) with respect to the following *per-example* loss functional (for any $\boldsymbol{\mu} \in \mathcal{P}, \boldsymbol{\xi} \in \mathcal{G}, t \in \mathcal{T}$),

$$\begin{aligned} \mathcal{L}_{CGA}(t, \boldsymbol{\mu}, \boldsymbol{\xi}) &= \|\mathcal{V}_\xi \circ \psi \circ \phi(t, \boldsymbol{\mu}, \boldsymbol{\xi}) - u(t, \boldsymbol{\mu}, \boldsymbol{\xi})\|_{L_\zeta^2(\Omega(\boldsymbol{\xi}))}^2 + \\ &\quad + \|\phi(t, \boldsymbol{\mu}, \boldsymbol{\xi}) - \psi^\dagger \circ \mathcal{V}_\xi^\dagger(u(t, \boldsymbol{\mu}, \boldsymbol{\xi}))\|_2^2 \end{aligned} \quad (5)$$

which is composed of two different additive terms. The purpose of the first term is to control the squared reconstruction error; on the other hand, the second term strives to constrain the architecture to find a suitable latent representation. It is worth emphasizing that the CGA basis function are learnt in an unsupervised manner and that the encoder network is required only during the training phase and can be discarded at inference time. As a final remark, we mention also that in practice we are not able to compute exactly the integrals over $\Omega(\boldsymbol{\xi})$, so that we replace (5) with its empirical counterpart.

We stress that, while the nonlinear autoencoder and the reduced network have been analyzed in previous works [7, 11], the impact of the novel proposed CGA projection and lifting is still to be studied.

4. Characterization of CGA basis functions

In the following section we propose an extensive analysis of CGA projection and lifting, providing also a characterization of its compressive capabilities. Specifically, our purpose is to show that only a small number N of CGA basis functions is sufficient to capture the variability of the solution manifold (see §4.1). This ensures a light neural network architecture and aims at enhancing the overall approximation capabilities. Moreover, since such basis functions are not analytically available, we provide in §4.2 a suitable approximation through neural networks.

4.1. The compressive capabilities of CGA basis functions

Before presenting rigorously the CGA approach, in order to fully appreciate the compressive capabilities of CGA basis functions, we first describe the classical POD framework, which is the baseline we want to compare our approach with. First, we stress that, thanks to the results of the previous sections, we are allowed to delineate the dimensionality reduction problem on the reference configuration $\tilde{\Omega}$. Thus, we characterize the POD formalism through a result cast in a Hilbert context, relying on the theoretical setting reported in [35, 51], namely,

Lemma 4.1. *We define the POD cost functional as*

$$\mathcal{J}_{POD}(\{w_n\}_{n=1}^N) = \left\| \mathcal{Z}_{\xi}^{-1}(u) - \sum_{k=1}^N (\mathcal{Z}_{\xi}^{-1}(u), w_k)_{L^2(\tilde{\Omega})} w_k \right\|_{L^2(\mathcal{T} \times \mathcal{P} \times \mathcal{G}; L^2(\tilde{\Omega}))}^2,$$

where $W_N = \{\{w_n\}_{n=1}^N \in L^2(\tilde{\Omega}) : (w_n, w_m)_{L^2(\tilde{\Omega})} = \delta_{nm}, \forall n, m \in \{1, \dots, N\}\}$. Then,

$$\mathcal{E}_{POD}(N) = \min_{\{w_n\}_{n=1}^N \in W_N} \mathcal{J}_{POD}(\{w_n\}_{n=1}^N) = \mathcal{J}_{POD}(\{v_n^{POD}\}_{n=1}^N) = \sum_{n>N} l_n < +\infty, \quad (6)$$

where $(v_n^{POD}, l_n)_n \subset L^2(\tilde{\Omega}) \times \mathbb{R}$ are the POD eigenpairs collecting the orthonormal basis of singular vectors and the singular values of the compact linear operator $K = L^2(\mathcal{T} \times \mathcal{P} \times \mathcal{G}) \rightarrow L^2(\tilde{\Omega})$ defined as $Kg = (u(\mathbf{x}), g)_{L^2(\mathcal{T} \times \mathcal{P} \times \mathcal{G})}$, for any $g \in L^2(\mathcal{T} \times \mathcal{P} \times \mathcal{G})$.

More informally, the POD approach ultimately aims at providing the best set of basis functions *on average* with respect to the physical and geometrical parameters, according to the projection error metric. However, it is worth noticing that geometrical parameters usually cause a slow eigenvalue decay, that is, l_N exhibits a polynomial decay as $N \rightarrow \infty$ [13]. The latter scenario requires the user to employ a large number N of reduced basis functions to suitably reconstruct the variability of solution manifold, which is extremely inefficient from a computational viewpoint. In this respect, the CGA approach aims at mitigating the aforementioned inefficiency by ultimately providing, *for any instance of the geometrical parameter* $\xi \in \mathcal{G}$, a set of modal basis functions that is optimal for the chosen parameter ξ . More formally, we characterize the CGA optimization problem through

Lemma 4.2. *We define CGA cost functional as*

$$\mathcal{J}_{CGA}(\xi, \{w_n\}_{n=1}^N) := \left\| \mathcal{Z}_{\xi}^{-1}(u) - \sum_{n=1}^N (\mathcal{Z}_{\xi}^{-1}(u), w_n)_{L^2(\tilde{\Omega})} w_n \right\|_{L^2(\mathcal{T} \times \mathcal{P}; L^2(\tilde{\Omega}))}^2,$$

where $B = \{v \in L^2(\tilde{\Omega}) : \|v\|_{L^2(\tilde{\Omega})} = 1\}$. Then, the CGA optimization problem

$$\min_{\{w_n\}_{n=1}^N \in [B]^N} \mathcal{J}_{CGA}(\xi, \{w_n\}_{n=1}^N), \quad (7)$$

is well-defined, that is, there exists $\xi \mapsto \{v_n^{CGA}(\xi)\}_{n=1}^N$ which attains the minimum.

We refer the interested reader to Appendix C for the proof. Nonetheless, we highlight that, thanks to the functional setting provided in §2.3, it is straightforward to prove through the change of variable formula

the following equalities, which show that the optimization problem defined in Eqs. (6) and (7) can be equivalently cast in the original configuration $\Omega(\boldsymbol{\xi})$, namely,

$$\begin{aligned}\mathcal{J}_{POD}(\{w_n\}_{n=1}^N) &= \left\| u - \sum_{n=1}^N (u, \mathcal{Z}_{\boldsymbol{\xi}}(w_n))_{L^2(\tilde{\Omega})} \mathcal{Z}_{\boldsymbol{\xi}}(w_n) \right\|_{L^2(\mathcal{T} \times \mathcal{P} \times \mathcal{G}; L^2_{\zeta}(\Omega(\boldsymbol{\xi})))}^2 \\ \mathcal{J}_{CGA}(\boldsymbol{\xi}, \{w_n\}_{n=1}^N) &= \left\| u - \sum_{n=1}^N (u, \mathcal{Z}_{\boldsymbol{\xi}}(w_n))_{L^2(\tilde{\Omega})} \mathcal{Z}_{\boldsymbol{\xi}}(w_n) \right\|_{L^2(\mathcal{T} \times \mathcal{P}; L^2_{\zeta}(\Omega(\boldsymbol{\xi})))}^2.\end{aligned}$$

Then, we can prove the main result of the present section, which showcases the superiority of the CGA basis functions when compared to POD modes.

Proposition 4.1. *For any $N < \infty$, we define the POD and CGA best approximation errors as*

$$\begin{aligned}\text{BAE}_{CGA}(N) &:= \int_{\mathcal{G}} \mathcal{J}_{CGA}(\boldsymbol{\xi}, \{v_n^{CGA}\}_{n=1}^N) d\boldsymbol{\xi} \\ \text{BAE}_{POD}(N) &:= \mathcal{J}_{POD}(\{v_n^{POD}\}_{n=1}^N).\end{aligned}$$

Then, we can prove that

$$\text{BAE}_{CGA}(N) \leq \text{BAE}_{POD}(N).$$

Proof. We recall that $W_N = \{\{w_n\}_{n=1}^N \in L^2(\tilde{\Omega}) : (w_n, w_m)_{L^2(\tilde{\Omega})} = 1, \forall n, m \in \{1, \dots, N\}\}$ and that B is the unit ball in $L^2(\tilde{\Omega})$. Then, it is easy to see that $W_N \subset [B]^N$, so that $v_{POD} \in [B]^N$. Thus, by an optimality argument, for any $\boldsymbol{\xi} \in \mathcal{G}$ it is valid that

$$\mathcal{J}_{CGA}(\boldsymbol{\xi}, \{v_n^{CGA}\}_{n=1}^N) = \min_{\{w_n\}_{n=1}^N \in [B]^N} \mathcal{J}_{CGA}(\boldsymbol{\xi}, \{w_n\}_{n=1}^N) \leq \mathcal{J}_{CGA}(\boldsymbol{\xi}, \{v_n^{POD}\}_{n=1}^N).$$

Then, by integrating all the sides over \mathcal{G} , we derive the inequality

$$\int_{\mathcal{G}} \mathcal{J}_{CGA}(\boldsymbol{\xi}, \{v_n^{CGA}\}_{n=1}^N) d\boldsymbol{\xi} \leq \int_{\mathcal{G}} \mathcal{J}_{CGA}(\boldsymbol{\xi}, \{v_n^{POD}\}_{n=1}^N) d\boldsymbol{\xi}.$$

Moreover, by definition of $\text{BAE}_{CGA}(N)$, we can state that,

$$\text{BAE}_{CGA}(N) \leq \int_{\mathcal{G}} \mathcal{J}_{CGA}(\boldsymbol{\xi}, \{v_n^{POD}\}_{n=1}^N) d\boldsymbol{\xi}.$$

Finally, since for any $\{w_n\}_{n=1}^N \in [B]^N$ it is valid that

$$\int_{\mathcal{G}} \mathcal{J}_{CGA}(\boldsymbol{\xi}, \{w_n\}_{n=1}^N) d\boldsymbol{\xi} = \mathcal{J}_{POD}(\boldsymbol{\xi}, \{w_n\}_{n=1}^N),$$

we derive the thesis, namely,

$$\text{BAE}_{CGA}(N) \leq \int_{\mathcal{G}} \mathcal{J}_{CGA}(\boldsymbol{\xi}, \{v_n^{POD}\}_{n=1}^N) d\boldsymbol{\xi} = \mathcal{J}_{POD}(\boldsymbol{\xi}, \{v_n^{POD}\}_{n=1}^N) = \text{BAE}_{POD}(N).$$

□

The superiority of the CGA basis functions when compared to the POD modes becomes evident in the cases where the geometrical parametrization is the only cause of the slow (polynomial) POD eigenvalue decay [13], which entails that $\mathcal{E}_{POD}(N) = O(N^{-\alpha})$ for some $\alpha > 0$. On the other hand, if we consider the CGA approach involving *ad hoc* basis functions for each geometrical parameter instance, the latter scenario entails that $\mathcal{J}_{CGA}(\boldsymbol{\xi}, \{v_n^{CGA}\}_{n=1}^N) = O(e^{-\beta N})$ for some $\beta > 0$, independently of $\boldsymbol{\xi} \in \mathcal{G}$. Then, by definition, $\mathcal{E}_{CGA}(N) = O(e^{-\beta N})$, so that we can conclude that a small amount of CGA basis functions suffices to capture the solution manifold variability in this scenario, where a large amount of POD modes are necessary to obtain the same level of compression.

4.2. Modeling CGA basis functions with neural networks

Despite having proved better compressive capabilities when compared to POD modes, in practical applications CGA basis functions are not available in a closed form, that is, the minimum in (7) cannot be explicitly attained and needs to be suitably approximated. In this respect, within the present section we aim at providing a way to surrogating such CGA basis functions with neural networks. Specifically, the main existence result consists in proving that it is possible to construct a suitable neural network $(\mathbf{x}, \boldsymbol{\xi}) \mapsto \{\hat{v}_n^{CGA}(\mathbf{x}, \boldsymbol{\xi})\}_{n=1}^N \approx \{\mathcal{Z}_\xi(v_n^{CGA}(\mathbf{x}, \boldsymbol{\xi}))\}_{n=1}^N$ such that the associated CGA projection error is arbitrarily close to the minimum in (7). More formally,

Proposition 4.2. *We let $N < \infty$ and assume that $(\mathbf{x}, \boldsymbol{\xi}) \mapsto \{v_n^{CGA}(\mathbf{x}, \boldsymbol{\xi})\}_{n=1}^N \in W^{1,2}(\tilde{\Omega} \times \mathcal{G})$. Then, there exists $\varepsilon^* > 0$ such that for any $\varepsilon \in (0, \varepsilon^*)$, it is possible to design a neural network architecture $(\mathbf{x}, \boldsymbol{\xi}) \mapsto \{\hat{v}_n^{CGA}(\mathbf{x}, \boldsymbol{\xi})\}_{n=1}^N \approx \{\mathcal{Z}_\xi(v_n^{CGA}(\mathbf{x}, \boldsymbol{\xi}))\}_{n=1}^N$ such that*

$$\int_{\mathcal{G}} \mathcal{J}_{CGA}(\boldsymbol{\xi}, \{\mathcal{Z}_\xi^{-1}(\hat{v}_n^{CGA}(\boldsymbol{\xi}))\}_{n=1}^N) d\boldsymbol{\xi} < \text{BAE}_{CGA}(N) + \varepsilon.$$

We refer the interested reader to Appendix C for the proof. Anyway, we emphasize that $(\mathbf{x}, \boldsymbol{\xi}) \mapsto \{\hat{v}_n^{CGA}(\mathbf{x}, \boldsymbol{\xi})\}_{n=1}^N \approx \{\mathcal{Z}_\xi(v_n^{CGA}(\mathbf{x}, \boldsymbol{\xi}))\}_{n=1}^N$ approximates both the geometry-aware basis functions and the action of the morphing operator \mathcal{Z}_ξ , so that we do not require the diffeomorphism to be explicitly known in practical applications.

As a final remark, we highlight that we focus solely on the analysis of \mathcal{J}_{CGA} and, by extension, of BAE_{CGA} . Indeed, we neglect the contribution of the approximation capabilities of the reduced network and the decoder, and the effectiveness of the representation entailed by the encoder, whose analyses are beyond the purpose of the present work; for further details on this latter aspect, we refer the reader to [5, 7, 11, 26]. Notwithstanding, it is possible to show that BAE_{CGA} is a lower bound for the approximation error, that is, through an optimality argument we obtain,

$$\begin{aligned} \mathcal{E}_{approx} &:= \|u - \hat{u}\|_{L^2(\mathcal{T} \times \mathcal{P} \times \mathcal{G}; L^2_\zeta(\Omega(\boldsymbol{\xi})))} \\ &= \left\| u - \mathcal{V}_\xi \circ \psi \circ \phi(t, \boldsymbol{\mu}, \boldsymbol{\xi}) \right\|_{L^2(\mathcal{T} \times \mathcal{P} \times \mathcal{G}; L^2_\zeta(\Omega(\boldsymbol{\xi})))} \\ &= \left\| u - \sum_{n=1}^N (u, \mathcal{Z}_\xi(\hat{v}_n^{CGA}))_{L^2(\Omega(\boldsymbol{\xi}))} \mathcal{Z}_\xi(\hat{v}_n^{CGA}) \right\|_{L^2(\mathcal{T} \times \mathcal{P} \times \mathcal{G}; L^2_\zeta(\Omega(\boldsymbol{\xi})))} \\ &= \left[\int_{\mathcal{G}} \mathcal{J}_{CGA}(\boldsymbol{\xi}, \{\hat{v}_n^{CGA}\}_{n=1}^N)^2 d\boldsymbol{\xi} \right]^{1/2} \\ &\geq \left[\int_{\mathcal{G}} \min_{\{v_n^{CGA}\}_{n=1}^N \in [B]^N} \mathcal{J}_{CGA}(\boldsymbol{\xi}, \{v_n^{CGA}\}_{n=1}^N)^2 d\boldsymbol{\xi} \right]^{1/2} \\ &= \text{BAE}_{CGA}(N), \end{aligned}$$

where we recall ψ, ϕ are the decoder and the reduced network, respectively.

5. Numerical experiments

Within this section we direct our efforts toward the implementation of numerical experiments supporting our theoretical findings and showcase the potential of the CGA-DL-ROM technique. More precisely,

- (i) we motivate the usage of CGA basis functions by demonstrating their superiority in terms of compressive capability when compared to POD;
- (ii) we compare the CGA-DL-ROM with suitable baselines present in the literature;
- (iii) we test the accuracy of CGA-DL-ROM when tackling applications stemming from more-involved industrial benchmarks and challenges.

We remark that the accuracy of the predictions is measured by means of two relative error metrics, namely,

$$\mathcal{E}_R = \sum_{j=1}^{N_s^{test}} \frac{\left(\sum_{i=1}^{N_h(\xi_j)} (u(x_i; \boldsymbol{\mu}_j, t_j, \boldsymbol{\xi}_j) - \hat{u}(x_i; \boldsymbol{\mu}_j, t_j, \boldsymbol{\xi}_j))^2 \right)^{1/2}}{\left(\sum_{i=1}^{N_h(\xi_j)} u(x_i; \boldsymbol{\mu}_j, t_j, \boldsymbol{\xi}_j)^2 \right)^{1/2}},$$

$$E = \left(\frac{\sum_{j=1}^{N_s^{test}} \sum_{i=1}^{N_h(\xi_j)} (u(x_i; \boldsymbol{\mu}_j, t_j, \boldsymbol{\xi}_j) - \hat{u}(x_i; \boldsymbol{\mu}_j, t_j, \boldsymbol{\xi}_j))^2}{\sum_{j=1}^{N_s^{test}} \sum_{i=1}^{N_h(\xi_j)} u(x_i; \boldsymbol{\mu}_j, t_j, \boldsymbol{\xi}_j)^2} \right)^{1/2},$$

where \hat{u} is the approximation of the ground truth u . We stress that \mathcal{E}_R is the most widely used metric in the field literature, while E is a metric which is consistent with POD/CGA optimization problems. For this reason the latter is only employed to test the superiority of CGA basis functions with respect to POD modes. Unless otherwise specified, (i) our numerical experiments are performed on a NVIDIA A100 80GB GPU, (ii) we employ Adam as optimizer, and (iii) we use 80% of the generated data for training and the rest is split in equal parts for validation and testing.

		Stenosis (§5.1)	Hyper-elasticity (§5.2)	NS obstacle (§5.3)	Heat exchanger (§5.4)
ROM dimensions	latent dim. l	4	20	20	17
	red. dim. N	4	30	50	70
Architecture specifics	$\{\hat{v}_n^{CGA}\}_{n=1}^N$	RDense(120, 10)	RDense(180, 11)	RDense(220, 10)	RDense(150, 10)
	encoder ψ^\dagger	Dense(150, 5)	Dense(150, 5)	Conv2d(4)	Dense(50, 5)
	decoder ψ	Dense(150, 5)	Dense(150, 5)	Conv2dT(4)	Dense(150, 5)
	red. network ϕ	Dense(50, 5)	Dense(50, 5)	Dense(100, 5)	Dense(80, 5)
Computational burden	# NN weights	196k	397k	943k	298k
	Training time	0.3h	0.8h	3h	4h
	Inference time (per instance)	1ms	0.5ms	1ms	5ms
Dataset specifics	# samples N_s	1000	2000	100	100
	splitting	80%/10%/10%	50%/40%/10%	80%/10%/10%	80%/10%/10%
	# timesteps N_t	–	–	300	60
	# dofs N_h	7.8k	972	31k ÷ 34k	100k ÷ 110k

Table 1. Hyperparameters of best results obtained with CGA-DL-ROM in every experiment. We denote with Dense(w, d) the dense network of d layers with maximum width w and with RDense(w, d) the same architecture with residual connections. On the other hand, Conv2d(d) (Conv2dT(d)) indicates a convolutional block consisting of d convolutional (transposed convolutional) layers and two dense layers at input and output to correctly match input and output dimensions.

5.1. Stationary flow in a stenotic channel

We first consider a benchmark test case to demonstrate the superior compressive and generalization capabilities of the proposed CGA-DL-ROM architecture when compared its POD counterpart. In this respect, we immediately stress that if we substitute CGA basis functions with POD modes, we obtain the POD-DL-ROM architecture [15], which indeed is the baseline we would like to compare our novel approach with.

The present numerical experiment involves a fluid flowing in a 2D blood vessel over a stenosis and is usually employed as a simplified model for studying coronary artery diseases, which are of great interest in

haemodynamics [50]. Specifically, we revisit numerical experiment 6.4 of [39], namely,

$$\begin{cases} -\nu\Delta\mathbf{u} + (\mathbf{u} \cdot \nabla)\mathbf{u} + \nabla p = \mathbf{0}, & \text{in } \Omega(\boldsymbol{\xi}) \\ \nabla \cdot \mathbf{u} = 0, & \text{in } \Omega \\ \mathbf{u} = [Ay(1-y), 0, 0], & \text{on } \Gamma_{IN} \\ \mathbf{u} = \mathbf{0}, & \text{on } \Gamma_W(\boldsymbol{\xi}) \\ -p\mathbf{n} + \nu(\nabla\mathbf{u})\mathbf{n} = \mathbf{0}, & \text{on } \Gamma_{OUT}, \end{cases} \quad (8)$$

where Γ_{IN} and Γ_{OUT} are the inlet and outlet of the blood vessel, whereas $\Gamma_W(\boldsymbol{\xi}) = \partial\Omega(\boldsymbol{\xi}) \setminus (\Gamma_{IN} \cup \Gamma_{OUT})$. Moreover, we emphasize that the problem is physically parametrized by the fluid viscosity $\nu \in [0.002, 0.004]$ and the inlet magnitude $A \in [2, 4]$. We parametrize the geometry of the artery's lower wall by means of the function

$$w(y; \boldsymbol{\xi}) = \begin{cases} \xi_1 \cos\left(\frac{\pi}{2\xi_2}(y - \xi_3)\right), & x \in (\xi_3 - \xi_2, \xi_3 + \xi_2) \\ 0, & \text{otherwise,} \end{cases}$$

where $\boldsymbol{\xi} = (\xi_1, \xi_2, \xi_3) \in [0.25, 0.4] \times [0.5, 0.75] \times [2, 3]$ collects the geometrical parameters which regulate the stenosis severity and position (see Fig. 5 for more details).

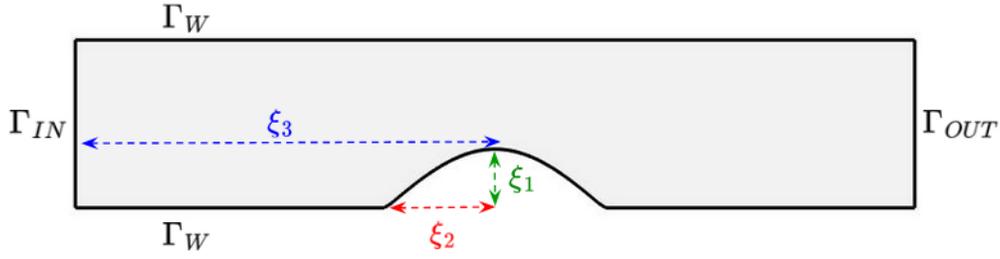


Figure 5. Stenosis test case: visualization of the effect of the geometrical parameters on the domain shape.

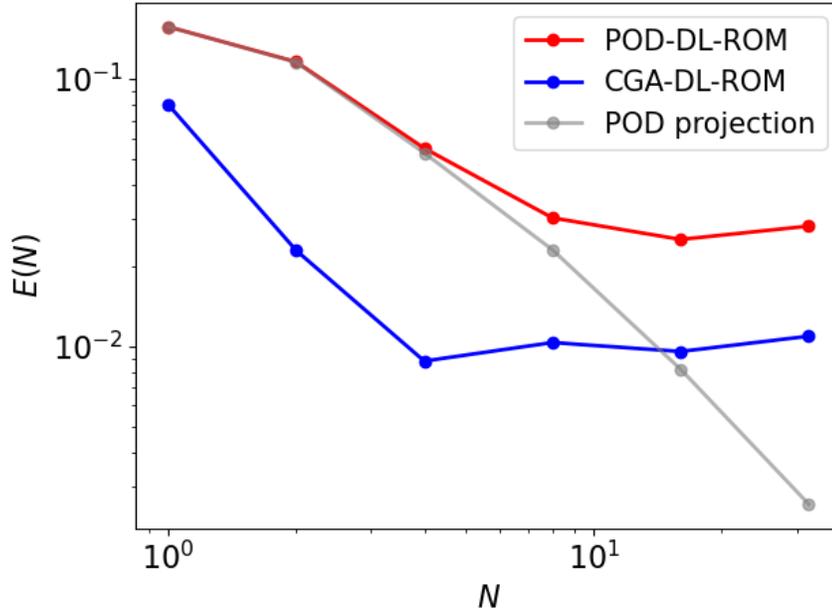


Figure 6. Stenosis test case: analysis of the compressive capabilities of CGA-DL-ROM and POD-DL-ROM. We also show the POD projection error for reference.

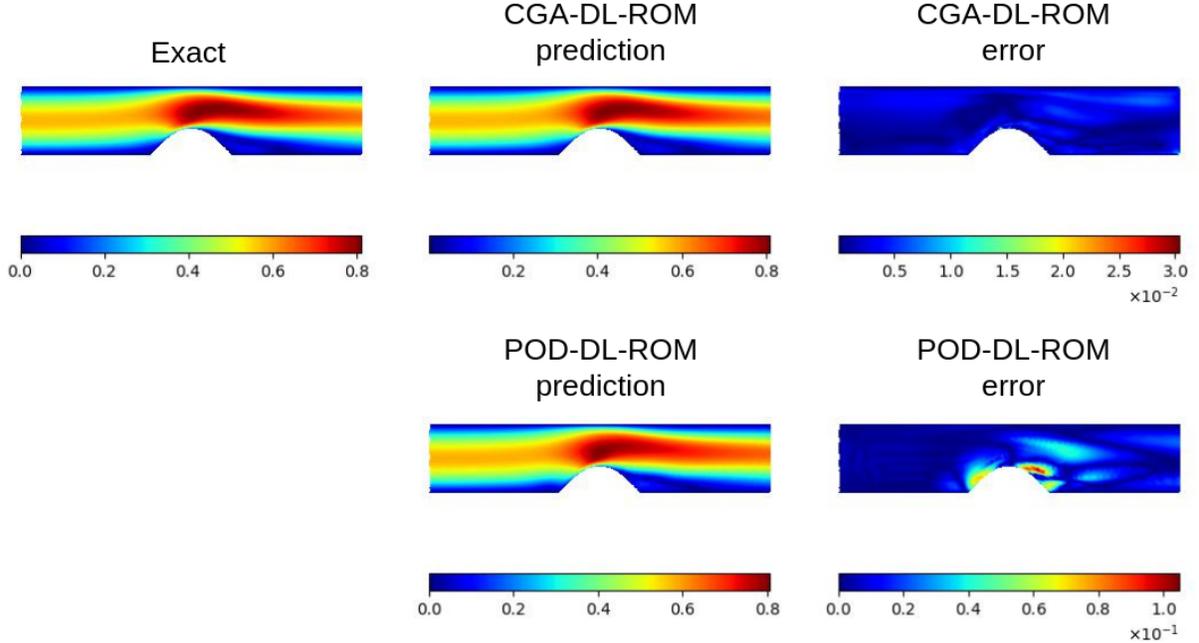


Figure 7. Stenosis test case: comparison of CGA-DL-ROM and POD-DL-ROM predictions and errors in the case of $N = 4$.

We generate a set of $N_s = 1000$ samples collecting only the x -velocity and y -velocity components, which are the fields that we would like to reconstruct. To do that, we make use of FEniCS [24] and gmsh [16], employing a mixed $\mathbb{P}^1\text{b} - \mathbb{P}^1$ finite element space and a Newton solver to deal with the nonlinearity. Generating all the samples takes 35 minutes on a Intel Core i7 13th gen 32GB RAM laptop.

We remark that, for the sake of simplicity, we make use of a *fixed-resolution* approach and a RBF strategy to morph the meshes, thus entailing a total of $N_h = 7600$ dofs for each sample. The latter simplification facilitates the comparison of CGA-DL-ROM with its POD counterpart because it is possible to pre-compute POD basis functions by means of SVD. Specifically, we train for 500 epochs a series of CGA-DL-ROMs and POD-DL-ROMs to reconstruct both the velocity components of for varying reduced dimension $N \in \{2^m : m = 0, \dots, 6\}$ and evaluate their accuracy on the test set. Also, we set the latent dimension $l = \min(N, 10)$. From the analysis' results displayed in in Fig. 6, it is evident that CGA-DL-ROMs are more accurate than POD-DL-ROMs for any value of N : this behaviour is particularly evident for lower values of N , thus corroborating our theoretical derivations. We also point out that $E(N)$ reaches a plateau for both the analyzed paradigms. However, CGA-DL-ROM saturates earlier ($N \geq 4$) than POD-DL-ROM ($N \geq 16$) and the plateau value of CGA-DL-ROM is $3\times$ smaller than the one relative to the POD counterpart. Finally, we emphasize that a only a small number of CGA basis functions ($N = 4$) is required to obtain a suitable accuracy. On the other hand, the same number of POD modes is not capable to suitably capture the solution manifold's variability. Indeed, POD-DL-ROM struggles especially in the proximity of the stenosis where the impact of the geometrical parameters is more evident (see Fig. 7).

5.2. Hyper-elasticity equations

Within the present numerical experiment we aim at comparing the proposed approach with baselines present in the literature. More precisely, we focus on the numerical experiment proposed in [29] that involves an hyper-elastic material described by the solid body equation

$$\frac{\partial^2 \mathbf{u}}{\partial t^2} + \nabla \cdot \boldsymbol{\sigma} = \mathbf{0}$$

where \mathbf{u} and $\boldsymbol{\sigma}$ are the displacement and the stress field, respectively, and the equation is endowed with a Rivlin-Saunders constitutive model. The body shape consists of a unit square with a parametric void at

the center; the void is centered in $(0.5, 0.5)$ and is characterized by the radius prior $r = 0.2 + \frac{0.2}{1 + \exp(r^*)}$ (where $r^* \sim \mathcal{N}(0, 4^2(-\nabla + 3^3)^{-1})$), yielding a total of 42 geometrical parameters per sample. The body is subject to the tensile traction $[0, 100]$ imposed at the top edge $y = 1$. Within the present test case we aim at reconstructing the stress given an input geometry consisting of a point cloud.

	Relative error	Latent space dimension	# NN weights and biases (in millions)
CGA-DL-ROM	1.24×10^{-2}	20	0.40
Geo-FNO [29]	2.20×10^{-2}	4608	1.55
GraphNO [29]	1.26×10^{-1}	–	<u>0.57</u>
DeepONet [29]	9.65×10^{-2}	<u>256</u>	1.03
DAFNO [34]	1.09×10^{-2}	4608	2.37
Coral [48]	1.64×10^{-2}	<u>128</u>	<u>0.53</u>
F-FNO [52]	1.74×10^{-2}	4608	8.87
GNOT [18]	8.65×10^{-3}	–	> 2

Table 2. Elasticity test case: comparison against baselines (less is better: *first*, *second*, *third*). The latent space dimension of FNO-based models is $l = \#\text{modes}^d \times \text{width}$, where d is the number of spatial dimensions. The latent space of DeepONet is its number of basis functions. The reported results are taken from the associated paper and/or code.

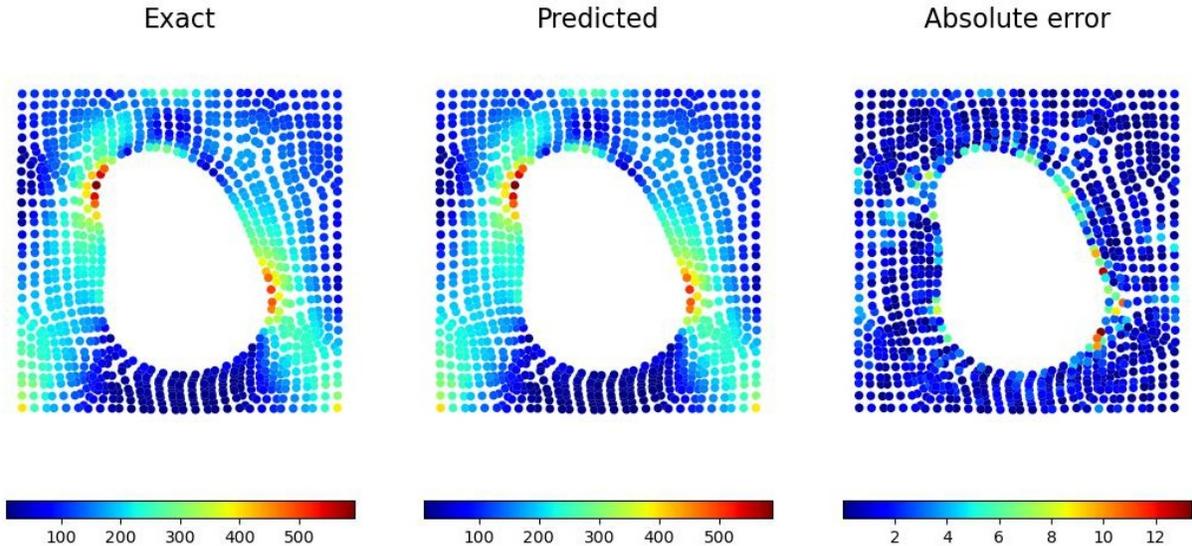


Figure 8. Hyper-elasticity test case: CGA-DL-ROM prediction and error for a test sample.

We train CGA-DL-ROM on 1000 data samples for 3000 epochs with a batch size of 8 and using a learning rate of 3×10^{-4} . As shown in Table 2, CGA-DL-ROMs are extremely efficient surrogate models. Indeed, they showcase an excellent accuracy on the test set (composed of 200 samples), almost on par with state-of-the-art paradigms, while having a smaller architecture (in terms of number of parameters) consisting of a remarkably low number of CGA basis functions ($N = 30$) and an even lower latent dimension ($l = 20$). On the other hand, transformer architectures deliver the best results in terms of accuracy, at the expense of heavy models. The only paradigm entailing a number of neural network parameters which is comparable with CGA-DL-ROM is Coral, but it features a worse accuracy and requires an inference-time optimization [48]. Finally, we further validate qualitatively the outstanding approximation and generalization capabilities of CGA-DL-ROM thanks to Fig. 8.

5.3. Navier-Stokes flow around an obstacle

The present numerical experiment involves the reconstruction of the velocity field of a fluid flowing in a 2D rectangular channel $[0, 3] \times [0, 0.5]$ past a circular obstacle. The domain shape is geometrically parameterized by means of $\boldsymbol{\xi} = \{y_c, r_c\}$, where $y_c \in [0.15, 0.35]$ and $r_c \in [0.05, 0.1]$ are respectively the y -position of the center of the obstacle and the radius of the obstacle. The inlet boundary is $\Gamma_{IN} = \{(x, y) \in \partial\Omega(\boldsymbol{\xi}) : x = 0\}$, while the outlet boundary is $\Gamma_{OUT} = \{(x, y) \in \partial\Omega(\boldsymbol{\xi}) : x = 3\}$. The problem formulation reads

$$\begin{cases} \frac{\partial \mathbf{u}}{\partial t} - \nu \Delta \mathbf{u} + (\mathbf{u} \cdot \nabla) \mathbf{u} + \nabla p = \mathbf{0}, & \text{in } \Omega(\boldsymbol{\xi}) \times (0, T] \\ \nabla \cdot \mathbf{u} = 0, & \text{in } \Omega(\boldsymbol{\xi}) \times (0, T] \\ \mathbf{u} = [h(y; A), 0], & \text{on } \Gamma_{IN}(\boldsymbol{\xi}) \times (0, T] \\ \mathbf{u} = \mathbf{0}, & \text{on } \partial\Omega(\boldsymbol{\xi}) \setminus (\Gamma_{IN} \cup \Gamma_{OUT}) \times (0, T] \\ -p\mathbf{n} + \nu(\nabla \mathbf{u})\mathbf{n} = \mathbf{0}, & \text{on } \Gamma_{OUT}(\boldsymbol{\xi}) \times (0, T] \\ \mathbf{u}_0 = \mathbf{u}_{stokes}, & \text{in } \Omega(\boldsymbol{\xi}) \end{cases} \quad (9)$$

where $T = 0.8$ is the time horizon, \mathbf{u}_{stokes} is the steady Stokes simulation on the same configuration, $\nu \in [1 \times 10^{-3}, 2 \times 10^{-3}]$ is the fluid viscosity and $h(y; A) = 4A(0.5 - y)y$ is the inlet profile that is regulated by the maximum amplitude $A \in [1.5^2, 5 \cdot 1.5^2]$. For the sake of clarity, we represent the physical parameters with the vector $\boldsymbol{\mu} = [A, \nu] \in \mathbb{R}^2$. We highlight that with the present choice of the physical parameters, since $Re \approx 550 \div 5625$, we witness vortex shedding phenomena for any of the sample instances, especially as t approaches the time horizon. The high-fidelity simulations are obtained with FEniCS [24] and gmsh [16] by discretizing the problem by means of $\mathbb{P}^2 - \mathbb{P}^1$ couple of finite element spaces and by employing the Chorin-Temam splitting scheme, yielding a total of more than 3×10^4 dofs. For the time-advancing scheme, we chose $\Delta t = T/3000$ to ensure the stability of the numerical solution, but we save the solution snapshots only each 10 iterations in order to limit the memory footprint of the dataset, while retaining the essential features of the temporal evolution of the solution fields. The elapsed time to generate the high-fidelity dataset is 23h on a Intel Core i7 13th gen 32GB RAM laptop.

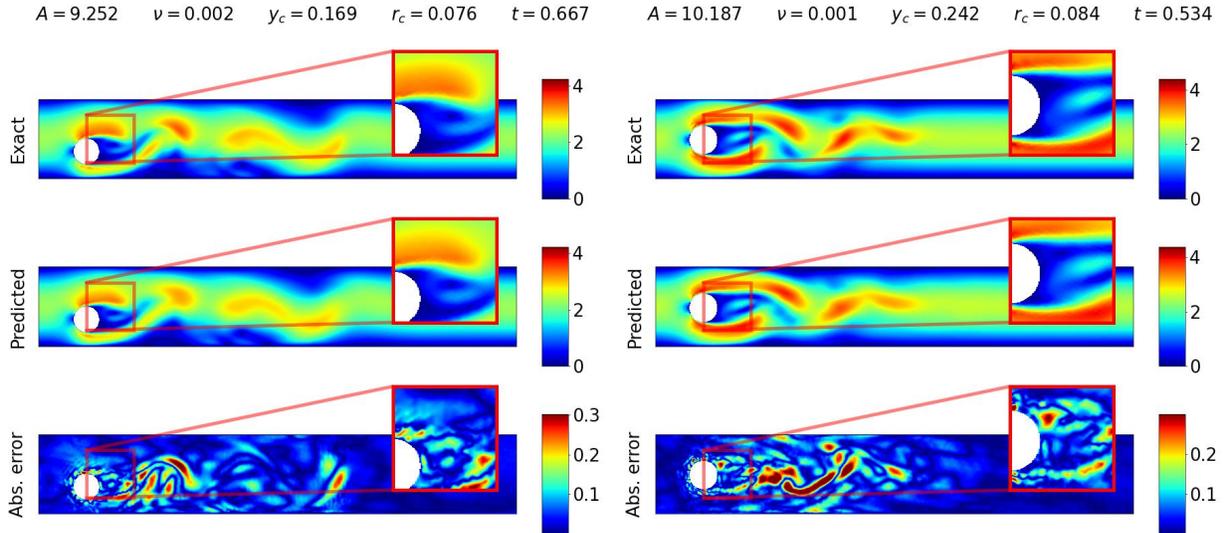


Figure 9. NS obstacle test case: accuracy of CGA-DL-ROM’s prediction for two different test instances. By comparing the two instances, it is evident the large variability of the solution manifold entailed by the physical and geometrical parametrization.

We train a CGA-DL-ROM for 65 epochs, using a batch size of 15. We evaluate the accuracy of the trained network on a test set consisting of unseen geometries, obtaining $\mathcal{E}_R = 4.99 \times 10^{-2}$, which suggests a precise reconstruction of the parameter-to-solution map despite the fact that only 80 different geometries are seen during the training phase. We also report in in Fig. 9 a visualization of the ROM prediction, highlighting

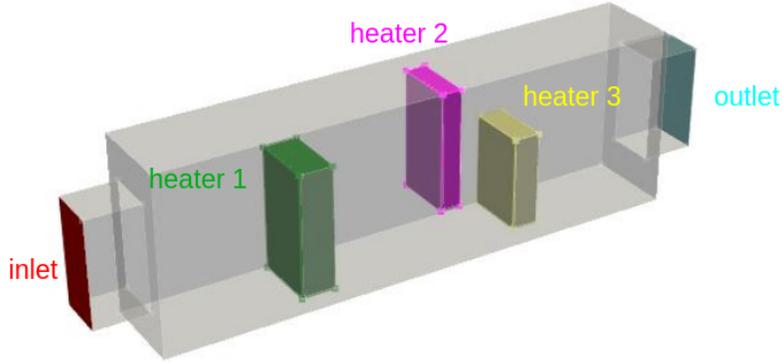


Figure 10. Heat exchanger test case: domain shape corresponding to an instance of the geometrical parameters.

that CGA-DL-ROM are able to suitably reconstruct even the low scale features occurring in the proximity of the obstacle’s wake, which is a region of the domain that is heavily influenced by the geometrical parameters. We stress that Fig. 9 shows the prediction of two different test instances so that the reader can appreciate the accurate reconstruction of the solution manifold despite the extremely high variability entailed by the physical and geometrical parametrization.

5.4. Fluid temperature in a heat exchanger

The purpose of this numerical experiment is to test the proposed framework with a relevant real-world application involving a large-scale 3D setting. More precisely, we are interested in describing the temperature field of a fluid flowing through the heat exchanger. We remark that the present test case extends the numerical experiment proposed in §5 of [7] by introducing a geometrical parameterization.

The heat exchanger shape $\Omega = \Omega(\boldsymbol{\xi})$ is parametrized by a set of geometrical parameters $\boldsymbol{\xi} = \{\xi_k\}_{k=1}^{14}$ (for details, we refer the reader to Table 3). The boundary of the heat exchanger is characterized by its three baffles $\Gamma_i(\boldsymbol{\xi})$ (for $i = 1, 2, 3$), the inlet $\Gamma_{IN}(\boldsymbol{\xi})$, the outlet $\Gamma_{OUT}(\boldsymbol{\xi})$ and the wall $\Gamma_W(\boldsymbol{\xi}) = \partial\Omega(\boldsymbol{\xi}) \setminus (\cup_{j=1}^3 \Gamma_j(\boldsymbol{\xi}) \cup \Gamma_{IN}(\boldsymbol{\xi}) \cup \Gamma_{OUT}(\boldsymbol{\xi}))$. A sample

Parameter	Description	Values
ξ_1	x -position of the first heater	[0.5, 0.9]
ξ_2	length of the first heater	[0.1, 0.2]
ξ_3	width of the first heater	[0.6, 0.7]
ξ_4	height of the first heater	[0.3, 0.4]
ξ_5	x -position of the second heater	[1.3, 1.7]
ξ_6	length of the second heater	[0.1, 0.2]
ξ_7	width of the second heater	[0.6, 0.7]
ξ_8	height of the first heater	[0.3, 0.4]
ξ_9	x -position of the third heater	[2.1, 2.5]
ξ_{10}	length of the third heater	[0.1, 0.2]
ξ_{11}	width of the third heater	[0.45, 0.55]
ξ_{12}	height of the third heater	[0.25, 0.35]
ξ_{13}	width of the inlet and the outlet	[0.4, 0.6]
ξ_{14}	height of the inlet and the outlet	[0.25, 0.35]

Table 3. Heat exchanger test case: details on the geometrical parametrization

To generate the high-fidelity dataset, we model the temperature u of the fluid with an advection-diffusion

equation, namely,

$$\begin{cases} \frac{\partial u}{\partial t} - D\Delta u + \mathbf{v} \cdot \nabla u = 0, & \text{in } \Omega(\boldsymbol{\xi}) \times (0, T] \\ u = \sum_{j=1}^3 g_j 1_{\Gamma_j(\boldsymbol{\xi})}, & \text{on } \cup_{j=1}^3 \Gamma_j(\boldsymbol{\xi}) \times (0, T] \\ u = 0, & \text{on } (\Gamma_{IN}(\boldsymbol{\xi}) \cup \Gamma_W(\boldsymbol{\xi})) \times (0, T] \\ \nabla u \cdot \mathbf{n} = 0, & \text{on } \Gamma_{OUT}(\boldsymbol{\xi}) \times (0, T] \\ u_0 = 0, & \text{in } \Omega(\boldsymbol{\xi}), \end{cases} \quad (10)$$

where $T = 2$ is the time horizon, $D \in [0.01, 0.1]$ is the thermal diffusivity of the fluid, while $g_1, g_2, g_3 \in [1, 11]$ are the imposed temperatures on the corresponding baffle. We remark that \mathbf{v} is the advection field due to the fluid flow, which is described by a stationary Navier-Stokes model,

$$\begin{cases} -\nu\Delta \mathbf{v} + (\mathbf{v} \cdot \nabla)\mathbf{v} + \nabla p = \mathbf{0}, & \text{in } \Omega(\boldsymbol{\xi}) \\ \nabla \cdot \mathbf{v} = 0, & \text{in } \Omega \\ \mathbf{v} = [h(y, z; A), 0, 0], & \text{on } \Gamma_{IN}(\boldsymbol{\xi}) \\ \mathbf{v} = \mathbf{0}, & \text{on } \cup_{j=1}^3 \Gamma_j \cup \Gamma_W(\boldsymbol{\xi}) \\ -p\mathbf{n} + \nu(\nabla \mathbf{v})\mathbf{n} = \mathbf{0}, & \text{on } \Gamma_{OUT}(\boldsymbol{\xi}), \end{cases} \quad (11)$$

where ν is the viscosity of the fluid and

$$h(y, z; A) = 0.15^{-2} \cdot 16A(0.75 - y)(y - 0.25)(0.4 - z)(z - 0.1).$$

represents the inlet profile regulated by the maximum amplitude $A \in [1, 2]$. For the sake of readability, we collect the physical parameters in the vector $\boldsymbol{\mu} = [A, \nu, g_1, g_2, g_3, D] \in \mathbb{R}^6$. We employ a three-step scheme for the computation of the high-fidelity solution, namely, for any parameter instance *(i)* we solve the Navier Stokes equations with $\mathbb{P}^1\mathbf{b} - \mathbb{P}^1$ couple of Finite Element spaces, handling the nonlinearity with a Newton solver; we remark that we choose the Stokes solution on the same configuration as the initial guess of the nonlinear solver, *(ii)* we interpolate the velocity field onto a \mathbb{P}^2 space on the same mesh, and *(iii)* we solve the advection-diffusion problem with \mathbb{P}^2 elements. We emphasize that the entire synthetic data generation phase is implemented in `python`, using `FEniCS` [24] and `gmsh` [16], and takes 5h on a Intel Core i7 13th gen 32GB RAM laptop.

We train the architecture for a total of 100 epochs with a learning rate of 10^{-3} and a batch size of 5. Then, we test the trained CGA-DL-ROM on a test dataset consisting of unseen geometries, thus obtaining an accuracy of $\mathcal{E}_R = 5.20 \times 10^{-2}$. The qualitative results displayed in Fig. 11 further validate the approximation and generalization capabilities of CGA-DL-ROMs: indeed the proposed architecture is able to suitably capture the moving fronts of the temperature due to the imposed advection field. Nonetheless, we emphasize that the variability of the solution manifold is correctly reproduced as CGA-DL-ROM is capable of grasping the strong effect of the geometrical parametrization and the diffusive/advective behaviour.

Conclusion

Within this work, we design and analyze the CGA-DL-ROM architecture to tackle problems featuring geometrical variability and parametrized domain shapes. Aside from the already analyzed DL-ROM core, consisting of a nonlinear autoencoder and a reduced network, the main novelties of CGA-DL-ROM consists in their characteristic CGA basis functions that make the proposed method

- *space-continuous*: we cast CGA-DL-ROM in an infinite-dimensional framework, allowing for *multi-resolution* datasets, which are quite common for problems featuring geometrical variability. Moreover, we complement the proposed framework with an abstract analysis of geometrically parametrized problems and a characterization of the main properties, including the well-posedness of the dimensionality reduction problem;

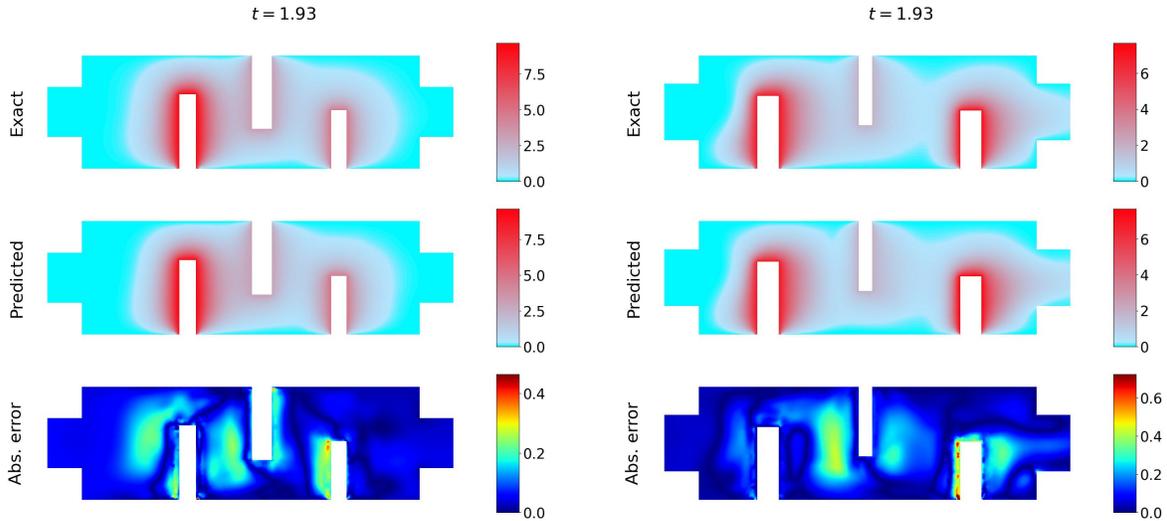


Figure 11. Heat exchanger test case: comparison between CGA-DL-ROM’s prediction and ground truth on the plane $z = 0.25$ at $t = 1.93s$. We emphasize the notable effects of the geometrical parametrization on the domain shape.

- *geometry-aware*: we inform such basis functions with the underlying geometrical parameters, thus creating a suitable inductive bias that strengthens the compressive capabilities, thus allowing domain practitioners to design light architectures with competitive prediction accuracy.

We showcase the versatility of the proposed framework through a series of numerical tests encompassing a diverse range of geometrically parametrized problems featuring generic *multi-resolution* datasets, thus validating CGA-DL-ROMs’ remarkable approximation and generalization capabilities. Then, encouraged by the promising results of this novel framework on industrial benchmarks (cf. NS obstacle and Heat Exchanger test cases), in the future we aim at testing our proposed framework with large scale real-world applications. Nonetheless, another line of research may entail the analysis and the design of an extension of CGA-DL-ROM to deal with non-parametric geometrical variability, where the domain shapes are parametrized by a set of unknown (possibly infinite-dimensional) geometrical parameters. Finally, we remark that the characterization of CGA basis functions is not limited to DL-ROMs and can be further extended to different architectures: this topic may constitute a promising direction for future research.

Acknowledgements

The authors are members of the Gruppo Nazionale Calcolo Scientifico-Istituto Nazionale di Alta Matematica (GNCS-INdAM) and acknowledge the project “Dipartimento di Eccellenza” 2023-2027, funded by MUR. SF acknowledges the Istituto Nazionale di Alta Matematica (INdAM) for the financial support received (Scholarships for visiting periods abroad, 2023-2024). SB acknowledges the support of European Union - NextGenerationEU within the Italian PNRR program (M4C2, Investment 3.3) for the PhD Scholarship “Physics-informed data augmentation for machine learning applications”. AM acknowledges the PRIN 2022 Project “Numerical approximation of uncertainty quantification problems for PDEs by multi-fidelity methods (UQ-FLY)” (No. 202222PACR), funded by the European Union - NextGenerationEU, and the Project “Reduced Order Modeling and Deep Learning for the real-time approximation of PDEs (DREAM)” (Starting Grant No. FIS00003154), funded by the Italian Science Fund (FIS) - Ministero dell’Università e della Ricerca. AM and SF acknowledge the project FAIR (Future Artificial Intelligence Research), funded by the NextGenerationEU program within the PNRR-PE-AI scheme (M4C2, Investment 1.3, Line on Artificial Intelligence). The authors thank Nicola R. Franco for the insightful discussions about parametric low-rank approximations.

References

- [1] Francesco Ballarin, Elena Faggiano, Sonia Ippolito, Andrea Manzoni, Alfio Quarteroni, Gianluigi Rozza, and Roberto Scrofani. Fast simulations of patient-specific haemodynamics of coronary artery bypass grafts based on a pod–galerkin method and a vascular shape parametrization. *Journal of Computational Physics*, 315:609–628, 2016.
- [2] Peter Benner, Stefano Grivet-Talocia, Alfio Quarteroni, Gianluigi Rozza, Wil Schilders, and Luís Miguel (Eds.) Silveira. *Model Order Reduction, Volume 1: System-and Data-Driven Methods and Algorithms*. De Gruyter, Berlin/Boston, 2021.
- [3] Peter Benner, Mario Ohlberger, Albert Cohen, and Karen (Eds.) Willcox. *Model reduction and approximation: theory and algorithms*. SIAM, Philadelphia, PA, 2017.
- [4] Peter Benner, Wil Schilders, Stefano Grivet-Talocia, Alfio Quarteroni, Gianluigi Rozza, and Luís Miguel Silveira (Eds.). *Model Order Reduction, Volume 2: Snapshot-Based Methods and Algorithms*. De Gruyter, Berlin/Boston, 2020.
- [5] Kaushik Bhattacharya, Bamdad Hosseini, Nikola B Kovachki, and Andrew M Stuart. Model reduction and neural networks for parametric pdes. *The SMAI journal of computational mathematics*, 7:121–157, 2021.
- [6] Haim Brezis. *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. Universitext. Springer New York, 2010.
- [7] Simone Brivio, Stefania Fresca, Nicola Rares Franco, and Andrea Manzoni. Error estimates for POD-DL-ROMs: a deep learning framework for reduced order modeling of nonlinear parametrized PDEs enhanced by proper orthogonal decomposition. *Adv. Comput. Math.*, 50(33), 2024.
- [8] Shengze Cai, Zhicheng Wang, Lu Lu, Tamer A. Zaki, and George Em Karniadakis. Deepm&mnnet: Inferring the electroconvection multiphysics fields based on operator approximation by neural networks. *Journal of Computational Physics*, 436:110296, July 2021.
- [9] Shuhao Cao. Choose a transformer: Fourier or galerkin. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, volume 34, pages 24924–24940. Curran Associates, Inc., 2021.
- [10] Fabien Casenave, Brian Staber, and Xavier Roynard. Mmgp: a mesh morphing gaussian process-based machine learning method for regression of physical problems under non-parameterized geometrical variability, 2023.
- [11] Nicola R. Franco, Andrea Manzoni, and Paolo Zunino. A deep learning approach to reduced order modelling of parameter dependent partial differential equations. *Math. Comp.*, 92(340):483–524, 2023.
- [12] Nicola Rares Franco, Stefania Fresca, Filippo Tombari, and Andrea Manzoni. Deep learning-based surrogate models for parametrized PDEs: Handling geometric variability through graph neural networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 33(12):123121, 12 2023.
- [13] Nicola Rares Franco, Andrea Manzoni, Paolo Zunino, and Jan S. Hesthaven. Deep orthogonal decomposition: a continuously adaptive data-driven approach to model order reduction, 2024.
- [14] Stefania Fresca, Luca Dede’, and Andrea Manzoni. A comprehensive deep learning-based approach to reduced order modeling of nonlinear time-dependent parametrized pdes. *Journal of Scientific Computing*, 87:61, 2021.
- [15] Stefania Fresca and Andrea Manzoni. Pod-dl-rom: Enhancing deep learning-based reduced order models for nonlinear parametrized pdes by proper orthogonal decomposition. *Computer Methods in Applied Mechanics and Engineering*, 388:114181, 2022.

- [16] Christophe Geuzaine and Jean-François Remacle. Gmsh: A 3-d finite element mesh generator with built-in pre- and post-processing facilities. *International Journal for Numerical Methods in Engineering*, 79(11):1309–1331, 2009.
- [17] Ingo Gühring and Mones Raslan. Approximation rates for neural networks with encodable weights in smoothness spaces. *Neural Networks*, 134:107–130, 2021.
- [18] Zhongkai Hao, Zhengyi Wang, Hang Su, Chengyang Ying, Yinpeng Dong, Songming Liu, Ze Cheng, Jian Song, and Jun Zhu. Gnot: A general neural operator transformer for operator learning. In *Proceedings of the 40th International Conference on Machine Learning, ICML’23*. JMLR.org, 2023.
- [19] AmirPouya Hemmasian and Amir Barati Farimani. Multi-scale time-stepping of partial differential equations with transformers. *Computer Methods in Applied Mechanics and Engineering*, 426:116983, 2024.
- [20] Pengzhan Jin, Shuai Meng, and Lu Lu. Mionet: Learning multiple-input operators via tensor product, 2022.
- [21] Efthymios N. Karatzas, Giovanni Stabile, Francesco Ballarin, and Gianluigi Rozza. *Chapter 13: Reduced Basis, Embedded Methods, and Parametrized Level-Set Geometry*, page 265–282. Society for Industrial and Applied Mathematics, January 2022.
- [22] Georgios Kissas, Jacob H. Seidman, Leonardo Ferreira Guilhoto, Victor M. Preciado, George J. Pappas, and Paris Perdikaris. Learning operators with coupled attention. *Journal of Machine Learning Research*, 23(215):1–63, 2022.
- [23] Nikola B Kovachki, Zongyi Li, Burigede Liu, Kamyar Azizzadenesheli, Kaushik Bhattacharya, Andrew M Stuart, and Anima Anandkumar. Neural operator: Learning maps between function spaces with applications to pdes. *Journal of Machine Learning*, 24(89), 2023.
- [24] Hans Petter Langtangen and Anders Logg. *Solving PDEs in Python*. Springer, 2017.
- [25] Samuel Lanthaler, Zongyi Li, and Andrew M. Stuart. The nonlocal neural operator: Universal approximation, 2023.
- [26] Samuel Lanthaler, Siddhartha Mishra, and George E Karniadakis. Error estimates for DeepONets: a deep learning framework in infinite dimensions. *Transactions of Mathematics and Its Applications*, 6(1):tnac001, 03 2022.
- [27] Toni Lassila and Gianluigi Rozza. Parametric free-form shape design with pde models and reduced basis method. *Computer Methods in Applied Mechanics and Engineering*, 199(23):1583–1592, 2010.
- [28] Zijie Li, Kazem Meidani, and Amir Barati Farimani. Transformer for partial differential equations’ operator learning, 2023.
- [29] Zongyi Li, Daniel Zhengyu Huang, Burigede Liu, and Anima Anandkumar. Fourier neural operator with learned deformations for pdes on general geometries, 2022.
- [30] Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations, 2020.
- [31] Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Multipole graph neural operator for parametric partial differential equations. In *Proceedings of the 34th International Conference on Neural Information Processing Systems, NIPS’20*, Red Hook, NY, USA, 2020. Curran Associates Inc.
- [32] Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Neural operator: Graph kernel network for partial differential equations, 2020.

- [33] Zongyi Li, Nikola Borislavov Kovachki, Chris Choy, Boyi Li, Jean Kossaifi, Shourya Prakash Otta, Mohammad Amin Nabian, Maximilian Stadler, Christian Hundt, Kamyar Azizzadenesheli, and Anima Anandkumar. Geometry-informed neural operator for large-scale 3d pdes, 2023.
- [34] Ning Liu, Siavash Jafarzadeh, and Yue Yu. Domain agnostic fourier neural operators, 2023.
- [35] Sarah Locke and John Singler. New proper orthogonal decomposition approximation theory for pde solution data. *SIAM Journal on Numerical Analysis*, 58(6):3251–3285, 2020.
- [36] Lu Lu, Pengzhan Jin, Guofei Pang, Zhongqiang Zhang, and George Em Karniadakis. Learning nonlinear operators via deeponet based on the universal approximation theorem of operators. *Nature Machine Intelligence*, 3:218–229, 2021.
- [37] Lu Lu, Xuhui Meng, Shengze Cai, Zhiping Mao, Somdatta Goswami, Zhongqiang Zhang, and George Em Karniadakis. A comprehensive and fair comparison of two neural operators (with practical extensions) based on fair data. *Computer Methods in Applied Mechanics and Engineering*, 393:114778, 2022.
- [38] Winfried Löttsch, Simon Ohler, and Johannes S. Otterbach. Learning the solution operator of boundary value problems using graph neural networks, 2023.
- [39] Andrea Manzoni. *Reduced Models for Optimal Control, Shape Optimization and Inverse Problems in Haemodynamics*. PhD thesis, EPFL, 2012.
- [40] Andrea Manzoni and Federico Negri. *Efficient Reduction of PDEs Defined on Domains with Variable Shape*, pages 183–199. Springer International Publishing, Cham, 2017.
- [41] Andrea Manzoni, Alfio Quarteroni, and Gianluigi Rozza. Shape optimization for viscous flows by reduced basis methods and free-form deformation. *International Journal for Numerical Methods in Fluids*, 70(5):646–670, 2012.
- [42] Nikolaj T. Mücke, Sander M. Bohté, and Cornelis W. Oosterlee. Reduced order modeling for parameterized time-dependent pdes using spatially and memory aware deep learning. *Journal of Computational Science*, 53:101408, 2021.
- [43] Pranshu Pant, Ruchit Doshi, Pranav Bahl, and Amir Barati Farimani. Deep learning for reduced order modelling and efficient temporal evolution of fluid simulations, 2021.
- [44] Tobias Pfaff, Meire Fortunato, Alvaro Sanchez-Gonzalez, and Peter W. Battaglia. Learning mesh-based simulation with graph networks, 2021.
- [45] Michael Prasthofer, Tim De Ryck, and Siddhartha Mishra. Variable-input deep operator networks, 2022.
- [46] Alfio Quarteroni, Andrea Manzoni, and Federico Negri. *Reduced Basis Methods for Partial Differential Equations*. Springer Cham, 2016.
- [47] B.N Sahney and S.P Singh. On best simultaneous approximation in banach spaces. *Journal of Approximation Theory*, 35(3):222–224, 1982.
- [48] Louis Serrano, Lise Le Boudec, Armand Kassai Koupaï, Thomas X Wang, Yuan Yin, Jean-Noël Vittaut, and Patrick Gallinari. Operator learning with neural fields: Tackling pdes on general geometries, 2023.
- [49] Nirav Vasant Shah, Martin Wilfried Hess, and Gianluigi Rozza. *Discontinuous Galerkin Model Order Reduction of Geometrically Parametrized Stokes Equation*, page 551–561. Springer International Publishing, August 2020.
- [50] Pierfrancesco Siena, Michele Girfoglio, Francesco Ballarin, and Gianluigi Rozza. Data-driven reduced order modelling for patient-specific hemodynamics of coronary artery bypass grafts with physical and geometrical parameters. *Journal of Scientific Computing*, 94, 2023.

- [51] John R. Singler. New POD error expressions, error bounds, and asymptotic results for reduced order models of parabolic PDEs. *SIAM Journal on Numerical Analysis*, 52(2):852–876, 2014.
- [52] Alasdair Tran, Alexander Mathews, Lexing Xie, and Cheng Soon Ong. Factorized fourier neural operators. In *The Eleventh International Conference on Learning Representations*, 2023.
- [53] Gege Wen, Zongyi Li, Kamyar Azizzadenesheli, Anima Anandkumar, and Sally M. Benson. U-fno—an enhanced fourier neural operator-based deep-learning model for multiphase flow. *Advances in Water Resources*, 163:104180, 2022.

Appendices

A. Well-posedness of geometrically parametrized problems

We stress that the proof of well-posedness of problem (2) is case-dependent and goes beyond the purpose of the present work: we refer the reader to [46] for more details on the subject. For the sake of simplicity, here we assume that there exists a unique classical solution such that

$$\sup_{\tilde{\mathbf{x}} \in \tilde{\Omega}} |\tilde{u}(\tilde{\mathbf{x}}; t_1, \boldsymbol{\mu}_1, \boldsymbol{\xi}_1) - \tilde{u}(\tilde{\mathbf{x}}; t_2, \boldsymbol{\mu}_2, \boldsymbol{\xi}_2)| \leq \tilde{L} \|(t_1, \boldsymbol{\mu}_1, \boldsymbol{\xi}_1) - (t_2, \boldsymbol{\mu}_2, \boldsymbol{\xi}_2)\|,$$

for $\tilde{L} > 0$ and for any $(t_1, \boldsymbol{\mu}_1, \boldsymbol{\xi}_1), (t_2, \boldsymbol{\mu}_2, \boldsymbol{\xi}_2) \in \mathcal{P} \times \mathcal{T} \times \mathcal{G}$. Then, we employ a push-forward argument to recover the well-posedness of the original formulation entailed by problem (1). In particular,

- $\mathbf{x} \mapsto u(\mathbf{x}; t, \boldsymbol{\mu}, \boldsymbol{\xi})$ is continuous. Letting $L_Z = \max_{\boldsymbol{\xi} \in \mathcal{G}} \text{Lip}(Z(\cdot; \boldsymbol{\xi}))$ and choosing an arbitrary $\mathbf{x}_1 \in \Omega(\boldsymbol{\xi})$, it is straightforward to prove that

$$\forall \varepsilon > 0 \quad \exists \delta > 0 \quad : \quad \|\mathbf{x}_1 - \mathbf{x}_2\| < \delta/L_Z \quad \text{such that} \quad |u(\mathbf{x}_1; t, \boldsymbol{\mu}, \boldsymbol{\xi}) - u(\mathbf{x}_2; t, \boldsymbol{\mu}, \boldsymbol{\xi})| < \varepsilon$$

independently of $t, \boldsymbol{\mu}, \boldsymbol{\xi}$. Indeed, since $\mathbf{x}_i = Z^{-1}(\tilde{\mathbf{x}}_i; \boldsymbol{\xi})$ for $i = 1, 2$, we have

$$\begin{aligned} |u(\mathbf{x}_1; t, \boldsymbol{\mu}, \boldsymbol{\xi}) - u(\mathbf{x}_2; t, \boldsymbol{\mu}, \boldsymbol{\xi})| &= |u(Z^{-1}(\tilde{\mathbf{x}}_1; \boldsymbol{\xi}); t, \boldsymbol{\mu}, \boldsymbol{\xi}) - u(Z^{-1}(\tilde{\mathbf{x}}_2; \boldsymbol{\xi}); t, \boldsymbol{\mu}, \boldsymbol{\xi})| \\ &= |\tilde{u}(\tilde{\mathbf{x}}_1; t, \boldsymbol{\mu}, \boldsymbol{\xi}) - \tilde{u}(\tilde{\mathbf{x}}_2; t, \boldsymbol{\mu}, \boldsymbol{\xi})|. \end{aligned}$$

Since $\|\tilde{\mathbf{x}}_1 - \tilde{\mathbf{x}}_2\| = \|Z(\mathbf{x}_1; \boldsymbol{\xi}) - Z(\mathbf{x}_2; \boldsymbol{\xi})\| \leq L_Z \|\mathbf{x}_1 - \mathbf{x}_2\| < \delta$, we conclude by continuity of $\mathbf{x} \mapsto \tilde{u}(\mathbf{x}; t, \boldsymbol{\mu}, \boldsymbol{\xi})$.

- $(t, \boldsymbol{\mu}, \boldsymbol{\xi}) \mapsto u(t, \boldsymbol{\mu}, \boldsymbol{\xi})$ is Lipschitz. Indeed, given an arbitrary choice of $(t_1, \boldsymbol{\mu}_1, \boldsymbol{\xi}_1), (t_2, \boldsymbol{\mu}_2, \boldsymbol{\xi}_2) \in \mathcal{P} \times \mathcal{T} \times \mathcal{G}$ and identifying $\mathbf{x}_1 = Z^{-1}(\tilde{\mathbf{x}}; \boldsymbol{\xi}_1)$, $\mathbf{x}_2 = Z^{-1}(\tilde{\mathbf{x}}; \boldsymbol{\xi}_2)$ for any $\tilde{\mathbf{x}} \in \tilde{\Omega}$, we have

$$\begin{aligned} \sup_{\substack{\mathbf{x}_1 \in \Omega(\boldsymbol{\xi}_1) \\ \mathbf{x}_2 \in \Omega(\boldsymbol{\xi}_2)}} |u(\mathbf{x}_1; t_1, \boldsymbol{\mu}_1, \boldsymbol{\xi}_1) - u(\mathbf{x}_2; t_2, \boldsymbol{\mu}_2, \boldsymbol{\xi}_2)| &= \sup_{\tilde{\mathbf{x}} \in \tilde{\Omega}} |u(Z^{-1}(\tilde{\mathbf{x}}; \boldsymbol{\xi}_1); t_1, \boldsymbol{\mu}_1, \boldsymbol{\xi}_1) - u(Z^{-1}(\tilde{\mathbf{x}}; \boldsymbol{\xi}_2); t_2, \boldsymbol{\mu}_2, \boldsymbol{\xi}_2)| \\ &= \sup_{\tilde{\mathbf{x}} \in \tilde{\Omega}} |\tilde{u}(\tilde{\mathbf{x}}; t_1, \boldsymbol{\mu}_1, \boldsymbol{\xi}_1) - \tilde{u}(\tilde{\mathbf{x}}; t_2, \boldsymbol{\mu}_2, \boldsymbol{\xi}_2)| \\ &\leq \tilde{L} \|(t_1, \boldsymbol{\mu}_1, \boldsymbol{\xi}_1) - (t_2, \boldsymbol{\mu}_2, \boldsymbol{\xi}_2)\|, \end{aligned}$$

which ensures continuous dependence from data.

B. Analysis of the morphing operators

Lemma B.1. *It is possible to prove that*

$$\begin{aligned} \mathcal{Z}_{\boldsymbol{\xi}} &: L^2(\tilde{\Omega}) \rightarrow L^2_{\zeta}(\Omega(\boldsymbol{\xi})) & \text{as } \tilde{f} \mapsto \tilde{f} \circ Z(\boldsymbol{\xi}) \\ \mathcal{Z}_{\boldsymbol{\xi}}^{-1} &: L^2_{\zeta}(\Omega(\boldsymbol{\xi})) \rightarrow L^2(\tilde{\Omega}) & \text{as } f \mapsto f \circ Z^{-1}(\boldsymbol{\xi}), \end{aligned}$$

are linear bounded functionals.

Proof. Let $\boldsymbol{\xi} \in \mathcal{G}$. Linearity of both $\mathcal{Z}_{\boldsymbol{\xi}}$ and $\mathcal{Z}_{\boldsymbol{\xi}}^{-1}$ is trivial to prove thanks to their definition. Then, owing to the change of variable formula (4), we have that

- $\mathcal{Z}_{\boldsymbol{\xi}}$ is bounded. Indeed,

$$\|\mathcal{Z}_{\boldsymbol{\xi}}\|_{\star} = \sup_{\substack{\tilde{f} \in L^2(\tilde{\Omega}): \\ \|\tilde{f}\|_{L^2(\tilde{\Omega})} \neq 0}} \frac{\|\mathcal{Z}_{\boldsymbol{\xi}}(\tilde{f})\|_{L^2(\Omega(\boldsymbol{\xi}))}}{\|\tilde{f}\|_{L^2(\tilde{\Omega})}} = \sup_{\substack{f \in L^2(\Omega(\boldsymbol{\xi})):\\ \|f\|_{L^2(\Omega(\boldsymbol{\xi}))} \neq 0}} \frac{\|\tilde{f}\|_{L^2(\tilde{\Omega})}}{\|f\|_{L^2(\tilde{\Omega})}} \leq 1.$$

- \mathcal{Z}_ξ^{-1} is bounded. Indeed,

$$\|\mathcal{Z}_\xi^{-1}\|_* = \sup_{\substack{f \in L^2_\zeta(\Omega(\xi)): \\ \|f\|_{L^2_\zeta(\Omega(\xi))} \neq 0}} \frac{\|\mathcal{Z}_\xi^{-1}(f)\|_{L^2(\tilde{\Omega})}}{\|f\|_{L^2_\zeta(\Omega(\xi))}} = \sup_{\substack{f \in L^2_\zeta(\Omega(\xi)): \\ \|f\|_{L^2_\zeta(\Omega(\xi))} \neq 0}} \frac{\|f\|_{L^2_\zeta(\Omega(\xi))}}{\|f\|_{L^2_\zeta(\Omega(\xi))}} \leq 1.$$

□

C. Supplementary proofs

C.1. Proof of Lemma 4.2

Proof. The proof is organized into three parts:

- (i) B is weakly compact since $L^2(\tilde{\Omega})$ is reflexive. Then, it is possible to endow B with a metric d_B that makes it compatible with the weak topology (see [6], Proposition (3.29)).
- (ii) Moreover, we show that $\mathcal{J}_{CGA} : \mathcal{G} \times [B]^N \rightarrow \mathbb{R}$ is weakly lower semi-continuous with respect to the product topology on $\mathcal{G} \times [B]^N$.
- (iii) Since a weakly lower semi-continuous functional attains its infimum on a weakly compact metric space [47], we conclude.

We are left to prove that $\mathcal{J}_{CGA} : \mathcal{G} \times [B]^N \rightarrow \mathbb{R}$ is lower semi-continuous. To this end, consider the sequences $\xi^{(k)} \rightarrow \xi$ and $w_n^{(k)} \rightharpoonup w_n$, for $n = 1, \dots, N$. Now, for the sake of readability we omit the dependencies of u on μ, t . Thanks to $\xi \mapsto u(\xi)$ being Lipschitz-continuous and \mathcal{Z}_ξ^{-1} being bounded for any $\xi \in \mathcal{G}$, then $\xi \mapsto \tilde{u}(\xi) = \mathcal{Z}_\xi^{-1}(u(\xi))$ is continuous. Thus, we obtain $\tilde{u}(\xi^{(k)}) \rightarrow \tilde{u}(\xi)$ and $\tilde{u}(\xi^{(k)}) \rightharpoonup \tilde{u}(\xi)$. Then, it is straightforward to see that for any $n = 1, \dots, N$,

- $(\tilde{u}(\xi^{(k)}), w_n^{(k)})_{L^2(\tilde{\Omega})} \rightarrow (\tilde{u}(\xi), w_n)_{L^2(\tilde{\Omega})}$, since

$$\begin{aligned} & (\tilde{u}(\xi^{(k)}), w_n^{(k)})_{L^2(\tilde{\Omega})} - (\tilde{u}(\xi), w_n)_{L^2(\tilde{\Omega})} = \\ & = (\tilde{u}(\xi^{(k)}), w_n^{(k)})_{L^2(\tilde{\Omega})} - (\tilde{u}(\xi), w_n^{(k)})_{L^2(\tilde{\Omega})} + (\tilde{u}(\xi), w_n^{(k)}) - (\tilde{u}(\xi), w_n)_{L^2(\tilde{\Omega})} \\ & \leq \|w_n^{(k)}\|_{L^2(\tilde{\Omega})} \|\tilde{u}(\xi^{(k)}) - \tilde{u}(\xi)\|_{L^2(\tilde{\Omega})} + (\tilde{u}(\xi), w_n^{(k)} - w_n)_{L^2(\tilde{\Omega})} \rightarrow 0 \end{aligned}$$

thanks to $\{w_n^{(k)}\}_k$ being weak convergent (thus bounded).

- $(\tilde{u}(\xi^{(k)}), w_n^{(k)})_{L^2(\tilde{\Omega})} w_n^{(k)} \rightharpoonup (\tilde{u}(\xi), w_n)_{L^2(\tilde{\Omega})} w_n$, since for any $z \in L^2(\tilde{\Omega})$

$$\begin{aligned} & ((\tilde{u}(\xi^{(k)}), w_n^{(k)})_{L^2(\tilde{\Omega})} w_n^{(k)}, z)_{L^2(\tilde{\Omega})} = \\ & = (\tilde{u}(\xi^{(k)}), w_n^{(k)})_{L^2(\tilde{\Omega})} (w_n^{(k)}, z)_{L^2(\tilde{\Omega})} \rightarrow (\tilde{u}(\xi), w_n)_{L^2(\tilde{\Omega})} (w_n, z)_{L^2(\tilde{\Omega})} = ((\tilde{u}(\xi), w_n)_{L^2(\tilde{\Omega})} w_n, z)_{L^2(\tilde{\Omega})} \end{aligned}$$

Thus, we have that

$$\sum_{n=1}^N (\tilde{u}(\xi^{(k)}), w_n^{(k)})_{L^2(\tilde{\Omega})} w_n^{(k)} \rightharpoonup \sum_{n=1}^N (\tilde{u}(\xi), w_n)_{L^2(\tilde{\Omega})} w_n$$

and by lower semi-continuity of $\|\cdot\|_{L^2(\tilde{\Omega})}$ with respect to the weak topology it is possible to prove that

$$\liminf_{k \rightarrow \infty} \|\tilde{u}(\xi^{(k)}) - \sum_{n=1}^N (\tilde{u}(\xi^{(k)}), w_n^{(k)})_{L^2(\tilde{\Omega})} w_n^{(k)}\|_{L^2(\tilde{\Omega})} \geq \|\tilde{u}(\xi) - \sum_{n=1}^N (\tilde{u}(\xi), w_n)_{L^2(\tilde{\Omega})} w_n\|_{L^2(\tilde{\Omega})},$$

which concludes the proof. □

C.2. Proof of Proposition 4.2

Proof. Fix $N < \infty$. Define $\Omega_{\mathcal{G}} := \overline{\cup_{\xi \in \mathcal{G}} \Omega(\xi)}$ and

$$\varepsilon^* := \min \left\{ 2N \|\mathcal{Z}_{\xi}^{-1}(u)\|_{L^2(\mathcal{T} \times \mathcal{P} \times \mathcal{G}; L^2(\tilde{\Omega}))}, \frac{1}{(2\text{BAE}_{CGA}(N)^{1/2})} \right\}.$$

Then, for any $\xi \in \mathcal{G}$, we define the zero-extension operator ϵ_{ξ} as

$$\epsilon_{\xi}(\mathcal{Z}_{\xi}(v_n^{CGA}(x, \xi))) := \begin{cases} \mathcal{Z}_{\xi}(v_n^{CGA}(\mathbf{x}, \xi)), & \mathbf{x} \in \Omega(\xi) \\ 0, & \mathbf{x} \in \Omega_{\mathcal{G}} \setminus \Omega(\xi), \end{cases}$$

for any $n = 1, \dots, N$. We immediately notice that $\epsilon_{\xi}(\mathcal{Z}_{\xi}(v_n^{CGA})) \in W^{1,2}(\Omega_{\mathcal{G}}) \times \mathcal{G}$. Indeed, setting $m = \min_{\xi \in \mathcal{G}} \min_{\mathbf{x} \in \Omega(\xi)} \zeta(\mathbf{x}; \xi) > 0$, we have

$$\begin{aligned} m \|\epsilon_{\xi}(\mathcal{Z}_{\xi}(w))\|_{L^2(\Omega_{\mathcal{G}} \times \mathcal{G})}^2 &= m \int_{\Omega_{\mathcal{G}} \times \mathcal{G}} \epsilon_{\xi}(\mathcal{Z}_{\xi}(w(\mathbf{x}; \xi)))^2 d\mathbf{x} d\xi \\ &= m \int_{\mathcal{G}} \int_{\Omega(\xi)} \epsilon_{\xi}(\mathcal{Z}_{\xi}(w(\mathbf{x}; \xi)))^2 d\mathbf{x} d\xi \\ &\leq \int_{\mathcal{G}} \int_{\Omega(\xi)} \zeta(\mathbf{x}; \xi) \mathcal{Z}_{\xi}(w(\mathbf{x}; \xi))^2 d\mathbf{x} d\xi \\ &= \int_{\mathcal{G}} \int_{\tilde{\Omega}} w(\tilde{\mathbf{x}}; \xi)^2 d\tilde{\mathbf{x}} d\xi \\ &= \|w\|_{L^2(\tilde{\Omega} \times \mathcal{G})}^2 < \infty, \end{aligned}$$

where w stands for either v_n^{CGA} or its weak derivative. Then, thanks to Gürhing's Theorem [17], there exists a set of neural networks $(\mathbf{x}, \xi) \mapsto \{\hat{v}_n^{CGA}(\mathbf{x}, \xi)\}_{n=1}^N$ such that

$$\|\hat{v}_n^{CGA} - \epsilon_{\xi}(\mathcal{Z}_{\xi}(v_n^{CGA}))\|_{L^2(\Omega_{\mathcal{G}} \times \mathcal{G})} < \frac{\varepsilon^2}{2MN \|\mathcal{Z}_{\xi}^{-1}(u)\|_{L^2(\mathcal{T} \times \mathcal{P} \times \mathcal{G}; L^2(\tilde{\Omega}))}}.$$

where $M = \sup_{\xi \in \mathcal{G}} \sup_{\mathbf{x} \in \Omega(\xi)} \zeta(\mathbf{x}; \xi)$. We note that

$$\begin{aligned} M \|\hat{v}_n^{CGA} - \epsilon_{\xi}(\mathcal{Z}_{\xi}(v_n^{CGA}))\|_{L^2(\Omega_{\mathcal{G}} \times \mathcal{G})} &= M \|\hat{v}_n^{CGA} - \epsilon_{\xi}(\mathcal{Z}_{\xi}(v_n^{CGA}))\|_{L^2(\mathcal{G}; L^2(\Omega_{\mathcal{G}}))} \\ &\geq \|\hat{v}_n^{CGA} - \epsilon_{\xi}(\mathcal{Z}_{\xi}(v_n^{CGA}))\|_{L^2(\mathcal{G}; L^2_{\zeta}(\Omega(\xi)))} \\ &= \|\hat{v}_n^{CGA} - \mathcal{Z}_{\xi}(v_n^{CGA})\|_{L^2(\mathcal{G}; L^2_{\zeta}(\Omega(\xi)))} \\ &= \|\mathcal{Z}_{\xi}^{-1}(\hat{v}_n^{CGA}) - v_n^{CGA}\|_{L^2(\mathcal{G}; L^2(\tilde{\Omega}))} \\ &= \|\mathcal{Z}_{\xi}^{-1}(\hat{v}_n^{CGA}) - v_n^{CGA}\|_{L^2(\tilde{\Omega} \times \mathcal{G})}, \end{aligned}$$

so that

$$\|\mathcal{Z}_{\xi}^{-1}(\hat{v}_n^{CGA}) - v_n^{CGA}\|_{L^2(\tilde{\Omega} \times \mathcal{G})} < \frac{\varepsilon^2}{2N \|\mathcal{Z}_{\xi}^{-1}(u)\|_{L^2(\mathcal{T} \times \mathcal{P} \times \mathcal{G}; L^2(\tilde{\Omega}))}}.$$

The whole architecture is obtained by stacking together the neural networks $\{\hat{v}_n^{CGA}\}_{n=1}^N$. Thus, by employing in sequence the lower triangular inequality, the triangular inequality, and the Cauchy-Schwarz inequality, we

obtain

$$\begin{aligned}
& \left[\int_{\mathcal{G}} \mathcal{J}_{CGA}(\xi, \{\mathcal{Z}_{\xi}^{-1}(\hat{v}_n^{CGA}(\xi))\}_{n=1}^N) \right]^{1/2} - \left[\text{BAE}_{CGA}(N) \right]^{1/2} = \\
& \stackrel{\text{def}}{=} \left\| \mathcal{Z}_{\xi}^{-1}(u) - \sum_{n=1}^N (\mathcal{Z}_{\xi}^{-1}(u), \mathcal{Z}_{\xi}^{-1}(\hat{v}_n^{CGA}))_{L^2(\tilde{\Omega})} \mathcal{Z}_{\xi}^{-1}(\hat{v}_n^{CGA}) \right\|_{L^2(\mathcal{T} \times \mathcal{P} \times \mathcal{G}; L^2(\tilde{\Omega}))} - \\
& \quad - \left\| \mathcal{Z}_{\xi}^{-1}(u) - \sum_{n=1}^N (\mathcal{Z}_{\xi}^{-1}(u), \mathcal{Z}_{\xi}^{-1}(v_n^{CGA}))_{L^2(\tilde{\Omega})} \mathcal{Z}_{\xi}^{-1}(v_n^{CGA}) \right\|_{L^2(\mathcal{T} \times \mathcal{P} \times \mathcal{G}; L^2(\tilde{\Omega}))} \\
& \stackrel{LTI}{\leq} \left\| \sum_{n=1}^N \left[(\mathcal{Z}_{\xi}^{-1}(u), \mathcal{Z}_{\xi}^{-1}(\hat{v}_n^{CGA}))_{L^2(\tilde{\Omega})} \mathcal{Z}_{\xi}^{-1}(\hat{v}_n^{CGA}) \right. \right. \\
& \quad \left. \left. - (\mathcal{Z}_{\xi}^{-1}(u), \mathcal{Z}_{\xi}^{-1}(v_n^{CGA}))_{L^2(\tilde{\Omega})} \mathcal{Z}_{\xi}^{-1}(v_n^{CGA}) \right] \right\|_{L^2(\mathcal{T} \times \mathcal{P} \times \mathcal{G}; L^2(\tilde{\Omega}))} \\
& = \left\| \sum_{n=1}^N \left[(\mathcal{Z}_{\xi}^{-1}(u), \mathcal{Z}_{\xi}^{-1}(\hat{v}_n^{CGA} - v_n^{CGA}))_{L^2(\tilde{\Omega})} \mathcal{Z}_{\xi}^{-1}(\hat{v}_n^{CGA}) \right. \right. \\
& \quad \left. \left. - (\mathcal{Z}_{\xi}^{-1}(u), \mathcal{Z}_{\xi}^{-1}(v_n^{CGA}))_{L^2(\tilde{\Omega})} \mathcal{Z}_{\xi}^{-1}(v_n^{CGA} - \hat{v}_n^{CGA}) \right] \right\|_{L^2(\mathcal{T} \times \mathcal{P} \times \mathcal{G}; L^2(\tilde{\Omega}))} \\
& \stackrel{TI}{\leq} \sum_{n=1}^N \left\| (\mathcal{Z}_{\xi}^{-1}(u), \mathcal{Z}_{\xi}^{-1}(\hat{v}_n^{CGA} - v_n^{CGA}))_{L^2(\tilde{\Omega})} \mathcal{Z}_{\xi}^{-1}(\hat{v}_n^{CGA}) \right\|_{L^2(\mathcal{T} \times \mathcal{P} \times \mathcal{G}; L^2(\tilde{\Omega}))} + \\
& \quad \left\| (\mathcal{Z}_{\xi}^{-1}(u), \mathcal{Z}_{\xi}^{-1}(v_n^{CGA}))_{L^2(\tilde{\Omega})} \mathcal{Z}_{\xi}^{-1}(v_n^{CGA} - \hat{v}_n^{CGA}) \right\|_{L^2(\mathcal{T} \times \mathcal{P} \times \mathcal{G}; L^2(\tilde{\Omega}))} \\
& \stackrel{CS}{\leq} \sum_{n=1}^N \|\mathcal{Z}_{\xi}^{-1}(u)\|_{L^2(\mathcal{T} \times \mathcal{P} \times \mathcal{G}; L^2(\tilde{\Omega}))} \|v_n^{CGA} - \hat{v}_n^{CGA}\|_{L^2(\tilde{\Omega} \times \mathcal{G})} \left[\|v_n^{CGA}\|_{L^2(\tilde{\Omega} \times \mathcal{G})} + \|\hat{v}_n^{CGA}\|_{L^2(\tilde{\Omega} \times \mathcal{G})} \right]
\end{aligned}$$

Now, we observe that

$$\begin{aligned}
\|\hat{v}_n^{CGA}\|_{L^2(\tilde{\Omega} \times \mathcal{G})} & \leq \|v_n^{CGA} - \hat{v}_n^{CGA}\|_{L^2(\tilde{\Omega} \times \mathcal{G})} + \|v_n^{CGA}\|_{L^2(\tilde{\Omega} \times \mathcal{G})} \\
& < \frac{\varepsilon^2}{2N \|\mathcal{Z}_{\xi}^{-1}(u)\|_{L^2(\mathcal{T} \times \mathcal{P} \times \mathcal{G}; L^2(\tilde{\Omega}))}} + 1 \\
& \leq \frac{(\varepsilon^*)^2}{2N \|\mathcal{Z}_{\xi}^{-1}(u)\|_{L^2(\mathcal{T} \times \mathcal{P} \times \mathcal{G}; L^2(\tilde{\Omega}))}} + 1 \leq 2,
\end{aligned}$$

which allows us to obtain the bound

$$\left[\int_{\mathcal{G}} \mathcal{J}_{CGA}(\xi, \{\mathcal{Z}_{\xi}^{-1}(\hat{v}_n^{CGA}(\xi))\}_{n=1}^N) \right]^{1/2} - \left[\text{BAE}_{CGA}(N) \right]^{1/2} < \varepsilon^2.$$

By isolating the first addendum of the l.h.s. and by squaring both sides of the latter equation, we can conclude

$$\int_{\mathcal{G}} \mathcal{J}_{CGA}(\xi, \{\mathcal{Z}_{\xi}^{-1}(\hat{v}_n^{CGA}(\xi))\}_{n=1}^N) < \varepsilon^4 + 2\varepsilon^2 \left[\text{BAE}_{CGA}(N) \right]^{1/2} + \text{BAE}_{CGA}(N) < \varepsilon + \text{BAE}_{CGA}(N),$$

since $\varepsilon < \varepsilon^* \leq 1/(2[\text{BAE}_{CGA}(N)]^{1/2})$. □

MOX Technical Reports, last issues

Dipartimento di Matematica
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

- 43/2025** Tomasetto, M.; Manzoni, A.; Braghin, F.
Real-time optimal control of high-dimensional parametrized systems by deep-learning based reduced order models
- 41/2025** Torzoni, M.; Maisto, D.; Manzoni, A.; Donnarumma, F.; Pezzulo, G.; Corigliano, A.
Active digital twins via active inference
- 42/2025** Franco, N. R.; Manzoni, A.; Zunino, P.; Hesthaven, J. S.
Deep orthogonal decomposition: a continuously adaptive neural network approach to model order reduction of parametrized partial differential equations
- 40/2025** Tentori, C.A.; Gregorio, C.; ...; Ieva, F.; Della Porta, M.G.
Clinical and Genomic-Based Decision Support System to Define the Optimal Timing of Allogeneic Hematopoietic Stem-Cell Transplantation in Patients With Myelodysplastic Syndromes
- 37/2025** Spreafico, M.; Ieva, F.; Fiocco, M.
Causal effect of chemotherapy received dose intensity on survival outcome: a retrospective study in osteosarcoma
- Gimenez Zapiola, A.; Boselli, A.; Menafoglio, A.; Vantini, S.
Hyper-spectral Unmixing algorithms for remote compositional surface mapping: a review of the state of the art
- 35/2025** Perotto, S.; Ferro, N.; Speroni, G.; Temellini, E.
Anisotropic recovery-based error estimators and mesh adaptation for real-life engineering innovation
- 34/2025** Bucelli, M.; Dede', L.
Coupling models of resistive valves to muscle mechanics in cardiac fluid-structure interaction simulations
- 32/2025** De Sanctis, M.F.; Di Battista, I.; Arnone, E.; Castiglione, C.; Palummo, A.; Bernardi, M.; Ieva, F.; Sangalli, L.M.
Exploring nitrogen dioxide spatial concentration via physics-informed multiple quantile regression
- 33/2025** Di Battista, I.; De Sanctis, M.F.; Arnone, E.; Castiglione, C.; Palummo, A.; Sangalli, L.M.
A semiparametric space-time quantile regression model