A Component-Based Data Assimilation Strategy with Applications to Vascular Flows

Duc-Quang Bui, Pierre Mollo, Fabio Nobile, Tommaso Taddei

▶ **To cite this version:**

Duc-Quang Bui, Pierre Mollo, Fabio Nobile, Tommaso Taddei. A Component-Based Data Assimilation Strategy with Applications to Vascular Flows. ESAIM: Proceedings and Surveys, In press, 73, pp.89-106. 10.1051/proc/202373089 . hal-03884154v1

HAL Id: hal-03884154

https://hal.science/hal-03884154v1

Submitted on 5 Dec 2022 (v1), last revised 11 Oct 2023 (v2)

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

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

ESAIM: PROCEEDINGS AND SURVEYS, Vol. ?, 2022, 1-10
Editors: Will be set by the publisher

A COMPONENT-BASED DATA ASSIMILATION STRATEGY WITH APPLICATIONS TO VASCULAR FLOWS *

Duc-Quang Bui[1], Pierre Mollo[2], Fabio Nobile[3] and Tommaso Taddei[4]

Abstract. We present a parameterized-background data-weak (PBDW) approach [Y Maday, AT Patera, JD Penn, M Yano, Int J Numer Meth Eng, 102(5), 933–965] to the steady-state variational data assimilation (DA) problem for systems modeled by partial differential equations (PDEs) and characterized by multiple interconnected components, with emphasis on vascular flows. We focus on the problem of reconstructing the state of the system in one specific component, based on local measurements. The PBDW approach does not require the solution of any PDE model at prediction stage (*projection-by-data*) and, as such, enables local state estimates on single components, as long as good background and update spaces for the estimation can be constructed. We discuss the application of PBDW to a two-dimensional steady Navier-Stokes problem for a family of parameterized geometries, and investigate instead the effects of enforcing no-slip boundary conditions and incompressibility constraints on the background and update spaces to enhance the state estimation. Furthermore, we show an actionable strategy to train local reduced-order bases (ROBs) for the background space that can later be used for DA tasks.

1. Introduction

Data assimilation (DA) refers to the process of integrating experimental observations into a possibly parameterized mathematical model to achieve better prediction of quantities of interest. In particular, state estimation is the DA task that addresses the problem of estimating the (unknown) state u^{true} of a physical system over a domain of interest $\Omega \subset \mathbb{R}^d$. This work deals with the problem of estimating the velocity flow field u^{true} based on sparse velocity measurements: the ultimate goal of the project is to devise a fast and efficient DA state estimation procedure to monitor blood flows in the vascular system.

For clinical applications, it is important to estimate the blood flow and derived quantities such as the wall shear stress over a region of interest. Towards this end, combination of sophisticated 3D models of the vascular flow with *in vivo* patient-specific measurements is key to properly inform clinical decisions. A major issue in this direction is the imposition of inflow and outflow boundary conditions for the 3D vascular model in the vessel of interest Ω, which are typically largely unknown (see, e.g., [29]): to reduce the impact of boundary conditions, several authors have proposed to perform simulations over a much larger region Ω_{net}, encompassing a network of vessels strictly containing the one of interest [7], with the assumption that boundary conditions

* *TT acknowledges the support of IdEx Bordeaux (projet EMERGENCE 2019).*

[1] Laboratoire Analyse, Géométrie et Applications, CNRS UMR 7539, USPN, email: bui@math.univ-paris13.fr

[2] Laboratoire de Mathématiques de Reims, CNRS UMR 9008, URCA, email: pierre.mollo@univ-reims.fr

[3] École Polytechnique Fédérale de Lausanne (EPFL), CSQI-MATH, email: fabio.nobile@epfl.ch

[4] IMB, UMR 5251, Univ. Bordeaux, 33400, Talence, France; INRIA, Inria Bordeaux Sud-Ouest, Team MEMPHIS, 33400, Talence, France, email: tommaso.taddei@inria.fr

© EDP Sciences, SMAI 2022

are better characterized for the larget network Ω_{net} (e.g. knowledge of pressure or flow rate at the level of the aortic valve or in peripheral vessels). Clearly, enlarging the computational domain increases the computational cost and ultimately hinders the ability of numerical simulations to provide reliable answers in a reasonable time frame.

Localization of the problem to the domain of interest allows to reduce the computational domain — and ultimately the cost of the DA task — at the price of introducing uncertainty at the (artificial) interface. In this work, we show that the availability of local measurements in the region of interest enables the solution of DA tasks without the need to solve global problems, nor even to explicitly characterize the global domain Ω_{net} at prediction stage.

Our point of departure is the parameterized-background data-weak (PBDW) approach to state estimation. PBDW was first introduced in [17] and further extended and analyzed in several subsequent works [2, 4, 8, 11, 18, 26]. The key idea of PBDW is to seek an approximation $\widehat{u} = \widehat{z} + \widehat{\eta}$ of the state in the region of interest Ω, employing projection-by-data. The first contribution to $\widehat{u}$, $\widehat{z} \in \mathcal{K}_N$, is the *deduced background* and is designed to address the uncertainty in the values of the model parameters; as discussed in section 2, given the solution manifold $\mathcal{M}$ that collects the solutions to the problem for all values of the parameters, $\mathcal{K}_N$ is a N-dimensional convex closed set that approximates the elements of $\mathcal{M}$. The second contribution to $\widehat{u}$, $\widehat{\eta} \in \mathcal{U}_M$, is the *update estimate* and is designed to address the non-parametric uncertainty due to model inadequacy; here, $\mathcal{U}_M$ is the M-dimensional linear space spanned by the Riesz representers of the observation functionals that model the action of the physical transducers on the true state.

PBDW relies on projection-by-data as opposed to projection-by-model (see, e.g., [21]). Projection-by-data approaches rely exclusively on experimental measurements to estimate the system's state at prediction stage; prior knowledge of the physical system, in the form of a mathematical model, is employed at training stage — i.e., before acquiring measurements — to define the approximation space and/or to define the loss function. On the other hand, projection-by-model approaches explicitly combine experimental data with a (possibly parameterized) mathematical model of the system at prediction stage. As observed in [17], projection-by-data largely eliminates many requirements related to boundary conditions and can accommodate norms that are considerably stronger than the norms required for well-posedness in projection-by-model. Furthermore, as discussed in [27], projection-by-data naturally accommodates localization over a subset (or even a manifold) of Ω_{net}.

The contribution of the present work is twofold. First, we discuss the application of PBDW to vascular flows at moderate Reynolds numbers in parametric geometries. Although PBDW is a projection-by-data approach and does not require the solution of any PDE (not even locally) at prediction stage, it may be convenient to incorporate physical constraints on the approximation (background and update) spaces for the state estimation. In this respect, we investigate the effects of imposing no-slip boundary conditions or enforcing the incompressibility constraint on the PBDW estimate $\widehat{u}$. Second, we illustrate the importance of localization methods for DA tasks and we present an actionable strategy to train local reduced-order bases (ROBs) to build the background set $\mathcal{K}_N$ for vascular flows.

We remark that several approaches for hemodynamics applications based on projection-by-data — but informed by prior knowledge in the form of a PDE — have been considered in the literature (e.g., [1,15]): we refer to the above mentioned literature and also to the recent lecture notes [20] for a thorough review of the subject. In particular, PBDW was previously applied to hemodynamic problems by Galarce et al. [8]: as opposed to this work, the authors of [8] consider the case of perfect (noiseless) measurements and do not address the issue of enforcing incompressibility of the state estimate. On the other hand, the application of PBDW to localized DA was first considered in [27] for an acoustic problem: in this work, we investigate the performance of localization methods for a class of relevant blood flow problems of interest in hemodynamics.

The outline of the paper is as follows. In section 2, we present the problem of interest, we introduce the localized state estimation problem, and we also present the general PBDW approach for state estimation. In section 3, we discuss the construction of the background set $\mathcal{K}_N$ through model order reduction (MOR, [13,22]) techniques, the choice of the update space for PBDW, and the enforcement of the incompressibility constraint.

In section 4, we present several numerical investigations to illustrate the performance of the proposed technique for a representative model problem.

2. Formulation

2.1. **Problem statement**

We denote by $\Omega_{\rm net} \subset \mathbb{R}^2$ the network of interest for which we assume to have a precise characterization of the boundary conditions; we further denote by $(u_{\rm net}, p_{\rm net})$ the velocity-pressure pair that solves the steady Navier-Stokes equations:

$$\begin{cases} -\nu\Delta u_{\rm net} + (u_{\rm net}\cdot\nabla)u_{\rm net} + \nabla p_{\rm net} = 0 & \text{in } \Omega_{\rm net} \\ \nabla\cdot u_{\rm net} = 0 & \text{in } \Omega_{\rm net} \\ u_{\rm net} = g & \text{on } \Gamma_{\rm net,in} \\ u_{\rm net} = 0 & \text{on } \Gamma_{\rm net,wall} \\ \nu\frac{\partial u_{\rm net}}{\partial \mathbf{n}} + p_{\rm net}\mathbf{n} = 0 & \text{on } \Gamma_{\rm net,out} \end{cases} \tag{1}$$

Here, ν is the kinematic viscosity, $\mathbf{n}$ denotes the outward unit normal, $\Gamma_{\rm net,out}$ is the outflow boundary, $\Gamma_{\rm net,wall}$ is the boundary associated with the vessel walls; finally, $\Gamma_{\rm net,in}$ denotes the inflow boundary where we impose the Dirichlet condition g such that

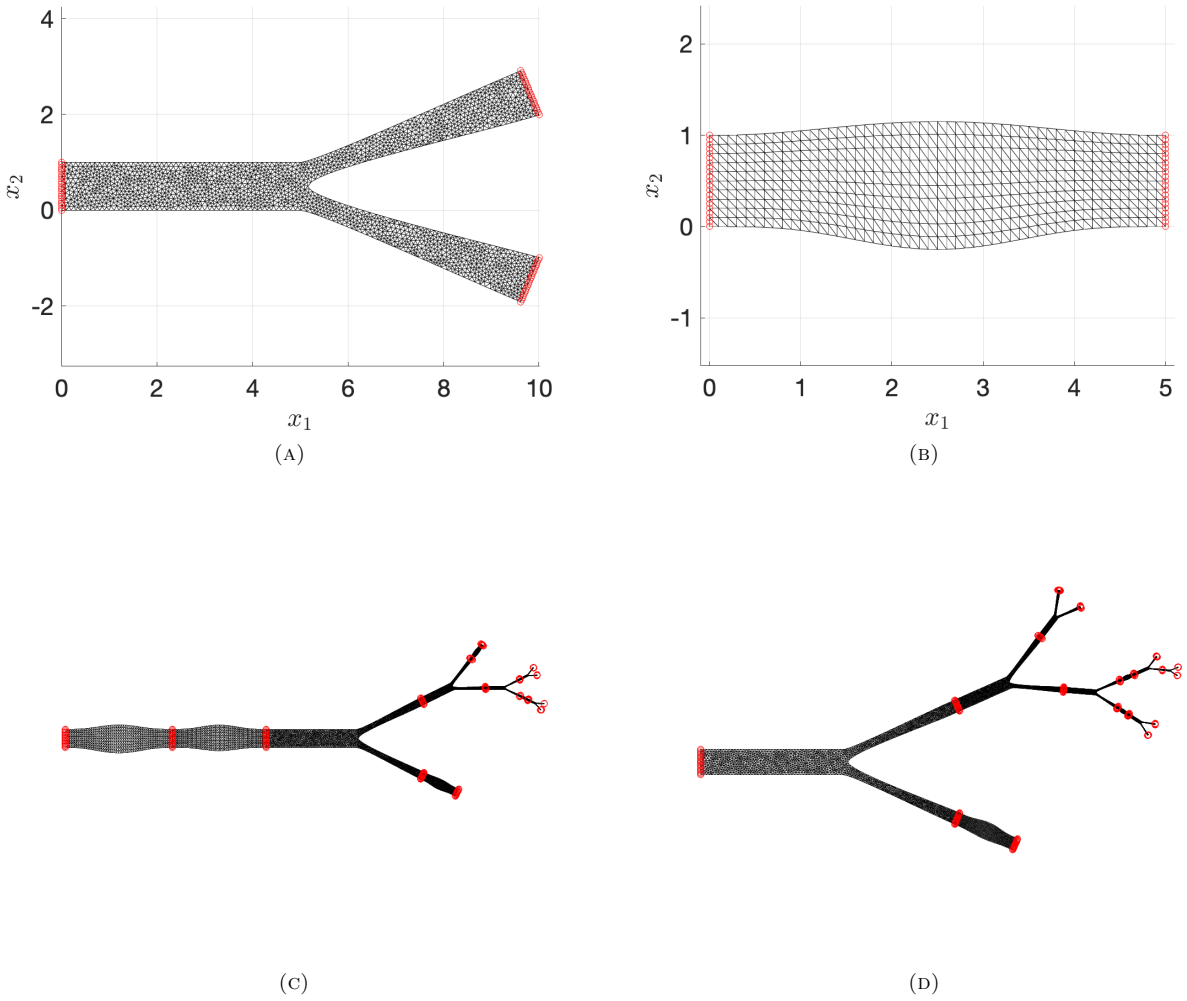
$$g(x_1,x_2) = \left(4\frac{(H-x_2)x_2}{H^2}\right)^{\mu_1^{\rm bnd}} \left(1+\mu_2^{\rm bnd}\sin(3\pi x_2)\right)\begin{bmatrix} 1 \\ 0 \end{bmatrix}. \tag{2}$$

Here, $H>0$ denotes the length of the inflow boundary, which is assumed to be perpendicular to the x_1 axis, and $\mu_1^{\rm bnd}, \mu_2^{\rm bnd}$ are unknown parameters that encode our uncertainty in the inflow BCs. Note that for $\mu_1^{\rm bnd}=1$ and $\mu_2^{\rm bnd}=0$ the inflow condition corresponds to the standard Poiseuille flow; the sinusoidal term mimics the effect of a non-symmetric disturbance, while the exponent $\mu_1^{\rm bnd}$ primarily influences the normal derivative of the velocity field at walls.

We interpret $\Omega_{\rm net}$ as the union of $N_{\rm dd}$ non-overlapping components $\omega_1,\ldots,\omega_{N_{\rm dd}}$. We assume that $\omega_1,\ldots,\omega_{N_{\rm dd}}$ are instantiations of two archetype components [14] $\widehat{\Omega}_1, \widehat{\Omega}_2$; that is, there exist two parametric bijections $\widehat{\Phi}^1, \widehat{\Phi}^2$ and $N_{\rm dd}$ parameters $\mu_1^{\rm net},\ldots,\mu_{N_{\rm dd}}^{\rm net}$ such that

$$\overline{\Omega}_{\rm net} = \bigcup_{j=1}^{N_{\rm dd}} \overline{\omega}_j, \qquad \text{with } \omega_j = \widehat{\Phi}^{\mathtt{L}_j}(\widehat{\Omega}^{\mathtt{L}_j}, \mu_j^{\rm net}),\ \ j=1,\ldots,N_{\rm dd}, \tag{3}$$

where $\mathtt{L}_j \in \{1,2\}$ is the label of the j-th component of the network. In Figure 1(A)-(B), we show the two archetype components — which we dub as "junction" and "channel" — and the associated finite element (FE) meshes. Archetype components are glued together at shared ports $\{\widehat{\Gamma}_{\rm p}^i\}_{i=1,2}$ [14] to form the global network: in Figures 1(C)-(D), we show two examples of vascular networks $\Omega_{\rm net}$ with 10 instantiated components considered in the numerical experiments. We observe that the solution $(u_{\rm net}, p_{\rm net})$ to (1)-(2) is uniquely characterized by (i) the set of $N_{\rm dd}$ labels $\{\mathtt{L}_j\}_{j=1}^{N_{\rm dd}}$, (ii) the set of geometric parameters $\overrightarrow{\mu}_{\rm net} = (\mu_1^{\rm net},\ldots,\mu_{N_{\rm dd}}^{\rm net}) \in \mathcal{P}_{\rm net} = \bigotimes_{j=1}^{N_{\rm dd}} \widehat{\mathcal{P}}^{\mathtt{L}_j}$ where $\widehat{\mathcal{P}}^1$ and $\widehat{\mathcal{P}}^2$ are suitable compact sets associated with the geometrical parameters, and (iii) the global parameter ν and the two parameters $\mu_1^{\rm bnd}, \mu_2^{\rm bnd}$ associated with the inflow BC g in (2). Practical values of the parameters $\nu, \mu_1^{\rm bnd}, \mu_2^{\rm bnd}$ considered in the numerical simulations as well as the definitions of the geometrical parameterization are provided in Appendix A.



(A) (B)

(C) (D)

FIGURE 1. Geometrical configurations. (A)-(B) archetype components for junction (left) and channel (right) and associated finite element meshes. (C)-(D) examples of vascular networks with N_{dd} =10 components. Red points denote the coordinates that belong to the components' ports.

Our goal is to estimate the velocity field $u^{\mathrm{true}} := u_{\mathrm{net}}|_{\Omega}$ in one of the deployed components $\Omega = \omega_{j^\star}$ based on M local measurements of the state

$$y_m = \ell_m^{\mathrm{o}}(u^{\mathrm{true}}) + \varepsilon_m, \quad m = 1, \dots, M, \tag{4a}$$

where $\varepsilon_1, \dots, \varepsilon_M$ are random disturbances, and $\ell_1^{\mathrm{o}}, \dots, \ell_M^{\mathrm{o}} : [H^1(\Omega)]^2 \to \mathbb{R}$ are given by

$$\ell_m^{\mathrm{o}}(v) = \frac{1}{2\pi\sigma_{\mathrm{o}}^2} \int_{\Omega} \exp\left(-\frac{1}{2\sigma_{\mathrm{o}}^2}\|x - x_m^{\mathrm{o}}\|_2^2\right) [\cos(\theta_m), \sin(\theta_m)] \cdot v(x)\, dx \tag{4b}$$

for some $\{x_m^{\rm o}\}_{m=1}^M \subset \Omega$, $\sigma_{\rm o} > 0$, $\{\theta_m\}_{m=1}^M \subset [0, 2\pi)$. In the numerical experiments, we generate $x_m^{\rm o} \overset{\rm iid}{\sim} \text{Uniform}(\Omega)$ and $\theta_m \overset{\rm iid}{\sim} \text{Uniform}(0, 2\pi)$; on the other hand, we consider random Gaussian noise of the form

$$\varepsilon_1, \dots, \varepsilon_M \overset{\rm iid}{\sim} \mathcal{N}\left(0, \left(\frac{\text{std}(\{\ell_m^{\rm o}(u^{\rm true})\}_m)}{\text{SNR}}\right)^2\right) \tag{4c}$$

where $\text{std}(\{x_m\}_m)$ is the sample standard deviation of the vector $\{x_m\}_m$ and SNR is the signal-to-noise ratio that will be specified in the numerical results.

Some comments are in order.

- The particular form of the observation functionals in (4b) is consistent with the choice in [8] for Doppler ultrasound imaging measurements. Note that $\ell_1^{\rm o}, \dots, \ell_M^{\rm o}$ are continuous linear functionals in $[H^1(\Omega)]^2$.
- If we denote by $\mu = \left(\nu, \mu_1^{\rm bnd}, \mu_2^{\rm bnd}, \overrightarrow{\mu}_{\rm net}\right)$ the vector of parameters and by $\mathcal{P}$ the corresponding parameter domain, we may introduce the solution manifold

$$\mathcal{M} := \{u_{\rm net}(\mu)|_\Omega : \mu \in \mathcal{P}\}. \tag{5}$$

  Note that evaluating an element of the manifold $\mathcal{M}$ requires to solve the equations (1) over the whole domain $\Omega_{\rm net}$ at prediction stage: it is thus extremely expensive for practical applications. Moreover, the exact topology of the network might be only partially known. These considerations justify the interest for localized DA techniques.
- In our presentation, we assumed that (i) the true state $u^{\rm true}$ solves the Navier-Stokes equations (1) at moderate Reynolds number for some unknown value of the physical and geometric parameters μ (i.e., non-parametric uncertainty is negligible for the problem at hand); and (ii) the geometry Ω of the domain of interest is fixed and is known exactly — even if the exact shape of the full network is unknown. The particular — steady-state, two-dimensional — PDE model considered is inadequate for realistic vascular flows: it is thus necessary to resort to more appropriate PDE models to achieve accurate blood flow reconstructions. The application of PBDW in combination with more realistic 3D unsteady models is beyond the scope of the present work. Nevertheless, since PBDW relies on projection-by-data, we envision that the present study might represent an important step towards the development of robust DA techniques for real-world hemodynamics applications.

2.2. **PBDW approach to state estimation**

We review the PBDW formulation for state estimation as presented in [9]. We introduce the Hilbert space $\mathcal{X}$ over Ω endowed with the inner product $(\cdot, \cdot)$ and the induced norm $\|\cdot\| = \sqrt{(\cdot, \cdot)}$. Then, the partial spline model (PSM, [31]) reads: find $\widehat{u} = \widehat{z} + \widehat{\eta}$ such that

$$(\widehat{z}, \widehat{\eta}) \in \arg \min_{(z,\eta) \in \mathcal{M} \times \mathcal{X}} \xi \|\eta\|^2 + \|\boldsymbol{\ell}^{\rm o}(z + \eta) - \mathbf{y}\|_2^2, \tag{6}$$

where $\boldsymbol{\ell}^{\rm o} = [\ell_1^{\rm o}, \dots, \ell_M^{\rm o}]^T : \mathcal{X} \to \mathbb{R}^M$ is the vector of observation functionals and $\mathcal{M}$ is the solution manifold (5). PSM reads as a regularized nonlinear least-square problem where the penalization term $\|\eta\|^2$ is designed to penalize deviations from the manifold $\mathcal{M}$ and $\xi > 0$ is a weighting parameter that balances the importance of prior knowledge (encoded in $\mathcal{M}$) with the data misfit $\|\boldsymbol{\ell}^{\rm o}(z+\eta) - \mathbf{y}\|_2$. We also remark that PSM reduces to the well-known 3D-VAR formulation [16] when $\mathcal{M}$ reduces to a singleton. Problem (6) is highly nonlinear and non-convex; furthermore, it requires solutions to the full model. It is thus ill-suited for real-time computations.

To address these issues, we introduce the rank-N approximation of the elements of $\mathcal{M}$:

$$\mathcal{M}_N := \left\{ \sum_{n=1}^N \widehat{\alpha}_n(\mu) \zeta_n : \mu \in \mathcal{P} \right\},$$

where $\{\zeta_n\}_{n=1}^N \subset \mathcal{X}$ is a suitable reduced-order basis that is designed to accurately approximate elements of $\mathcal{M}$ and $\widehat{\boldsymbol{\alpha}} : \mathcal{P} \to \mathbb{R}^n$ is a vector-valued function of generalized coordinates that can be obtained using projection-by-model — i.e., by solving a suitable (Petrov-)Galerkin reduced-order model. If we introduce the convex set A that approximates the range of the function $\widehat{\boldsymbol{\alpha}}$, $\widehat{\boldsymbol{\alpha}}(\mathcal{P})$, we obtain the convex approximation of $\mathcal{M}_N$

$$\mathcal{K}_N := \left\{ \sum_{n=1}^N \alpha_n \zeta_n : \boldsymbol{\alpha} \in A \right\}. \tag{7a}$$

Substitution of $\mathcal{M}$ with $\mathcal{K}_N$ in (6) leads to the general PBDW formulation: find $\widehat{u} = \widehat{z} + \widehat{\eta}$ such that

$$(\widehat{z}, \widehat{\eta}) \in \arg \min_{(z,\eta)\in\mathcal{K}_N\times\mathcal{X}} \xi\|\eta\|^2 + \|\boldsymbol{\ell}^{\mathrm{o}}(z+\eta) - \mathbf{y}\|_2^2. \tag{7b}$$

If $\boldsymbol{\ell}^{\mathrm{o}}(z) = 0$ implies $z \equiv 0$ for any $z \in \mathcal{K}_N$, then it is possible to show that the solution to (7b) exists and is unique (cf. [9, Proposition 2.1]); furthermore, the optimal update $\widehat{\eta}$ belongs to the M-dimensional update space

$$\mathcal{U}_M := \mathrm{span}\{q_m\}_{m=1}^M, \quad \text{where } (q_m, v) = \ell_m^{\mathrm{o}}(v) \quad \forall\, v \in \mathcal{X}, \;\; m = 1, \ldots, M. \tag{7c}$$

The solution to (7) does not rely on an explicit characterization of the elements of the manifold $\mathcal{M}$, and can thus be solved without having to query the underlying PDE model (*projection-by-data*) — in particular, we do not have to prescribe boundary conditions on $\partial\Omega$. We further observe that, if the set A in (7a) is of the form $\{\boldsymbol{\alpha} \in \mathbb{R}^N : \mathbf{A}\boldsymbol{\alpha} \le \mathbf{b}\}$ for some matrix-vector pair $(\mathbf{A}, \mathbf{b})$, problem (7b) is equivalent to a quadratic programming problem with linear constraints of size $N + M$. In Appendix B, we discuss in detail the algebraic counterpart of the PBDW statement in (7) and we provide further details concerning the actual implementation of the method.

We observe that PBDW depends on several elements that should be carefully selected: (i) the ROB $\{\zeta_n\}_{n=1}^N$ and the convex set A in (7a); (ii) the inner product $(\cdot,\cdot)$ that informs the update space $\mathcal{U}_M$ in (7c); and (iii) the choice of the hyper-parameter ξ. Furthermore, for incompressible flows, we should also discuss the practical enforcement of the incompressibility constraint. We address these issues in the subsequent section.

We further observe that the effectiveness of the PBDW approach strongly depends on the existence of accurate low-dimensional rank-N approximations (cf. (7a)) of the solution manifold $\mathcal{M}$: in the setting of localized data assimilation, for diffusion-dominated problems, this property is ensured by the rapid spatial evanescence of high-frequency modes (see, e.g., [24]); on the other hand, we envision that a linear reduction approach might not be effective in the high-Reynolds number regime.

3. Methodology

In this section, we discuss the construction of the many elements of the PBDW formulation for the two-dimensional model problem introduced in section 2. In section 3.1, we discuss the finite element (FE) discretization; in section 3.2, we present the data compression strategy for the construction of the background sets $\mathcal{K}_N$; then, in section 3.3, we propose an actionable strategy for the enforcement of the incompressibility constraint. Finally, in section 3.4, we discuss the choice of the ambient space $\mathcal{X}$ and of the hyper-parameter ξ in (7).

3.1. Finite element formulation

We denote by $\widehat{\mathcal{T}}^{(i)} = \left(\left\{ \widehat{x}_\ell^{(i)} \right\}_{\ell=1}^{N_{\mathrm{hf},(\mathrm{i})}}, \mathtt{T}^{(i)} \right)$ the FE mesh associated with the i-th archetype component with nodes $\left\{ \widehat{x}_\ell^{(i)} \right\}_{\ell=1}^{N_{\mathrm{hf},(\mathrm{i})}}$ and connectivity matrix $\mathtt{T}^{(i)}$. We rely on a P2-P1 Taylor-Hood (TH) FE discretization [28]; we denote by $\widehat{\mathcal{X}}_{\mathrm{hf}}^{(i)}$ and $\widehat{\mathcal{Q}}_{\mathrm{hf}}^{(i)}$ the TH-FE spaces for velocity and pressure, respectively, associated with the mesh

$\widehat{\mathcal{T}}^{(i)}$. Given $u \in \widehat{\mathcal{X}}_{\rm hf}^{(i)}$, we denote by $\mathbf{u} \in \mathbf{R}^{2N_{\rm hf,(i)}}$ the corresponding FE vector of nodal values. Then, we define the deformed meshes (we omit explicit dependence on parameter for simplicity)

$$\mathcal{T}_{\rm net}^{(j)} = \left(\left\{ \Phi_j \left(\widehat{x}_\ell^{(\mathtt{L}_j)} \right) \right\}_{\ell=1}^{N_{\rm net,(j)}}, \mathtt{T}^{(\mathtt{L}_j)} \right), \quad \text{where } \Phi_j = \widehat{\Phi}^{\mathtt{L}_j}(\cdot; \mu_j^{\rm net}), \;\; N_{\rm net,(j)} = N_{\rm hf,(\mathtt{L}_j)}, \tag{8}$$

and the associated FE spaces $\mathcal{X}_{\rm hf}^{(j)}$ and $\mathcal{Q}_{\rm hf}^{(j)}$, for $j = 1, \ldots, N_{\rm dd}$. Finally, we introduce the global FE mesh $\mathcal{T}_{\rm net}$ as the union of the local deployed meshes $\{\mathcal{T}_{\rm net}^{(j)} : j = 1, \ldots, N_{\rm dd}\}$ and the associated velocity and pressure spaces $\mathcal{X}_{\rm hf}$ and $\mathcal{Q}_{\rm hf}$. Given $u \in \mathcal{X}_{\rm hf}$, we denote by $u_j \in \mathcal{X}_{\rm hf}^{(j)}$ the corresponding restriction to ω_j, for $j = 1, \ldots, N_{\rm dd}$.

By construction, given the FE mesh-vector pair $(\mathcal{T}_{\rm net}^{(j)}, \mathbf{u})$ associated with the field $u \in \mathcal{X}_{\rm hf}^{(j)}$, we have that the corresponding pair $(\widehat{\mathcal{T}}^{(\mathtt{L}_j)}, \mathbf{u})$ identifies the mapped field $u \circ \Phi_j \in \widehat{\mathcal{X}}_{\rm hf}^{(\mathtt{L}_j)}$. We further observe that if $u \in \mathcal{X}_{\rm hf}$ satisfies the weak incompressibility constraint

$$\int_{\Omega_{\rm net}} (\nabla \cdot u)\, q \, dx = 0 \quad \forall\, q \in \mathcal{Q}_{\rm hf}, \tag{9a}$$

then, for all $j = 1, \ldots, N_{\rm dd}$, the restriction $u_j = u|_{\omega_j}$ should satisfy

$$\int_{\omega_j} (\nabla \cdot u_j)\, q \, dx = 0 \quad \forall\, q \in \mathcal{Q}_{\rm hf,0}^{(j)} := \{q' \in \mathcal{Q}_{\rm hf}^{(j)} : q'|_{\Gamma_{\rm p}^{(j)}} = 0\}, \tag{9b}$$

where $\Gamma_{\rm p}^{(j)} = \Phi_j(\widehat{\Gamma}_{\rm p}^{(\mathtt{L}_j)})$ is the union of all ports of ω_j (cf. Figure 1). Condition (9b) can be interpreted as an additional piece of information that the PBDW state estimate should (at least approximately) satisfy.

3.2. **Construction of the background set**

We rely on global FE solves to construct reduced order bases (ROBs) $\{\zeta_n^{(i)}\}_{n=1}^N$ and sets $A^{(i)} \subset \mathbb{R}^N$ for each archetype component[1]. Algorithm 1 outlines the computational procedure: we generate $n_{\rm train}$ random networks $\Omega_{\rm net}$ and we approximate the corresponding solution $(u_{\rm net}, p_{\rm net})$ to (1) using the TH-FE method (as in section 3.1, we omit explicit dependence on the parameter to simplify notation); we extract the local solutions, map them back onto the reference configuration, and populate the datasets of local solutions for each archetype component; then, we apply proper orthogonal decomposition (POD, [23, 30]) with respect to the $H^1(\widehat{\Omega}^i)$ inner product to obtain the local ROBs and we define the sets $\{A^{(i)}\}_{i=1,2}$ as the minimal hyper-rectangles that contain the projected training data onto the POD subspace. In Algorithm 1, notation $\mathtt{POD}(\mathcal{D}, (\cdot,\cdot), N)$ signifies that we apply POD on the set $\mathcal{D}$ based on the inner product $(\cdot,\cdot)$ and we extract the first N dominant modes. We rotate the velocity field as follows: $\widetilde{u}^j = \mathtt{Rot}(-\theta_j)(u_{\rm net}^{\rm hf}|_{\omega_j} \circ \Phi_j)$, where θ_j is the angle between the inlet port of the j-th deformed component and the x_1 axis, and $\mathtt{Rot}(\theta)$ is the two-dimensional rotation matrix $\mathtt{Rot}(\theta) = [\cos(\theta), -\sin(\theta); \sin(\theta), \cos(\theta)]$.

[1]To shorten notation, we here assume that the size N of the ROBs is the same for all components.

Algorithm 1 Construction of the background sets (7a)

1: Generate n_{train} global parameters $\{\mu^{(k)}\}_{k=1}^{n_{\text{train}}}$.
2: Set $\mathcal{D}^{(i)} = \emptyset$ for $i = 1, 2$.
3: **for** $k = 1, 2 \dots, n_{\text{train}}$ **do**
4: Solve the TH-FE problem associated to (1) for $\mu = \mu^{(k)}$.
5: Update the local datasets: $\mathcal{D}^{(i)} = \mathcal{D}^{(i)} \cup \{\texttt{Rot}(-\theta_j)(u_{\text{net}}^{\text{hf}}|_{\omega_j} \circ \Phi_j), j = 1, \dots, N_{\text{dd}} : \mathtt{L}_j = i\}$ for $i = 1, 2$.
6: **end for**
7: Generate the ROBs: $\{\widetilde{\zeta}_n^{(i)}\}_{n=1}^N = \texttt{POD}\left(\mathcal{D}^{(i)}, (\cdot,\cdot)_{H^1(\widehat{\Omega}^i)}, N\right)$ for $i = 1, 2$
8: Define $A^{(i)} = \bigotimes_{i=1}^N [a_n^{(i)}, b_n^{(i)}]$ with $a_n^{(i)} = \min_{w \in \mathcal{D}^{(i)}} \left(w, \widetilde{\zeta}_n^{(i)}\right)_{H^1(\widehat{\Omega}^i)}$, $b_n^{(i)} = \max_{w \in \mathcal{D}^{(i)}} \left(w, \widetilde{\zeta}_n^{(i)}\right)_{H^1(\widehat{\Omega}^i)}$, for $i = 1, 2$, $n = 1, \dots, N$.

Algorithm 1 relies on the generation of global networks and global parameters; we postpone the definition of the sampling distribution used to generate the training set to Appendix A. We also observe that the approach relies on n_{train} global solves and might thus be particularly expensive. In the MOR literature, this motivated the development of localized training techniques [3, 5, 24, 25]: a detailed analysis of these approaches as well as their application to the model problem at hand is beyond the scope of this paper. In the numerical experiments, we empirically investigate the impact of the size N_{dd} of the network used at training stage.

Algorithm 1 generates the "master" reduced sets $\widetilde{\mathcal{K}}_N^{(i)} = \{\sum_{n=1}^N \alpha_n \widetilde{\zeta}_n^{(i)} : \boldsymbol{\alpha} \in A^{(i)}\} \subset \widehat{\mathcal{X}}_{\text{hf}}^{(i)}$. Given a new geometry Ω of type $i \in \{1, 2\}$ such that $\Omega = \widehat{\Phi}^i(\widehat{\Omega}^{(i)}; \mu^{\text{net}})$ for some $\mu^{\text{net}} \in \widehat{\mathcal{P}}^i$, we consider the background set[2]

$$\mathcal{K}_N := \left\{ \sum_{n=1}^N \alpha_n \zeta_n : \boldsymbol{\alpha} \in A^{(i)} \right\}, \text{ with } \zeta_n = \texttt{Rot}(\theta) \left(\widetilde{\zeta}_n^{(i)} \circ \left(\widehat{\Phi}^i(\cdot; \mu^{\text{net}}) \right)^{-1} \right), \quad n = 1, \dots, N, \tag{10}$$

where θ denotes the angle between the inlet boundary of Ω and the x_1 axis.

We observe that each element of the ROBs $\{\widetilde{\zeta}_n^{(i)}\}_n$ for $i = 1, 2$ can be stored as a FE mesh-vector pair and then instantiated at prediction stage by deforming the mesh as described in (8). Clearly, the set $\mathcal{K}_N$ defined in (10) is not weakly-divergence free in the sense of (9b). To address this issue, in the framework of MOR of parametric PDEs in parameterized geometries, several authors have proposed the use of Piola transformations (see, e.g., [6] and the references therein). A thorough investigation of the impact of the use of the Piola transformation in our framework is beyond the scope of the present paper.

3.3. **Enforcing the incompressibility constraint through artificial measurements**

In order to weakly enforce the incompressibility constraint, we propose to introduce a set of "artificial" measurements of the form

$$\boldsymbol{\ell}^{\text{div}}(u) = \left[\int_\Omega \psi_1 \left(\nabla \cdot u\right) dx, \dots, \int_\Omega \psi_K \left(\nabla \cdot u\right) dx \right]^T, \quad \text{for } \psi_1, \dots, \psi_K \in L^2(\Omega),\ K \in \mathbb{N}. \tag{11a}$$

We further define the space $\mathcal{Q}_K = \text{span}\{\psi_k\}_{k=1}^K$. Recalling (9b), we have that $\boldsymbol{\ell}^{\text{div}}(u^{\text{true}}) = 0$ if $\psi_1, \dots, \psi_K \in \mathcal{Q}_{\text{hf},0}(\Omega)$ whereas, for arbitrary choices of $\psi_1, \dots, \psi_K \in L^2(\Omega)$, in general $\boldsymbol{\ell}^{\text{div}}(u^{\text{true}}) \neq 0$, yet $\|\boldsymbol{\ell}^{\text{div}}(u^{\text{true}})\|_2$ converges to zero as the mesh size decreases. We consider $\psi_1, \dots, \psi_K$ to be orthonormal in L^2: this implies that

$$\|\boldsymbol{\ell}^{\text{div}}(u)\|_2 = \sup_{\psi \in \mathcal{Q}_K,\, \|\psi\|_{L^2} = 1} (\nabla \cdot u, \psi)_{L^2(\Omega)} = \left(\sum_{j=1}^K (\nabla \cdot u, \psi_j)_{L^2(\Omega)}^2 \right)^{\frac{1}{2}}. \tag{11b}$$

[2]Once again, we omit dependence on the parameter μ^{net}.

Furthermore, given the instantiated ROB $\{\zeta_n\}_{n=1}^N$ (cf. (10)), we choose

$$\mathcal{Q}_{K=N} = \text{span}\{\nabla \cdot \zeta_n\}_{n=1}^N, \quad n = 1, \dots, N; \tag{11c}$$

given the orthonormal basis $\{\psi_n^{\text{div}}\}_{n=1}^N$ of $\mathcal{Q}_N$, we set $\ell_n^{\text{div}}(v) = \int_\Omega \psi_n^{\text{div}}(\nabla \cdot v)\, dx$ for all $v \in \mathcal{X}$ and $n = 1, \dots, N$. Notice that, with this choice, $\boldsymbol{\ell}^{\text{div}}(u^{\text{true}}) \neq 0$ since $q_n^{\text{div}} \notin \mathcal{Q}_{\text{hf},0}(\Omega)$. Since $\ell_1^{\text{div}}, \dots, \ell_N^{\text{div}}$ are linear continuous functionals in $\mathcal{X} \subset H^1(\Omega)$, the PBDW formulation with artificial measurements reads exactly as (7) with $M+N$ measurements and thus an $M+N$-dimensional update space: more precisely, we have

$$(\widehat{z}, \widehat{\eta}) \in \arg \min_{(z,\eta) \in \mathcal{K}_N \times \mathcal{U}_{M+N}} \xi \|\eta\|^2 + \|\boldsymbol{\ell}^{\text{o}}(z+\eta) - \mathbf{y}\|_2^2 + \|\boldsymbol{\ell}^{\text{div}}(z+\eta)\|_2^2. \tag{12a}$$

where

$$\mathcal{U}_{M+N} = \text{span}\{q_m\}_{m=1}^M \cup \text{span}\{q_n^{\text{div}}\}_{n=1}^N, \quad \text{with } (q_n^{\text{div}}, v) = \ell_n^{\text{div}}(v) \ \ \forall\, v \in \mathcal{X}. \tag{12b}$$

In the numerical experiments, we investigate the impact of the artificial measurements on performance.

3.4. **Choice of the ambient space $\mathcal{X}$ and the hyper-parameter ξ**

In this work, we choose $\mathcal{X} = \{v \in \mathcal{X}_{\text{hf}}(\Omega) : v|_{\Gamma_{\text{net,wall}} \cap \partial\Omega} = 0\}$ equipped with the $H^1(\Omega)$ inner product

$$(w, v) = \int_\Omega \nabla w : \nabla v \, + \, wv \, dx.$$

This choice is justified by the fact that, since the elements of $\mathcal{K}_N$ are not divergence-free, the correction $\widehat{\eta}$ might also contribute to ensure the fulfillment of the incompressibility constraint. On the other hand, since the elements of the background space satisfy the no-slip conditions, it is important to ensure that also the update space belongs to $H^1_{0,\Gamma_{\text{net,wall}} \cap \partial\Omega}(\Omega)$. Note that the particular choice of the inner product depends on the physical domain; therefore, the update space cannot be precomputed at training stage and should thus be computed for any new configuration. For this reason, as currently implemented, the method is not well-suited for real-time mesh-independent computations: in section 4, we comment on the computational cost of the state estimation procedure. We envision that, provided that the measurements' locations are known *a priori*, we might employ MOR techniques to the solution to the parameterized Riesz problems and ultimately speed up the assembly of the update space.

We observe that the PBDW statement (7) (and equivalently (12)) depends on the choice of the regularization parameter ξ. As discussed in [26], the optimal choice of ξ depends on the ratio between the experimental noise, the number of measurements, and the accuracy of the background: it is thus highly problem-dependent and difficult to estimate *a priori*. In this work, we rely on holdout validation (see, e.g., [12]) to tune ξ: we randomly split the dataset of measurements into a training set and a validation set; we solve the PBDW problem for several values of ξ in a given set $\Xi_{\text{train}} \subset \mathbb{R}_+$ and we pick $\xi^\star$ that minimizes the mean square error over the validation set; finally, we solve PBDW over the whole dataset of measurements with $\xi = \xi^\star$. In the numerical experiments, we consider a $80\% - 20\%$ training-validation split.

4. Numerical results

We present the results of the numerical investigations for the model problem introduced in section 2. We assess performance through the vehicle of a dataset of simulations, which are constructed using $n_{\text{test}} = 10$ networks with $N_{\text{dd}} = 10$ components; In more detail, we randomly pick five components of type "junction" and five components of type "channel" among the ones present in the dataset of test networks. The figures below provide the behavior of the relative L^2 or H^1 prediction error

$$E[u^{\text{true}}] = \frac{\|u^{\text{true}} - \widehat{u}\|_\star}{\|u^{\text{true}}\|_\star}, \quad \star = H^1(\Omega) \text{ or } \star = L^2(\Omega); \tag{13}$$

To properly take into account the experimental noise, we repeat the estimation $N_{\text{rep}} = 100$ times and we show the $25 - 50 - 75$ quantiles of the error over the selected instantiated components and the evaluation of the random noise.

Figure 2 investigates the influence of the size N_{dd} of the networks used at training stage. In more detail, we build the background space $\mathcal{K}_N$ with $A = \mathbb{R}^N$ based on n_{train} simulations on networks with N_{dd} components and then we measure the relative projection error

$$E^{\text{proj}}[u^{\text{true}}] = \frac{\inf_{\zeta \in \mathcal{K}_N} \|u^{\text{true}} - \zeta\|_{H^1(\Omega)}}{\|u^{\text{true}}\|_{H^1(\Omega)}}$$

on the test set. To ensure that the training sets are all of comparable size, we set $n_{\text{train}} = \left\lfloor \frac{200}{N_{\text{dd}}} \right\rfloor$. We observe that the projection error reaches a plateau for $N_{\text{dd}} \gtrsim 5$ or 6 for both archetype components and all values of N; this result suggests that — for the considered operating conditions — the training can be performed on networks of modest size.

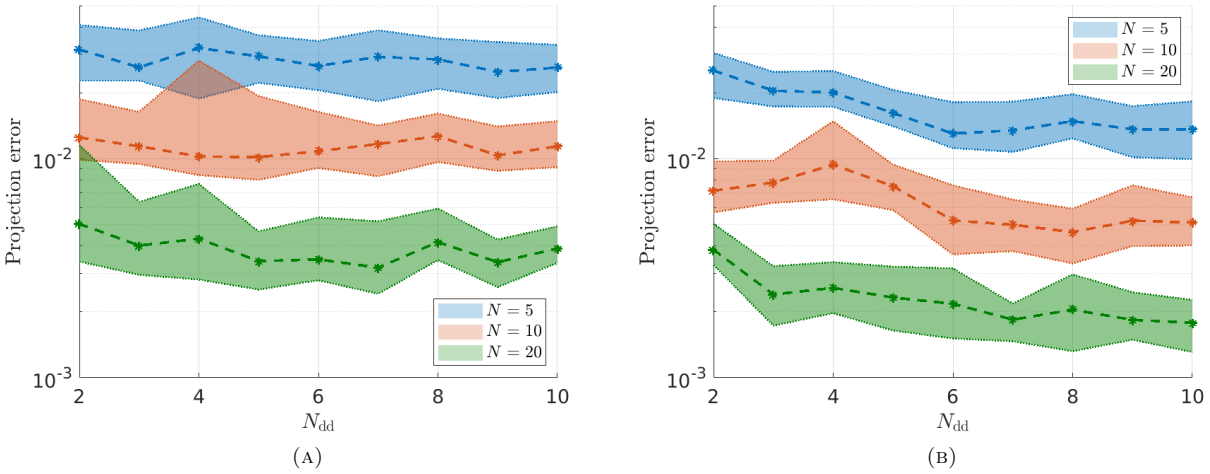


FIGURE 2. Relative H^1 projection error for the two archetype components: junction (A) and channel (B). Training set is based on global solves in networks with $N_{\text{dd}} = 2, \ldots, 10$ components; the test set is extracted from $n_{\text{test}} = 10$ simulations in networks with 10 components.

Figure 3 shows the behavior of the L^2 relative prediction error (13) of the PBDW estimate for the channel component with respect to the size N of the background set and for two choices of M and SNR. As in (12), M denotes the number of local measurements (4) in the component of interest; the number of artificial measurements is set equal to N. We recall that PBDW is ill-posed for $N > M$. We observe that the optimal value of N weakly depends on the amount of measurements and also on the experimental noise. We remark that the introduction of the box constraints (see definition of $\mathcal{K}_N$ in Algorithm 1) is important to stabilize results for large values of N (see [9]). In all cases considered, the introduction of artificial measurements improves performance of the state estimation procedure.

Figure 4 shows the behavior of the relative L^2 prediction error with respect to the signal-to-noise ratio SNR, for two choices of M and N and for the two archetype components. We observe that the introduction of artificial measurements is beneficial, particularly for noisy (i.e., small SNR) measurements. The impact of artificial measurements on performance is also a function of the number of available measurements M.

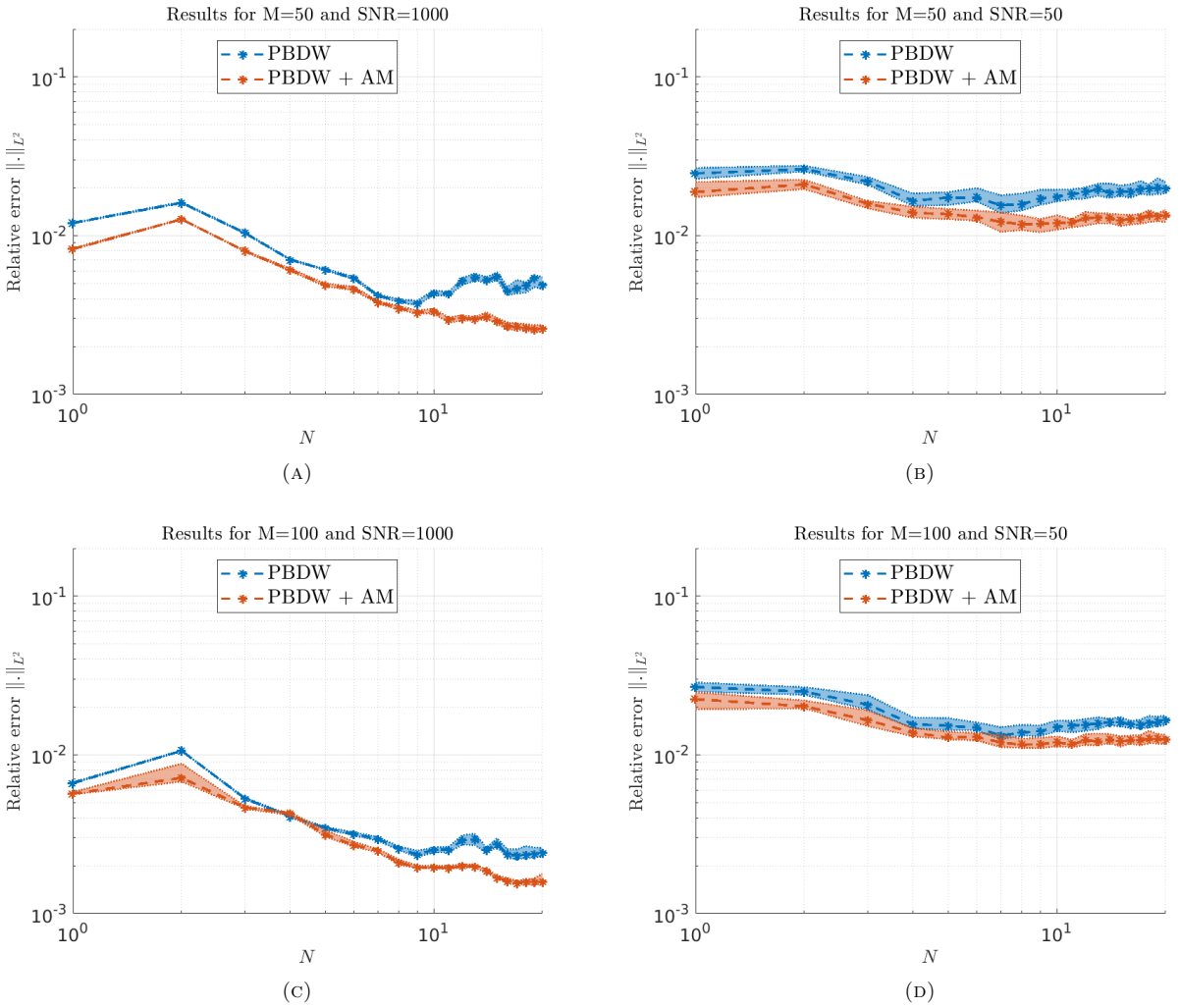


FIGURE 3. Relative L^2 prediction error with respect to the background set dimension N, for two choices of SNR and two choices of the number of measurements in the component of interest M. Results refer to the channel components. The acronym AM stands for artificial measurements; the number of artificial measurements is set equal to N.

Note that the results in Figure 4(B) may be justified by the fact that $\ell^{\mathrm{div}}(u^{\mathrm{true}}) \neq 0$ due to the numerical discretization error.

Figure 5 shows the behavior of the relative L^2 prediction error with respect to the number M of measurements in the range $[25, 100]$ for two values of N and SNR. As for the previous test cases, we observe that the introduction of artificial measurements is beneficial for moderate values of M.

Figure 6 investigates the effect of artificial measurements on the divergence of the predicted field. For consistency with the underlying numerical scheme, we compute the dual norm:

$$E^{\mathrm{div}}[\widehat{u}] = \sup_{q \in \mathcal{Q}_{\mathrm{hf}} : q|_{\Gamma_{\mathrm{p}}} = 0} \frac{\int_\Omega (\nabla \cdot \widehat{u})\, q \, dx}{\|q\|_{L^2(\Omega)}}. \tag{14}$$

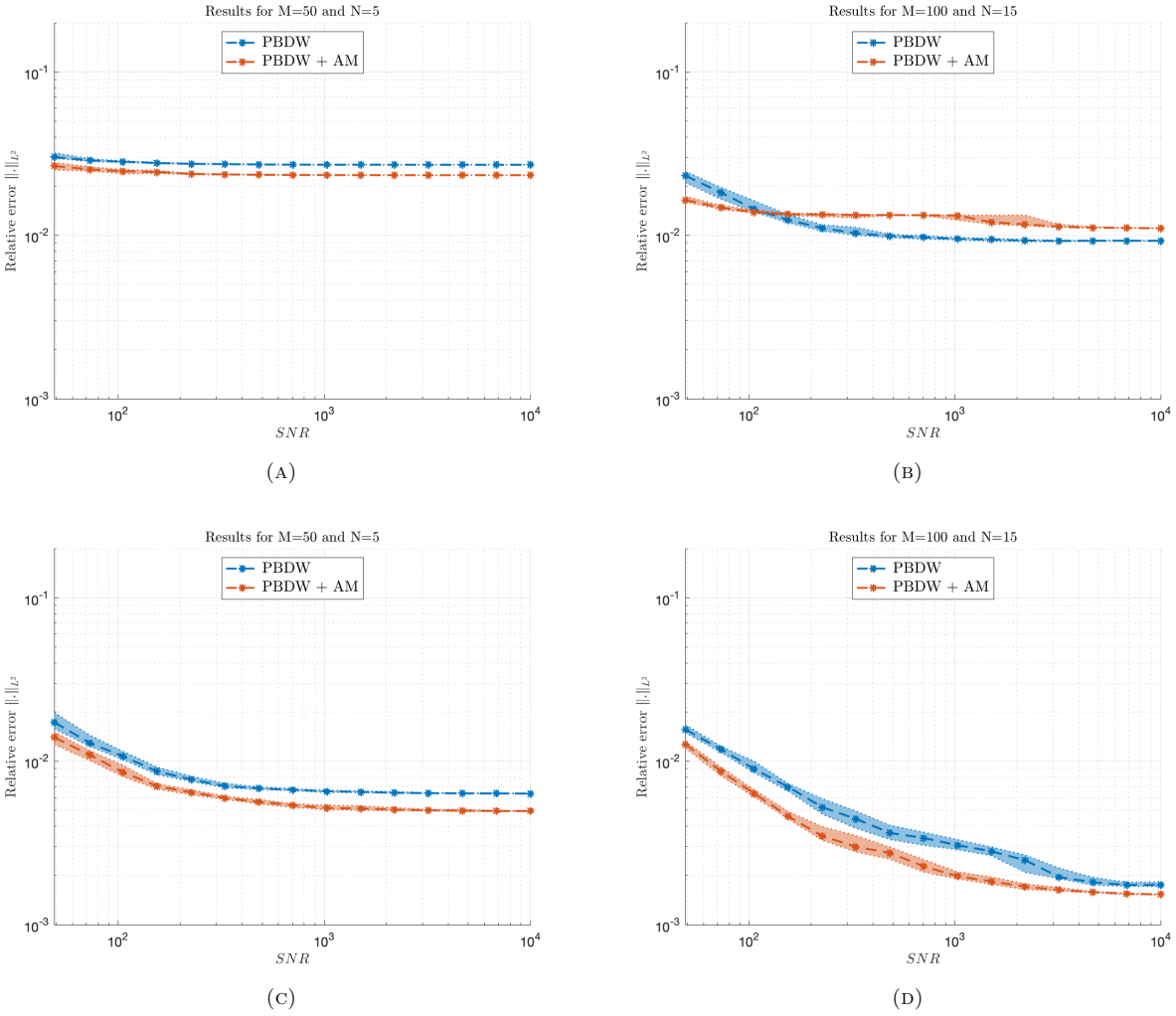


FIGURE 4. Relative L^2 prediction error with respect to the signal-to-noise ratio SNR, for two choices of M and N, and both junction (top) and channel (bottom) components.

Note that the TH FE solution u^{true} satisfies $E^{\text{div}}[u^{\text{true}}] = 0$ for any selected component Ω. Figures 6 (A) and (B) show the results for the junction component with respect to the background dimension N, for $M = 50$, and for $SNR = 50$ and $SNR = 1000$, respectively; Figures 6 (C) and (D) replicate the test for the channel component. We observe that the introduction of the artificial measurements improves the satisfaction of the divergence-free constraint for the junction component; on the other hand, we find that artificial measurements are less impactful for the channel component. We conjecture that this is due to the particular parametric deformations considered for the two archetype components.

Figure 7 shows the wall-clock computational cost in seconds of the PBDW procedure; computations are performed using the cluster ROMEO[3] of the University of Reims. We show results with respect to the number of measurements for various choices of the background dimension N, for both PBDW and PBDW with artificial measurements. To facilitate the interpretation, we report both the total prediction cost and the cost of

[3]https://romeo.univ-reims.fr/

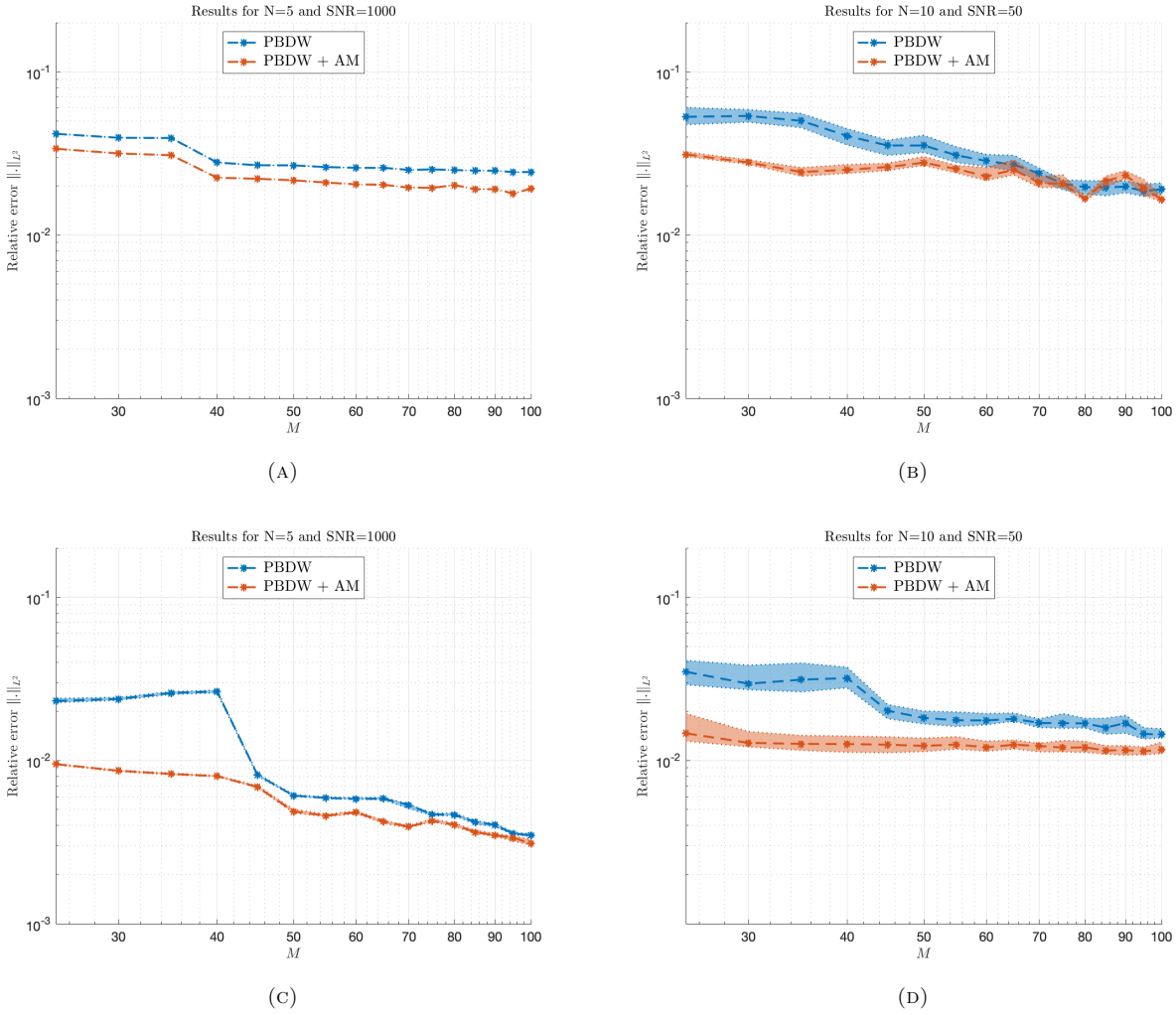
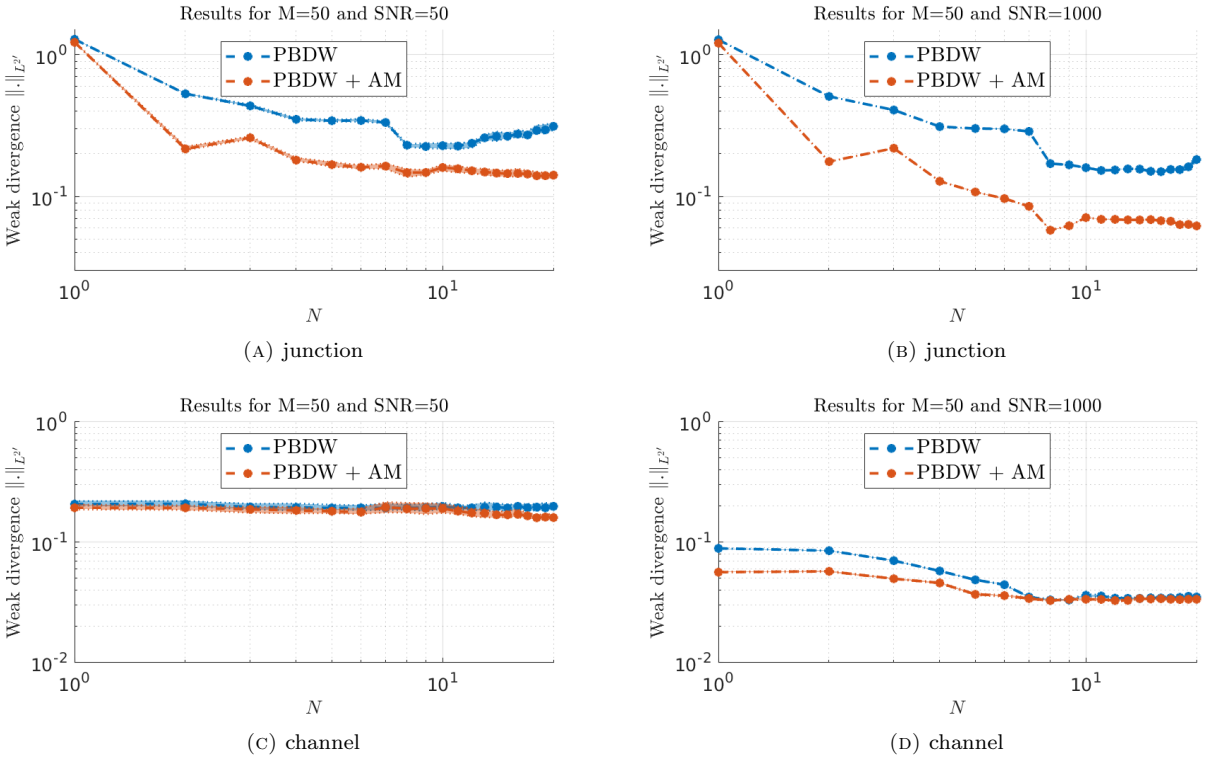


FIGURE 5. Relative L^2 prediction error with respect to the number of measurements M, for two choices of N and SNR, and both junction (top) and channel (bottom) components.

computing the update space. We observe that for this problem the cost of PBDW is below 0.5 seconds for all choices of M, N, with and without cross-validation and artificial measurements. We also observe that the cost of computing the update space exceeds 75% of the total prediction cost in the absence of holdout validation, while it is approximately 50% of the total prediction cost in the presence of holdout validation for the selection of the Tykhonov regularization parameter ξ.

5. CONCLUSION

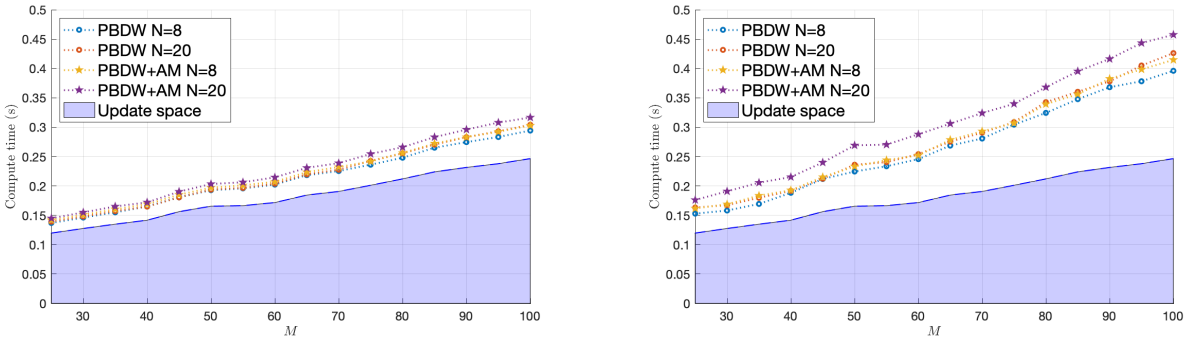
We illustrated the application of the PBDW approach to a two-dimensional incompressible flow problem in parametric geometries. In more detail, we discussed the localization of the state estimation problem to a subregion of interest of the global network, and the weak enforcement of the incompressibility constraint through the introduction of suitable artificial measurements. No-slip BCs at vessels' walls are enforced by



(A) junction (B) junction

(C) channel (D) channel

FIGURE 6. Behavior of $E^{\text{div}}[\widehat{u}]$ with respect to the background set dimension N, for two choices of SNR and for both junction and channel components.



(A) junction, without holdout validation (B) junction, with holdout validation

FIGURE 7. Average prediction computational cost (in seconds). Results refer to the junction component.

properly selecting the ambient space $\mathcal{X}$. Numerical results of section 4 show that the introduction of artificial measurements improves performance, particularly for moderate values of M and SNR; furthermore, they also

suggest that training of localized spaces might be performed on significantly smaller networks than the ones considered at prediction stage.

In view of the application of our approach to real-world hemodynamics problems, much work is still needed. First, we wish to extend our approach to three-dimensional unsteady problems, with fully-automated geometry registration techniques for the construction of the mapping $\Phi : \widehat{\Omega}^\star \to \Omega$ for new configurations. Second, we wish to develop localized training techniques to avoid expensive global solves (cf. Algorithm 1): in this respect, recent works on localized training for nonlinear PDEs [3, 25] might represent the point of departure for the definition of effective background sets $\mathcal{K}_N$. Third, we here propose to build the set $\mathcal{K}_N$ purely based on approximation considerations. As discussed in the literature, approximation performance depends on the product between approximation and stability (see, e.g., [17, Proposition 2]). If the observation functionals in (7) are known *a priori*, it might be important to devise specialized data compression techniques that are explicitly informed by the available observations [19].

A. Parameterization of the model problem

We consider $\nu \in [0.05, 0.2]$ and we set $(\mu_1^{\text{bnd}}, \mu_2^{\text{bnd}}) \in [0.4, 1] \times [-0.1, 0.1]$ in (1); note that the Reynolds number ranges from 5 to 20 for all test cases. To define the deformation maps, we first prescribe parametric analytic expressions of the boundaries in terms of 6 independent parameters β for each archetype component in the parameter regions $\widetilde{\mathcal{P}}^1$ and $\widetilde{\mathcal{P}}^2$, and then we extend the map to the whole domain through transfinite interpolation [10] (cf. Figure 8); to ensure continuity, we also consider a rotation parameter θ, a shift $\bar{x}$, and a scaling factor s; it is thus convenient to write the maps as

$$\widehat{\Phi}^i(\widehat{x}; \mu^{\text{net}} := [\beta, \theta, \bar{x}, s]) = \bar{x} + s\,\text{Rot}(\theta)\widetilde{\Phi}^i(\widehat{x}; \mu), \quad i = 1, 2, \;\; \beta \in \widetilde{\mathcal{P}}^i,$$

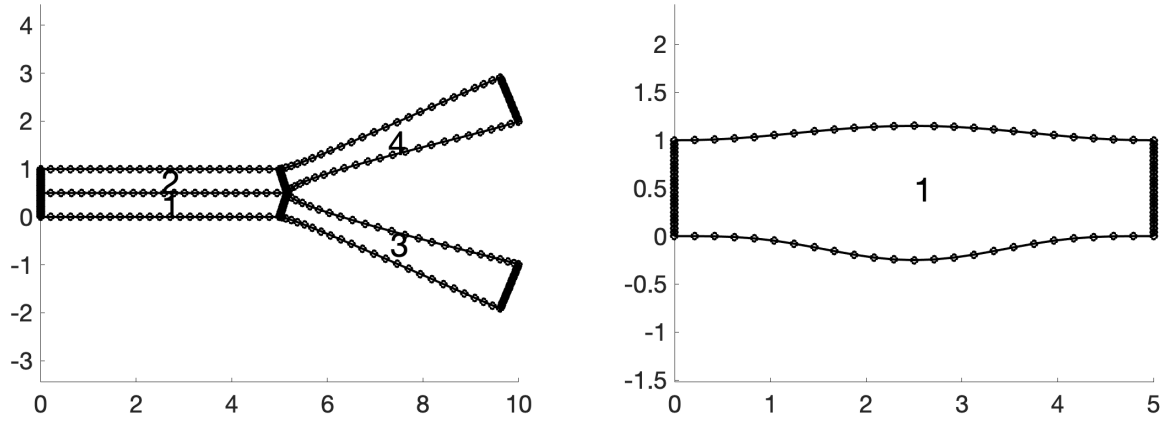
where $\widetilde{\Phi}^1, \widetilde{\Phi}^2$ are the piece-wise transfinite maps associated with the partitions Figure 8. To generate each global network, we proceed iteratively as described in Algorithm 2.

Algorithm 2 Generation of global networks.

Initialize $\Omega_{\text{net}} = \emptyset$ with $N_{\text{p}} = 0$ ports.
for $j = 1, \dots, N_{\text{dd}}$ **do**
  Sample the component index $i \sim \text{Uniform}(\{1, 2\})$ and the parameter $\beta_j \sim \text{Uniform}(\widetilde{\mathcal{P}}^i)$.
  if $i = 1$ **then**
    Set $\omega_j = \widetilde{\Phi}^i(\widehat{\Omega}^i; \beta_j)$.
  else
    Sample the port $\ell \in \text{Uniform}\,(\{1, \dots, N_{\text{p}}\})$ to which the new component will be attached.
    Determine the orientation θ, the scale s and the shift $\bar{x}$ of the new component.
    Set $\omega_j = \bar{x} + s\,\text{Rot}(\theta)\widetilde{\Phi}^i\left(\widehat{\Omega}^i; \beta_j\right)$.
  end if
  Update $\Omega_{\text{net}} = \Omega_{\text{net}} \cup \omega_j$ and update the list of ports $\{\Gamma_\ell\}_{\ell=1}^{N_{\text{p}}}$.
end for

B. Algebraic PBDW formulation

We review the algebraic counterpart of the PBDW formulation (7) as implemented in our code; we refer to [9, section 2.2] for further details. Towards this end, given the trial ROB $\{\zeta_n\}_{n=1}^N$ and the Riesz representers



(A) (B)

FIGURE 8. Geometrical mappings. Maps $\widetilde{\Phi}^1, \widetilde{\Phi}^2$ are defined by prescribing parametric analytic expressions of the boundaries via transfinite interpolation.

$\{q_m\}_{m=1}^M$, we define the matrices $\mathbf{L} \in \mathbb{R}^{M\times N}$ and $\mathbf{K} \in \mathbb{R}^{M\times M}$ such that

$$\mathbf{L}_{m,n} = \ell_m^{\mathrm{o}}(\zeta_n), \quad \mathbf{K}_{m,m'} = \ell_m^{\mathrm{o}}(q_{m'}) = (q_m, q_{m'}), \quad m, m' = 1, \dots, M,\ n = 1, \dots, N.$$

Then, we define $\widehat{\boldsymbol{\alpha}} \in \mathbb{R}^N$ and $\widehat{\boldsymbol{\eta}} \in \mathbb{R}^M$ such that

$$\widehat{u}(x) = \sum_{n=1}^{N} \left(\widehat{\boldsymbol{\alpha}}\right)_n \zeta_n(x) \ + \ \sum_{m=1}^{M} \left(\widehat{\boldsymbol{\eta}}\right)_m q_m(x), \quad x \in \Omega.$$

Finally, we obtain the algebraic counterpart of (7):

$$(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\eta}}) \in \arg \min_{(\boldsymbol{\alpha},\boldsymbol{\eta})\in A\times\mathbb{R}^M} \ \xi \boldsymbol{\eta}^T \mathbf{K} \boldsymbol{\eta} \ + \ \|\mathbf{K}\boldsymbol{\eta} + \mathbf{L}\boldsymbol{\alpha} - \mathbf{y}\|_2^2. \tag{15}$$

Since problem (15) is quadratic in $\boldsymbol{\eta}$, we have the following relationship between $\widehat{\boldsymbol{\eta}}$ and $\widehat{\boldsymbol{\alpha}}$:

$$\widehat{\boldsymbol{\eta}} = \mathbf{W}_\xi \left(\mathbf{y} - \mathbf{L}\widehat{\boldsymbol{\alpha}}\right), \quad \mathbf{W}_\xi := \left(\xi \mathbf{Id} + \mathbf{K}\right)^{-1}, \tag{16a}$$

which can be restated as $\mathbf{K}\widehat{\boldsymbol{\eta}} + \mathbf{L}\widehat{\boldsymbol{\alpha}} - \mathbf{y} = -\xi\widehat{\boldsymbol{\eta}}$. By substituting (16a) in (15), we then find that

$$\xi \widehat{\boldsymbol{\eta}}^T \mathbf{K} \widehat{\boldsymbol{\eta}} \ + \ \|\mathbf{K}\widehat{\boldsymbol{\eta}} + \mathbf{L}\widehat{\boldsymbol{\alpha}} - \mathbf{y}\|_2^2 = \xi \left(\mathbf{y} - \mathbf{L}\widehat{\boldsymbol{\alpha}}\right)^T \mathbf{W}_\xi \mathbf{W}_\xi^{-1} \mathbf{W}_\xi \left(\mathbf{y} - \mathbf{L}\widehat{\boldsymbol{\alpha}}\right),$$

which implies

$$\widehat{\boldsymbol{\alpha}} \in \arg\min_{\boldsymbol{\alpha}\in A} \ \|\mathbf{y} - \mathbf{L}\boldsymbol{\alpha}\|_{\mathbf{W}_\xi}. \tag{16b}$$

We conclude that the solution to (15) can be decomposed in the solution to an N-dimensional constrained optimization problem followed by an M-dimensional linear problem.

Algorithm 3 summarizes the computational procedure. If (15) needs to be solved for several values of ξ, it is convenient to compute and store the eigendecomposition of $\mathbf{K}$ to speed up the computation of $\mathbf{W}_\xi$.

Algorithm 3 Solution to (7).

1: Compute $\mathbf{W}_\xi := (\xi\mathbf{Id} + \mathbf{K})^{-1}$.
2: Solve $\widehat{\boldsymbol{\alpha}} \in \arg\min_{\boldsymbol{\alpha}\in A} \|\mathbf{y} - \mathbf{L}\boldsymbol{\alpha}\|_{\mathbf{W}_\xi}$.
3: Compute $\widehat{\boldsymbol{\eta}} = \mathbf{W}_\xi\,(\mathbf{y} - \mathbf{L}\widehat{\boldsymbol{\alpha}})$.

References

[1] L. Azzimonti, L. M. Sangalli, P. Secchi, M. Domanin, and F. Nobile, *Blood flow velocity field estimation via spatial regression with PDE penalization*, Journal of the American Statistical Association, 110 (2015), pp. 1057–1071.

[2] A. Benaceur, *Reducing sensors for transient heat transfer problems by means of variational data assimilation*, The SMAI journal of computational mathematics, 7 (2021), pp. 1–25.

[3] A. Benaceur and A. T. Patera, *Port-reduced reduced-basis component method for steady state Navier–Stokes and passive scalar equations*, tech. rep., MIT, 2022.

[4] P. Binev, A. Cohen, W. Dahmen, R. DeVore, G. Petrova, and P. Wojtaszczyk, *Data assimilation in reduced modeling*, SIAM/ASA Journal on Uncertainty Quantification, 5 (2017), pp. 1–29.

[5] J. L. Eftang and A. T. Patera, *Port reduction in parametrized component static condensation: approximation and a posteriori error estimation*, International Journal for Numerical Methods in Engineering, 96 (2013), pp. 269–302. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/nme.4543.

[6] E. Fonn, H. van Brummelen, T. Kvamsdal, and A. Rasheed, *Fast divergence-conforming reduced basis methods for steady navier–stokes flow*, Computer Methods in Applied Mechanics and Engineering, 346 (2019), pp. 486–512.

[7] L. Formaggia, A. Quarteroni, and A. Veneziani, *Cardiovascular Mathematics: Modeling and simulation of the circulatory system*, vol. 1, Springer Science & Business Media, 2010.

[8] F. Galarce, D. Lombardi, and O. Mula, *Reconstructing haemodynamics quantities of interest from doppler ultrasound imaging*, International Journal for Numerical Methods in Biomedical Engineering, 37 (2021), p. e3416.

[9] H. Gong, Y. Maday, O. Mula, and T. Taddei, *PBDW method for state estimation: error analysis for noisy data and nonlinear formulation*, arXiv preprint arXiv:1906.00810, (2019).

[10] W. J. Gordon and C. A. Hall, *Construction of curvilinear co-ordinate systems and applications to mesh generation*, International Journal for Numerical Methods in Engineering, 7 (1973), pp. 461–477.

[11] J. K. Hammond, R. Chakir, F. Bourquin, and Y. Maday, *PBDW: A non-intrusive Reduced Basis Data Assimilation method and its application to an urban dispersion modeling framework*, Applied Mathematical Modelling, 76 (2019), pp. 1–25.

[12] T. Hastie, R. Tibshirani, and J. H. Friedman, *The elements of statistical learning: data mining, inference, and prediction*, vol. 2, Springer, 2009.

[13] J. S. Hesthaven, G. Rozza, and B. Stamm, *Certified Reduced Basis Methods for Parametrized Partial Differential Equations*, SpringerBriefs in Mathematics, Springer International Publishing, Cham, 2016.

[14] D. B. P. Huynh, D. J. Knezevic, and A. T. Patera, *A static condensation reduced basis element method: Complex problems*, Computer Methods in Applied Mechanics and Engineering, 259 (2013), pp. 197–216.

[15] G. Kissas, Y. Yang, E. Hwuang, W. R. Witschey, J. A. Detre, and P. Perdikaris, *Machine learning in cardiovascular flows modeling: Predicting arterial blood pressure from non-invasive 4d flow mri data using physics-informed neural networks*, Computer Methods in Applied Mechanics and Engineering, 358 (2020), p. 112623.

[16] A. C. Lorenc, *Analysis methods for numerical weather prediction*, Quarterly Journal of the Royal Meteorological Society, 112 (1986), pp. 1177–1194.

[17] Y. Maday, A. T. Patera, J. D. Penn, and M. Yano, *A parameterized-background data-weak approach to variational data assimilation: formulation, analysis, and application to acoustics*, International Journal for Numerical Methods in Engineering, 102 (2015), pp. 933–965.

[18] Y. Maday and T. Taddei, *Adaptive PBDW approach to state estimation: noisy observations; user-defined update spaces*, SIAM Journal on Scientific Computing, 41 (2019), pp. B669–B693.

[19] L. Mathelin, K. Kasper, and H. Abou-Kandil, *Observable dictionary learning for high-dimensional statistical inference*, Archives of Computational Methods in Engineering, 25 (2018), pp. 103–120.

[20] O. Mula, *Inverse problems: A deterministic approach using physics-based reduced models*, arXiv preprint arXiv:2203.07769, (2022).

[21] F. Naets, J. Croes, and W. Desmet, *An online coupled state/input/parameter estimation approach for structural dynamics*, Computer Methods in Applied Mechanics and Engineering, 283 (2015), pp. 1167–1188.

[22] A. Quarteroni, A. Manzoni, and F. Negri, *Reduced Basis Methods for Partial Differential Equations*, vol. 92 of UNITEXT, Springer International Publishing, Cham, 2016.

[23] L. Sirovich, *Turbulence and the dynamics of coherent structures. I. Coherent structures*, Quarterly of applied mathematics, 45 (1987), pp. 561–571.

[24] K. Smetana and A. T. Patera, *Optimal local approximation spaces for component-based static condensation procedures*, SIAM Journal on Scientific Computing, 38 (2016), pp. A3318–A3356.
[25] K. Smetana and T. Taddei, *Localized model reduction for nonlinear elliptic partial differential equations: localized training, partition of unity, and adaptive enrichment*, arXiv preprint arXiv:2202.09872, (2022).
[26] T. Taddei, *An adaptive parametrized-background data-weak approach to variational data assimilation*, ESAIM: Mathematical Modelling and Numerical Analysis, 51 (2017), pp. 1827–1858.
[27] T. Taddei and A. T. Patera, *A Localization Strategy for Data Assimilation; Application to State Estimation and Parameter Estimation*, SIAM Journal on Scientific Computing, 40 (2018), pp. B611–B636. Publisher: Society for Industrial and Applied Mathematics.
[28] C. Taylor and P. Hood, *A numerical solution of the Navier-Stokes equations using the finite element technique*, Computers & Fluids, 1 (1973), pp. 73–100.
[29] I. E. Vignon-Clementel, C. A. Figueroa, K. E. Jansen, and C. A. Taylor, *Outflow boundary conditions for three-dimensional finite element modeling of blood flow and pressure in arteries*, Computer methods in applied mechanics and engineering, 195 (2006), pp. 3776–3796.
[30] S. Volkwein, *Model reduction using proper orthogonal decomposition*, Lecture Notes, Institute of Mathematics and Scientific Computing, University of Graz. see http://www. uni-graz. at/imawww/volkwein/POD. pdf, 1025 (2011).
[31] G. Wahba, *Spline models for observational data*, SIAM, 1990.