

Model reduction techniques for fast blood flow simulation in parametrized geometries

Andrea Manzoni^{‡,*}, Alfio Quarteroni^{‡,†}, Gianluigi Rozza[‡]

June 8, 2011

[‡] CMCS - Modelling and Scientific Computing,
MATHICSE - Mathematics Institute of Computational Science and Engineering,
EPFL - Ecole Polytechnique Fédérale de Lausanne,
Station 8, CH-1015 Lausanne, Switzerland.

[†] MOX - Modellistica e Calcolo Scientifico,
Dipartimento di Matematica “F. Brioschi”,
Politecnico di Milano,
via Bonardi 9, 20133 Milano, Italy.

Keywords: real-time simulation; model reduction; geometrical and computational reduction; reduced basis methods; radial basis functions; haemodynamics; Navier-Stokes equations

Abstract

In this paper we propose a new model reduction technique aimed at real-time blood flow simulations on a given family of geometrical shapes of arterial vessels. Our approach is based on the combination of a low-dimensional shape parametrization of the computational domain and the reduced basis method to solve the associated parametrized flow equations. We propose a preliminary analysis carried on a set of arterial vessel geometries, described by means of a radial basis functions parametrization. In order to account for patient-specific arterial configurations, we reconstruct the latter by solving a suitable parameter identification problem. Real-time simulation of blood flows are thus performed on each reconstructed parametrized geometry, by means of the reduced basis method. We focus on a family of parametrized carotid artery bifurcations, by modelling blood flows using Navier-Stokes equations and measuring distributed outputs such as viscous energy dissipation or vorticity. The latter are indexes that might be correlated with the assessment of pathological risks. The approach advocated here can be applied to a broad variety of (different) flow problems related with geometry/shape variation, for instance related with shape sensitivity analysis, parametric exploration, and shape design.

*Corresponding Author, E-mail: andrea.manzoni@epfl.ch

1 Introduction

Numerical simulation of blood flow problems have become quite popular in the past two decades, mainly because of *(i)* increasing computational power, *(ii)* progress in imaging and geometry extraction/reconstruction techniques [1] as well as to *(iii)* the availability of always more efficient numerical algorithms [2, 3, 4]. The driving factor behind this development is the awareness that numerical models can provide quantitative descriptions of blood behaviour in important vascular districts or in vessel networks, and to explain and assess the relationships between vessels shape, haemodynamics, and a family of clinical indicators. Among the latter we mention wall shear stresses, vorticity, viscous energy dissipations: they can be correlated to risk of failure in bypass grafting [5], or that of artery occlusion in presence of stenosis [6], or the one of aneurysm rupture [7]. Such kind of analyses can also help in understanding how different surgical solutions may affect blood circulation; to mention an example, shape optimization techniques may be used for designing better prosthetic devices [8, 9, 10]. Ideally, numerical simulations ought to be carried out in a very fast way, to rapidly provide a quantitative output/response on each new patient specific geometrical configuration. The goal is twofold, since we may be interested in either *(i)* a shape sensitivity analysis and evaluation of output related to blood flows or *(ii)* spanning a large family of different vessel configurations and simulating blood flows in simplified reconstructed geometries, where each configuration can be easily represented as a deformation of a reference configuration. Several examples dealing with the application of the first setting to the case of a family of carotid bifurcations are presented in [11, 12], while a recent application of the second framework for fast patient-specific simulations can be found in [13]. To afford the above computationally intensive tasks, pursuing computational reduction strategies can be determinant whenever interested in *real-time* simulations and/or repeated *output* evaluations for different values of some inputs of interest. A large number of model order reduction techniques have been proposed in the past few years, depending on the mathematical problems to be faced; basically, these techniques try to quickly capture, in an automatic way, the essential features, or modes, of a given process. Reduced basis methods [14, 15] as well as the proper orthogonal decomposition (or, equivalently, principal component analysis or, again, Karhunen-Loève transformation) technique [16] are two important examples in this field. Widely analyzed and developed in the last two decades, these techniques nowadays provide a rigorous framework for numerical simulations of complex systems arising not only in life sciences, but also in many other engineering contexts.

In this work, we apply the reduced basis method for the solution of parametrized partial differential equations (PDEs) to provide a preliminary approach for the real-time simulation of blood flows in a family of carotid artery bifurcations, in order to provide a shape sensitivity analysis of a specific output related

to blood flows, for example in presence of growing stenoses. We also consider the problem of geometrical reconstruction from data, ideally suited for the application of the proposed methodology in a patient-dependent framework. In order to achieve a strong geometrical reduction and describe the carotid configurations by using only a small number of significant parameters, we introduce a general class of shape parametrizations based on the radial basis function technique, which is a general interpolation method based on the choice and the displacement of a small set of control points [17]. In this way, it is possible to span a broad family of carotid configurations where either a localized variation (such as a stenosis in one of the branches or at the bifurcation) or a global shape can be described. Thanks to the reformulation of the original problem in a reference (parameter-independent) domain, reduced basis approximations of blood flows in different configurations are obtained by combining previously computed flow solutions, for a small set of properly selected parameter values. These basis solutions are computed and stored just once, during the so-called *offline* computational stage (a step to prepare a “database” of significant solutions); for each new value of the input parameters, corresponding to a different geometrical configuration, a new very small problem has to be solved during the so-called *online* stage, obtained by means of a Galerkin projection on the reduced basis space [14, 18]. This computational splitting, provided by suitable assumptions on differential operators and shape parametrization, entails great savings and real-time flow simulations by solving a very reduced linear system. Moreover, we discuss a (reduced version) of a shape reconstruction problem by introducing a suitable parameter identification problem, in order to identify the closest parametrized version of a given geometrical configuration. The interest of the approach presented in this work goes beyond the preliminary application here addressed. Indeed, our reduced model can be applied to a broad variety of problems for which real-time simulation in complex parametrized geometries is the key issue.

The structure of the paper is as follows. In Sect. 2 we introduce the mathematical model for the blood flows description, as well as the computational reduction framework for real-time simulation. In Sect. 3 we discuss a general class of shape parametrization techniques which allows to obtain quite an effective geometrical reduction, characterizing the case of the radial basis functions technique, while in Sect. 4 we briefly discuss a simple parameter identification problem for shape reconstruction through parametrized configurations. In Sect. 5 we introduce the general reduced basis framework for parametrized PDEs, focusing on the nonlinear steady Navier-Stokes case; in particular, both the certification of the methodology through a posteriori error estimations (to guarantee accuracy and reliability of results) and the *offline-online* stratagem for computational reduction are briefly discussed. Then, in Sect. 6 we introduce the problems related with the real-time description of blood flows in carotid artery bifurcations. In the end, we present some preliminary numerical results in Sect. 7, and discuss some possible improvements in Sect. 8.

2 Real time flow simulation for haemodynamics

It is recognized that correlation exists between haemodynamic forces and the location, development and morphology of plaque deposits within the lumen, which are responsible of the narrowing of blood vessels (or stenosis) and of serious cardiovascular diseases, such as atherosclerosis and intimal hyperplasia. Ruptures of the plaques may induce thrombus formation in the lumen, consequent detachment and occlusion of further smaller branches (thromboembolism), or even the complete closure of the lumen, implying chronic pathologies as ischemia, infarction or stroke (in the case of carotid arteries).

It is therefore interesting to analyze the sensitivity of physical outputs related to flow with respect to shape and shape variations, such as – for example – wall shear stresses, wall shear stress gradients, vorticity, dissipated energy; all these quantities are influenced by lumen geometry, characterize local haemodynamics effects and may assess a risk of artery occlusion. Rather than numerical simulations on a relatively small number of different configurations, it becomes crucial to explore a wide family of geometries – thus spanning a broad variety of shapes – in order to take into account their variability among patients and provide a more complete representation of blood flows and related outputs with respect to shape variation. Moreover, given a new geometry, we may be asked to obtain a numerical approximation of the flow in a very small amount of time, say order of one second, that can be considered as a real-time condition. Thus, we need to face both with a *real-time* problem (because of the rapid evaluation) and a *many-query* context (because of the need to span a large set of configuration), for which suitable model reduction techniques are requested, since full simulations may result very expensive if they have to be carried out for each new geometry.

With the sole aim of illustrating the use of low order methods for numerical simulation in parametrized domains, we adopt the steady nonlinear Navier-Stokes equations for modelling moderate Reynolds flows in mid/large size arteries [2], which in a parametrized domain $\Omega_o(\boldsymbol{\mu})$ read [19]:

$$\left\{ \begin{array}{ll} -\nu \Delta \tilde{\mathbf{u}}(\boldsymbol{\mu}) + \nabla p(\boldsymbol{\mu}) + (\tilde{\mathbf{u}}(\boldsymbol{\mu}) \cdot \nabla) \tilde{\mathbf{u}}(\boldsymbol{\mu}) = \mathbf{f} & \text{in } \Omega_o(\boldsymbol{\mu}) \\ \nabla \cdot \tilde{\mathbf{u}}(\boldsymbol{\mu}) = 0 & \text{in } \Omega_o(\boldsymbol{\mu}) \\ \tilde{\mathbf{u}}(\boldsymbol{\mu}) = \mathbf{0} & \text{on } \Gamma_w^o \\ \tilde{\mathbf{u}}(\boldsymbol{\mu}) = \mathbf{g}^D & \text{on } \Gamma_{in}^o \\ -p(\boldsymbol{\mu}) \mathbf{n} + \nu \frac{\partial \tilde{\mathbf{u}}(\boldsymbol{\mu})}{\partial \mathbf{n}} = \mathbf{g}^N & \text{on } \Gamma_{out}^o; \end{array} \right. \quad (1)$$

$\boldsymbol{\mu} \in \mathcal{D} \subset \mathbb{R}^p$ is a vector of *input* parameters, $(\tilde{\mathbf{u}}(\boldsymbol{\mu}), p(\boldsymbol{\mu}))$ are the velocity and the pressure fields, for some given \mathbf{f} , \mathbf{g}^D , \mathbf{g}^N ; \mathbf{n}_o is the normal unit vector on the boundary $\partial\Omega_o(\boldsymbol{\mu}) = \Gamma_{in}^o \cup \Gamma_{out}^o \cup \Gamma_w^o$, where Γ_{in}^o is the inflow boundary, Γ_{out}^o the outflow boundary and Γ_w^o the boundary corresponding to the arterial wall. Nevertheless, further extensions of the methodology addressed in this work to

time-dependent problems – to take into account blood flow pulsatility – and treatment of fluid-structure interaction are optimistically foreseen in view of an effective reduction for nonlinear problems.

The weak formulation of problem (1) can be obtained by introducing the functional spaces $\mathcal{X}(\boldsymbol{\mu}) = (H_{0,\Gamma_D}^1(\Omega_o(\boldsymbol{\mu})))^2$ and $\mathcal{Q}(\boldsymbol{\mu}) = L^2(\Omega_o(\boldsymbol{\mu}))$ for velocity and pressure, respectively, being $\Gamma_D = \Gamma_{in}^o \cup \Gamma_w^o$, $H_{0,\Gamma_D}^1(\Omega) = \{v \in H^1(\Omega) : v|_{\Gamma_D} = 0\}$, and integrating by parts the expression obtained by multiplying (1) by a test function $\mathbf{w} \in \mathcal{X}(\boldsymbol{\mu})$. We introduce a lifting function $R_o \in (H^1(\Omega_o(\boldsymbol{\mu})))^2$ such that $R_o|_{\Gamma_{in}^o} = \mathbf{g}^D$ and denote $\mathbf{u} = \tilde{\mathbf{u}} - R_o$, so that $\mathbf{u}|_{\Gamma_D} = \mathbf{0}$ and $\nabla \mathbf{u} = \nabla \tilde{\mathbf{u}} - \nabla R_o$. Then, denoting $\mathcal{Y}(\Omega_o(\boldsymbol{\mu})) = \mathcal{X}(\boldsymbol{\mu}) \times \mathcal{Q}(\boldsymbol{\mu})$ and $W = (\mathbf{w}, q)$, the weak formulation reads: find $Y(\boldsymbol{\mu}) = (\mathbf{u}(\boldsymbol{\mu}), p(\boldsymbol{\mu})) \in \mathcal{Y}(\Omega_o(\boldsymbol{\mu}))$ s.t.

$$\mathcal{A}_o(Y(\boldsymbol{\mu}), W; \boldsymbol{\mu}) + \mathcal{C}_o(Y(\boldsymbol{\mu}), Y(\boldsymbol{\mu}), W; \boldsymbol{\mu}) = \mathcal{F}_o(W; \boldsymbol{\mu}), \quad (2)$$

where

$$\mathcal{A}_o(Y, W; \boldsymbol{\mu}) = a_o(\mathbf{u}, \mathbf{w}; \boldsymbol{\mu}) + d_o(\mathbf{u}, \mathbf{w}; \boldsymbol{\mu}) + b_o(p, \mathbf{w}; \boldsymbol{\mu}) + b_o(q, \mathbf{u}; \boldsymbol{\mu}), \quad (3)$$

$$\mathcal{C}_o(Y, Y, W; \boldsymbol{\mu}) = c_o(\mathbf{u}, \mathbf{u}, \mathbf{w}; \boldsymbol{\mu}), \quad (4)$$

$$\mathcal{F}_o(W; \boldsymbol{\mu}) = \langle F_o^s, \mathbf{w} \rangle + \langle F_o^l, \mathbf{w} \rangle + \langle G_o^l, q \rangle; \quad (5)$$

the bilinear forms corresponding to viscous and pressure terms are given by

$$a_o(\mathbf{v}, \mathbf{w}; \boldsymbol{\mu}) = \int_{\Omega_o(\boldsymbol{\mu})} \frac{\partial \mathbf{v}}{\partial x_{o,i}} \nu_{ij}^o \frac{\partial \mathbf{w}}{\partial x_{o,j}} d\Omega_o, \quad b_o(q, \mathbf{w}; \boldsymbol{\mu}) = - \int_{\Omega_o(\boldsymbol{\mu})} q \chi_{ij}^o \frac{\partial \mathbf{w}_i}{\partial x_{o,j}} d\Omega_o,$$

where summation over i and j is understood, $\nu_{ij}^o = \nu \delta_{ij}$, $\chi_{ij}^o = \delta_{ij}$ and δ_{ij} indicates the Kronecker symbol; the trilinear form corresponding to the convective term is given by

$$c_o(\mathbf{u}, \mathbf{v}, \mathbf{w}; \boldsymbol{\mu}) = \int_{\Omega_o(\boldsymbol{\mu})} u_i \eta_{ij}^o \frac{\partial v_k}{\partial x_{o,j}} w_k d\Omega_o,$$

being $\eta_{ij}^o = \chi_{ij}^o$; $\langle F_o^s, \cdot \rangle$ is a source term given by

$$\langle F_o^s, \mathbf{w} \rangle = \int_{\Omega_o(\boldsymbol{\mu})} \mathbf{f} \cdot \mathbf{w} d\Omega_o + \int_{\Gamma_{out}^o} \mathbf{g}^N \cdot \mathbf{w} d\Gamma_o,$$

while $d_o(\cdot, \cdot; \boldsymbol{\mu})$, F_l^o and G_l^o are terms due to non-homogeneous Dirichlet boundary conditions:

$$d_o(\mathbf{v}, \mathbf{w}; \boldsymbol{\mu}) = c_o(L_o \mathbf{g}^D, \mathbf{v}, \mathbf{w}) + c_o(\mathbf{v}, L_o \mathbf{g}^D, \mathbf{w}),$$

$$\langle F_l^o, \mathbf{w} \rangle = -a_o(L_o \mathbf{g}^D, \mathbf{w}) - c_o(L_o \mathbf{g}^D, L_o \mathbf{g}^D, \mathbf{w}), \quad \langle G_l^o, q \rangle = -b_o(q, L_o \mathbf{g}^D),$$

is a source term. Moreover, we are interested in the evaluation of a physical output depending on the field variables, say

$$s_o(\boldsymbol{\mu}) = \int_{\Omega_o(\boldsymbol{\mu})} Q(Y(\boldsymbol{\mu})) d\Omega_o,$$

being $Q = Q(\cdot)$ a quadratic function of velocity and/or pressure variables. In particular, we will focus on the viscous energy dissipation, for which we have $Q(Y(\boldsymbol{\mu})) = \nu |\nabla \mathbf{u}(\boldsymbol{\mu})|^2$, and then

$$s_o(\boldsymbol{\mu}) = \int_{\Omega_o(\boldsymbol{\mu})} \nu |\nabla \mathbf{u}(\boldsymbol{\mu})|^2 d\Omega_o \equiv a_o(\mathbf{u}(\boldsymbol{\mu}), \mathbf{u}(\boldsymbol{\mu}); \boldsymbol{\mu}).$$

2.1 Parametrized formulation of Navier-Stokes equations

Our idea is to describe a given family of geometrical configurations by means of a suitable shape parametrization; then, we will carry out full flow simulation only on a small number of configurations – corresponding to the basis solutions – and exploit the reduced basis method in order to perform real-time reduced simulations for any new configuration. In order to combine the basis solutions, we need to introduce a reference (parameter-independent) domain Ω , on which ultimately all the computations are performed (and compared). To do this, problem (1)-(5) has to be traced back to the reference domain Ω , which is related to the original domain $\Omega_o(\boldsymbol{\mu})$ through a parametric mapping $T(\cdot; \boldsymbol{\mu})$; its construction will be discussed in Sect. 3. By change of variables, we obtain the parametrized formulation of the Navier-Stokes problem (2): find $Y(\boldsymbol{\mu}) = (\mathbf{u}(\boldsymbol{\mu}), p(\boldsymbol{\mu})) \in \mathcal{Y}$ s.t.

$$\mathcal{A}(Y(\boldsymbol{\mu}), W; \boldsymbol{\mu}) + \mathcal{C}(Y(\boldsymbol{\mu}), Y(\boldsymbol{\mu}), W; \boldsymbol{\mu}) = \mathcal{F}(W; \boldsymbol{\mu}), \quad \forall W \in \mathcal{Y}. \quad (6)$$

Here $\mathcal{Y} = \mathcal{X} \times \mathcal{Q}$, being $\mathcal{X} = (H_{0,\Gamma_D}^1(\Omega))^2$ and $\mathcal{Q} = L^2(\Omega)$;

$$\mathcal{A}(Y, W; \boldsymbol{\mu}) = a(\mathbf{u}, \mathbf{w}; \boldsymbol{\mu}) + d(\mathbf{u}, \mathbf{w}; \boldsymbol{\mu}) + b(p, \mathbf{w}; \boldsymbol{\mu}) + b(q, \mathbf{w}; \boldsymbol{\mu}) + c(\mathbf{u}, \mathbf{u}, \mathbf{w}; \boldsymbol{\mu}), \quad (7)$$

$$\mathcal{C}(Y, Y, W; \boldsymbol{\mu}) = c(\mathbf{u}, \mathbf{u}, \mathbf{w}; \boldsymbol{\mu}), \quad (8)$$

$$\mathcal{F}(W; \boldsymbol{\mu}) = \langle F^s(\boldsymbol{\mu}), \mathbf{w} \rangle + \langle F^l(\boldsymbol{\mu}), \mathbf{w} \rangle + \langle G^l(\boldsymbol{\mu}), q \rangle. \quad (9)$$

In this case,

$$a(\mathbf{v}, \mathbf{w}; \boldsymbol{\mu}) = \int_{\Omega} \frac{\partial \mathbf{v}}{\partial x_i} \nu_{ij}(\mathbf{x}, \boldsymbol{\mu}) \frac{\partial \mathbf{w}}{\partial x_j} d\Omega, \quad \nu(\mathbf{x}, \boldsymbol{\mu}) = \mathbf{J}_T^{-1} \boldsymbol{\nu}^o \mathbf{J}_T^{-T} |\mathbf{J}_T|, \quad (10)$$

$$b(p, \mathbf{w}; \boldsymbol{\mu}) = - \int_{\Omega} p \chi_{ij}(\mathbf{x}, \boldsymbol{\mu}) \frac{\partial w_j}{\partial x_i} d\Omega, \quad \chi(\mathbf{x}, \boldsymbol{\mu}) = \mathbf{J}_T^{-1} |\mathbf{J}_T|, \quad (11)$$

$$c(\mathbf{u}, \mathbf{v}, \mathbf{w}; \boldsymbol{\mu}) = \int_{\Omega} u_i \eta_{ij}(\mathbf{x}, \boldsymbol{\mu}) \frac{\partial v_k}{\partial x_j} w_k d\Omega, \quad \eta(\mathbf{x}, \boldsymbol{\mu}) \equiv \chi(\mathbf{x}, \boldsymbol{\mu}) = \mathbf{J}_T^{-1} |\mathbf{J}_T|, \quad (12)$$

being $\mathbf{J}_T = \mathbf{J}_T(\mathbf{x}, \boldsymbol{\mu})$ the Jacobian of $T(\mathbf{x}, \boldsymbol{\mu})$ and $|\mathbf{J}_T|$ its determinant. In the same way, we have

$$\langle F^s(\boldsymbol{\mu}), \mathbf{w} \rangle = \int_{\Omega} \mathbf{f} \cdot \mathbf{w} |\mathbf{J}_T| d\Omega + \int_{\Gamma_{out}} \mathbf{g}^N \cdot \mathbf{w} |\mathbf{J}_T \mathbf{t}| d\Gamma,$$

while $d(\cdot, \cdot; \boldsymbol{\mu})$, $F_l(\boldsymbol{\mu})$ and $G_l(\boldsymbol{\mu})$ are terms due to non-homogeneous Dirichlet boundary conditions:

$$d(\mathbf{v}, \mathbf{w}; \boldsymbol{\mu}) = c(L\mathbf{g}^D, \mathbf{v}, \mathbf{w}; \boldsymbol{\mu}) + c(\mathbf{v}, L\mathbf{g}^D, \mathbf{w}; \boldsymbol{\mu}),$$

$$\langle F_l(\boldsymbol{\mu}), \mathbf{w} \rangle = -a(L\mathbf{g}^D, \mathbf{w}; \boldsymbol{\mu}) - c(L\mathbf{g}^D, L\mathbf{g}^D, \mathbf{w}; \boldsymbol{\mu}), \quad \langle G_l(\boldsymbol{\mu}), q \rangle = -b_o(q, L\mathbf{g}^D; \boldsymbol{\mu});$$

as before, $L\mathbf{g}_D \in (H^1(\Omega))^2$ is a lifting function such that $L\mathbf{g}_D|_{\Gamma_D} = \mathbf{g}^D$, while \mathbf{t} denotes the unit tangential vector to $\partial\Omega$. In the end, the parametrized form of the output of interest is given by

$$s(\boldsymbol{\mu}) = \int_{\Omega} Q \circ T(Y(\boldsymbol{\mu})) |J_T| d\Omega, \quad (13)$$

being $Y(\boldsymbol{\mu})$ the solution of (6). In particular, for the case of the energy dissipation, $s(\boldsymbol{\mu}) = a(\mathbf{u}(\boldsymbol{\mu}), \mathbf{u}(\boldsymbol{\mu}); \boldsymbol{\mu})$.

We briefly remind the conditions for the well-posedness of problem (6) and the existence of nonsingular branches of solutions [20, 21]. Let us endow \mathcal{Y} by the scalar product $(\cdot, \cdot)_{\mathcal{Y}} = (\cdot, \cdot)_{\mathcal{X}} + (\cdot, \cdot)_{\mathcal{Q}}$, where $(\mathbf{u}, \mathbf{v})_{\mathcal{X}} = (\nabla \mathbf{u}, \nabla \mathbf{v})_{(L^2(\Omega))^2}$, $(p, q)_{\mathcal{Q}} = (p, q)_{L^2(\Omega)}$ for all $\mathbf{u}, \mathbf{v} \in \mathcal{X}$, $p, q \in \mathcal{Q}$; the graph norm of $Y = (\mathbf{u}, p)$ is given by $\|Y\|_{\mathcal{Y}}^2 = \|\mathbf{u}\|_{\mathcal{X}}^2 + \|p\|_{\mathcal{Q}}^2$, where $\|\mathbf{u}\|_{\mathcal{X}} = (\mathbf{u}, \mathbf{u})_{\mathcal{X}}^{1/2}$, $\|p\|_{\mathcal{Q}} = (p, p)_{\mathcal{Q}}^{1/2}$. Moreover, defining the Navier-Stokes operator $\mathcal{G} : \mathcal{Y} \rightarrow \mathcal{Y}'$ as

$$\mathcal{G}(Y(\boldsymbol{\mu}), V; \boldsymbol{\mu}) = \mathcal{A}(Y(\boldsymbol{\mu}), V; \boldsymbol{\mu}) + \mathcal{C}(Y(\boldsymbol{\mu}), Y(\boldsymbol{\mu}), V; \boldsymbol{\mu}), \quad (14)$$

its Fréchet derivative (about W) with respect to the first variable Y is given by

$$d\mathcal{G}(W; \boldsymbol{\mu})(Y, V) = \mathcal{A}(Y, V; \boldsymbol{\mu}) + \mathcal{C}(Y, W, V; \boldsymbol{\mu}) + \mathcal{C}(W, Y, V; \boldsymbol{\mu}), \quad \forall \boldsymbol{\mu} \in \mathcal{D}.$$

We then associate to $d\mathcal{G}(W; \boldsymbol{\mu})(\cdot, \cdot)$ the Babuška inf-sup and continuity “constants”

$$\beta(Y(\boldsymbol{\mu})) = \inf_{W \in \mathcal{Y}} \sup_{V \in \mathcal{Y}} \frac{d\mathcal{G}(Y(\boldsymbol{\mu}); \boldsymbol{\mu})(W, V)}{\|W\|_{\mathcal{Y}} \|V\|_{\mathcal{Y}}},$$

$$\gamma(Y(\boldsymbol{\mu})) = \sup_{W \in \mathcal{Y}} \sup_{V \in \mathcal{Y}} \frac{d\mathcal{G}(Y(\boldsymbol{\mu}); \boldsymbol{\mu})(W, V)}{\|W\|_{\mathcal{Y}} \|V\|_{\mathcal{Y}}};$$

under the following stability and continuity conditions [22, 23],

$$\beta(Y(\boldsymbol{\mu})) \geq \beta_0 > 0 \quad \gamma(Y(\boldsymbol{\mu})) < +\infty, \quad \forall \boldsymbol{\mu} \in \mathcal{D},$$

in the neighborhood of $Y(\boldsymbol{\mu})$ the solution of (6) is unique. We can also verify that (for homogeneous Dirichlet conditions) $\gamma(Y(\boldsymbol{\mu})) \leq \gamma_a(\boldsymbol{\mu}) + \gamma_b(\boldsymbol{\mu})(2 + \rho^2 \|Y(\boldsymbol{\mu})\|_{\mathcal{Y}})$, being $\gamma_a(\boldsymbol{\mu})$ and $\gamma_b(\boldsymbol{\mu})$ the continuity constants of $a(\cdot, \cdot; \boldsymbol{\mu})$ and $b(\cdot, \cdot; \boldsymbol{\mu})$, respectively, and

$$\rho = \sqrt{2} \sup_{\mathbf{v} \in \mathcal{X}} \|\mathbf{v}\|_{(L^4(\Omega))^2} / \|\mathbf{v}\|_{\mathcal{X}} \quad (15)$$

a Sobolev embedding constant [24], where $\|\mathbf{v}\|_{(L^p(\Omega))^2} = (\int_{\Omega} (v_i v_i)^{p/2})^{1/p}$.

3 Radial basis functions techniques for shape parametrization

Efficient shape representation and the description of its deformation is an issue arising in computational geometry, e.g. in image-based segmentation, tracking or registration, shape reconstruction. Our goal is to set up a mathematical model which is flexible – in order to represent complex configurations – and low-dimensional – i.e. depending on a small number of input parameters. Free shape representations, obtained by introducing a small set of control points – and possibly some interpolation strategies – are a good compromise between flexibility and low-dimensionality and they can be properly coupled with the reduced basis method for parametrized PDEs [25, 9]. In particular, control points can be located either on a regular lattice or scattered in the domain, and their displacements induce a local deformation of the domain. In this way, a global deformation mapping $T(\cdot; \boldsymbol{\mu})$ is obtained as a linear combination of the control points displacements, treated as input parameters. Two strategies in this class are the Free-Form Deformations (FFD) technique and the method based on Radial Basis Functions (RBF).

Free-Form Deformations [26, 27] are obtained as the tensor product of deformations performed along each direction on a regular lattice of control points. Usually, only small subsets of these deformations are selected (by means of problem-dependent criteria) as input parameters if we want a geometrical reduction; in this way, each parameter μ_1, \dots, μ_p identifies a selected directional displacement of a corresponding control point. This technique has been widely used in computer graphics and also applied to optimal design and shape optimization problems, thanks to its flexibility and easiness of handling (see for example [28, 9] for two applications in the reduced basis context). However, FFD suffer from some limitations, since *(i)* the control points cannot be chosen freely but must be the nodes of a lattice *(ii)* it is not possible to perform a boundary control and *(iii)* the process is not interpolatory. In particular, using a rectangular lattice to describe deformations of irregular or complex shapes makes the choice of control points crucially important; for example, control points located far from the boundary to be optimized have less influence. Choosing the subset of active points is therefore a critical issue, highly problem-dependent.

In order to overcome these limitations, we introduce a different shape parametrization technique, still based on a set of control points, but which is interpolatory: the Radial Basis Function (RBF) technique. For a general introduction see [17, 29], while applications to shape optimization and mesh deformation are described in [30]. This technique is used for nonlinear interpolation (e.g. of scattered data); in the two-dimensional case, a map $\tau : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ is defined as

follows:

$$\tau(\mathbf{x}) = P(\mathbf{x}) + \sum_{i=1}^k w_i \sigma(\|\mathbf{x} - \mathbf{X}_i\|), \quad (16)$$

being $P(\cdot)$ a low-degree polynomial function, $\{\mathbf{w}_i\}_{i=1}^k$, $\mathbf{w}_i \in \mathbb{R}^2$, a set of weights corresponding to the (a-priori selected) k control points, whose reference positions are $[\mathbf{X}_1, \dots, \mathbf{X}_k]$, a set of distinct points in \mathbb{R}^2 , and $\sigma(\cdot)$ a (translated) radially symmetric function. Common choices introduced for modelling 2D (or 3D) shapes are, for example,

$$\sigma(h) = \begin{cases} \exp(-h^2/\sigma^2) & \text{Gaussian,} \\ (h^2 + \gamma^2)^{1/2} & \text{multiquadratic,} \\ h^\gamma & \text{power, } \gamma = 1, 3 \\ h^2 \log(h) & \text{thin-plate splines.} \end{cases}$$

The asymptotic behaviour of $\sigma(h)$ is different among the selected basis functions, and the choice on the various possibilities is performed according to shape regularity and to convergence properties of the numerical method used to compute the coefficients appearing in (16). For example, the linear RBFs exhibit better convergence properties, while the cubic RBFs guarantee an enhanced shape smoothness. Usually, a polynomial function of degree 1 is chosen, $P(\mathbf{x}) = \mathbf{c} + \mathbb{A}\mathbf{x}$, being $\mathbf{c} \in \mathbb{R}^2$ and $\mathbb{A} \in \mathbb{R}^{2 \times 2}$; thus, the map can be rewritten as

$$\tau(\mathbf{x}) = \mathbf{c} + \mathbb{A}\mathbf{x} + \mathbb{W}^T s(\mathbf{x}), \quad (17)$$

being $s(\mathbf{x}) = (\sigma(\|\mathbf{x} - \mathbf{X}_1\|), \dots, \sigma(\|\mathbf{x} - \mathbf{X}_k\|))^T$ and $\mathbb{W} = [\mathbf{w}_1, \dots, \mathbf{w}_k]^T \in \mathbb{R}^{k \times 2}$. In this way, $P(\cdot)$ is the affine part (rotation and/or scaling), while the term $\mathbb{W}^T s(\mathbf{x})$ depending on the control points adds a nonaffine contribution to the deformation. The map (17) is function of $2k+6$ coefficients in the two-dimensional case; to compute them we look for a transformation such that each control point of the initial (or reference) shape is mapped onto the corresponding control point of the target (or deformed) one. Introducing the initial position $\mathbb{X} = [\mathbf{X}_1, \dots, \mathbf{X}_k]^T \in \mathbb{R}^{k \times 2}$ and the deformed position $\mathbb{Y} = [\mathbf{Y}_1, \dots, \mathbf{Y}_k]^T \in \mathbb{R}^{k \times 2}$ of the control points, the weights $\{\mathbf{w}_i\}_{i=1}^k$ in (16) are found by fulfilling the interpolation constraints

$$\tau(\mathbf{X}_i) = \mathbf{Y}_i, \quad \forall i = 1, \dots, k, \quad (18)$$

When a polynomial term $P(\mathbf{x}) = \mathbf{c} + \mathbb{A}\mathbf{x}$ is included, the system is completed by the additional constraint [17, 29]

$$\sum_{i=1}^k \mathbf{w}_i = \mathbf{0} \quad \sum_{i=1}^k X_{i1} \mathbf{w}_i = \sum_{i=1}^k X_{i2} \mathbf{w}_i = \mathbf{0}, \quad (19)$$

being $\mathbf{X}_j = (X_{j1}, X_{j2})^T$, which can represent the conservation of total force and momentum [30]. More in general, if $P(\mathbf{x}) \in \mathbb{P}_q = \text{span}\{p_1, \dots, p_Q\}$, space of all

polynomials of degree up to $q \geq 1$ in 2 unknowns, the side constraints can be expressed as [17]

$$\sum_{i=1}^k \mathbf{w}_i p_l(\mathbf{X}_i) = \mathbf{0} \quad \forall l = 1, \dots, Q = \dim(\mathbb{P}_q).$$

In order to fit the RBF technique in our parametrized framework, let us express the deformed positions $[\mathbf{Y}_1, \dots, \mathbf{Y}_k]^T$ of the control points as

$$\mathbf{Y}_i(\boldsymbol{\mu}_i) = \mathbf{X}_i + \boldsymbol{\mu}_i, \quad i = 1, \dots, k,$$

being $\{\boldsymbol{\mu}_i = (\mu_{i1}, \mu_{i2})\}_{i=1}^k$ the displacements of the k control points, which are usually chosen on the boundary of the shape which has to be deformed (see Fig. 1). The input parameters $\boldsymbol{\mu} = (\mu_1, \dots, \mu_p) \in \mathbb{R}^p$ are then a subset of $p \leq 2k$ degrees of freedom, chosen accordingly to some problem-dependent criteria. Thus, the parametric mapping $T(\cdot; \boldsymbol{\mu}) : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ will be given by

$$T(\mathbf{x}; \boldsymbol{\mu}) = \mathbf{c}(\boldsymbol{\mu}) + \mathbb{A}(\boldsymbol{\mu})\mathbf{x} + \mathbb{W}(\boldsymbol{\mu})^T s(\mathbf{x}), \quad (20)$$

where the coefficients $\mathbf{c}(\boldsymbol{\mu})$, $\mathbb{A}(\boldsymbol{\mu})$, $\mathbb{W}(\boldsymbol{\mu})$ satisfy the constraints (18)-(19). By denoting \mathbb{S} the interpolation matrix whose elements are $\mathbb{S}_{ij} = s_i(\mathbf{X}_j) = \sigma(\|\mathbf{X}_j - \mathbf{X}_i\|)$ and $\mathbb{I}_k = [1, 1, \dots, 1]^T \in \mathbb{R}^k$, we can rewrite (18)-(19) in a compact form, where the coefficients are obtained from the solution of a linear system, such that the parametrization is residing only on the right-hand-side:

$$\begin{bmatrix} \mathbb{S} & \mathbb{I}_k & \mathbb{X} \\ \mathbb{I}_k^T & \mathbb{O} & \mathbb{O} \\ \mathbb{X}^T & \mathbb{O} & \mathbb{O} \end{bmatrix} \begin{bmatrix} \mathbf{W} \\ \mathbf{c}^T \\ \mathbb{A}^T \end{bmatrix} = \begin{bmatrix} \mathbf{Y}(\boldsymbol{\mu}) \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}.$$

By adding a polynomial function $P(\mathbf{x})$ in (16) the interpolation problem always admits a unique solution, since the global matrix appearing in (3) is symmetric and positive definite^a. For a small number of control points, as in our approach, linear system (3) can be efficiently solved by a suitable direct method (matrix factorization is not depending on the parameters). Nevertheless, since it may be badly conditioned and non-sparse (because of the global character of RBF), difficulties may arise especially when using a large number of control points, as for example in fluid-structure interaction coupled problems or, more generally, when dealing with mesh motion through RBF. In these cases, suitable scaling or preconditioning strategies may help, as discussed for example in [17].

Once the map $T(\cdot; \boldsymbol{\mu})$ has been built, parametrized tensors $\mathbf{J}_T(\boldsymbol{\mu})$, $\boldsymbol{\nu}(\boldsymbol{\mu})$ and $\boldsymbol{\chi}(\boldsymbol{\mu}) \equiv \boldsymbol{\eta}(\boldsymbol{\mu})$ appearing in (10)-(12) are computed symbolically by means of a Computer Algebra System.

^aThe interpolation matrix \mathbb{S} is in general symmetric but positive semidefinite; it is positive definite if $\sigma(r) = \phi(r^2)$, being $\phi : \mathbb{R}_+ \rightarrow \mathbb{R}$ a continuous completely monotonic function, i.e. such that $(-1)^l \phi^{(l)}(t) \geq 0$ for all positive t . By imposing that the only polynomial vanishing on the control points is zero we obtain a RBF (conditionally) positive definite [17].

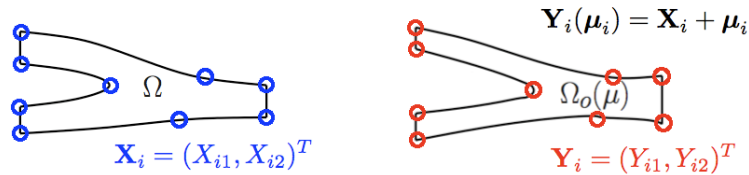


Figure 1: Schematic diagram of the RBF technique: on the left, the reference (or initial) configuration Ω and the unperturbed control points \mathbf{X}_i , $i = 1, \dots, k$; on the right, the deformed (or target) configuration $\Omega_o(\boldsymbol{\mu})$ and the displaced control points \mathbf{Y}_i , $i = 1, \dots, k$. *A priori* each control point \mathbf{X}_i can be translated by a vector $\boldsymbol{\mu}_i \in \mathbb{R}^2$; actually, only a subset of the $2k$ possible degrees of freedom are chosen as *input* parameters.

4 Parameter identification for shape reconstruction

Shape reconstruction consists of recovering a transformation that establishes some desired correspondences between two geometrical configurations, according to some similarity/distance measures. This process, and more generally, representation of shapes and their deformations, is central in many applications of image processing and medical image analysis, such as segmentation (extraction of geometrical structures from images), tracking (recovering the temporal evolution from consecutive frames), or registration. Our (limited) ambition is to set a (much simpler) preliminary procedure in order to reconstruct a parametrized version of a given shape configuration, by means of the RBF technique discussed in the previous section, on which we aim to perform a flow simulation. In order to be compatible with the reduced real-time framework, the reconstruction procedure has to be *(i)* based on a small number of parameters and *(ii)* performed in a very small amount of time. Let us denote \mathcal{S} the initial shape which has to be deformed in order to reconstruct the target shape \mathcal{T}_d by means of the parametric map $T(\cdot; \boldsymbol{\mu})$. Moreover, let us introduce a shape representation $\mathcal{R}(\cdot)$, a distance (or similarity/dissimilarity metric) $d(\cdot)$ defined in the space of representations and a loss function $\rho(\cdot)$. The reconstruction process can be in general expressed as a minimization problem:

$$\hat{\boldsymbol{\mu}} = \arg \min_{\boldsymbol{\mu} \in \mathcal{D}_{ad}} \rho(d(\mathcal{R}(T(\mathcal{S}; \boldsymbol{\mu})), \mathcal{R}(\mathcal{T}_d))) + \beta \omega(\boldsymbol{\mu}) \quad (21)$$

where $\omega(\boldsymbol{\mu})$ is a suitable regularization term and $\beta > 0$ is a weighting parameter. A usual choice is $\rho(s) = s^2$ for the loss function and the Euclidean distance for $d(\mathbf{x})$; more complex transformations or weighting can be introduced in order to improve the robustness of the matching. The most difficult task is the choice of the shape representation $\mathcal{R}(\cdot)$. Many strategies have been developed [31, 32]. In view of model reduction, we represent a shape configuration by means of a set of N_r *registration points* (or landmarks) $\{\mathbf{x}_j^r\}_{j=1}^{N_r} \in \mathcal{S}$, such that $\mathcal{R}(T(\mathcal{S}; \boldsymbol{\mu})) = T(\mathbf{x}_j^r; \boldsymbol{\mu})$, $r = 1, \dots, N_r$; in the same way, we suppose to know the target position $\mathcal{R}(\mathcal{T}_d) = \{\mathbf{y}_j^r\}_{j=1}^{N_r}$ of these registration points in the target configuration \mathcal{T}_d .

Moreover, we introduce a set of control points $\{\mathbf{X}_i\}_{i=1}^{N_C}$ on the initial shape in order to define a parametric RBF mapping $T(\cdot; \boldsymbol{\mu})$, whose image is given by $\{\mathbf{Y}_i\}_{i=1}^{N_C}$ (see Fig. 2 for a summarizing scheme). The idea is that, since the RBF mapping may fail in matching the two configurations, the registration points can be used to drive the mapping in order to enhance its fitting capabilities. In this way, the parameter identification problem (21) becomes a matching problem between two point-sets, that can be written as a *least squares* minimization problem

$$\arg \min_{\boldsymbol{\mu} \in \mathcal{D}_{ad}} J(\boldsymbol{\mu}) = \sum_{j=1}^{N_r} \|T(\mathbf{x}_j^r; \boldsymbol{\mu}) - \mathbf{y}_j^r\|^2 + \beta \sum_{i=1}^{N_c} \|T(\mathbf{X}_i^c; \boldsymbol{\mu}) - \mathbf{Y}_i^c\|^2, \quad (22)$$

where the regularization term (a quadratic function of $\boldsymbol{\mu}$) is given by the distance between the images $\{T(\mathbf{X}_i^c; \boldsymbol{\mu})\}$ of the control points and their target positions $\{\mathbf{Y}_i^c\}_{i=1}^{N_C}$. Note that the introduction of the registration points does not increase the problem dimension, as the number of input parameters of shape representation remains unchanged. Regarding instead the control points, they need to be scattered all over the domain in order to describe a wide family of shapes and global deformations.

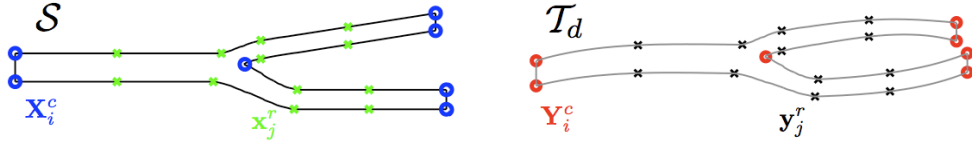


Figure 2: Schematic diagram of the shape reconstruction: on the left, the reference (or initial) configuration \mathcal{S} , the unperturbed control points \mathbf{X}_i^c , $i = 1, \dots, N_c$ (in blue), and the unperturbed registration points \mathbf{x}_j^r , $i = 1, \dots, N_r$ (in green); on the right, the target configuration \mathcal{T}_d , the target control points \mathbf{Y}_i^c , $i = 1, \dots, N_c$ (in red), and the target registration points \mathbf{y}_j^r , $i = 1, \dots, N_r$ (in black).

5 Reduced basis solution of parametrized Navier-Stokes equations

The general theory of Brezzi, Rappaz, Raviart [20, 21, 33] for approximating non singular branches of solution of the steady Navier-Stokes equations provides the framework for the analysis of the reduced basis approximation, as well as a tool to derive a suitable a posteriori error estimation. In that theory, the solution was parametrized w.r.t. the (inverse of the) Reynolds number, whereas here the parametrization is w.r.t. the parameters $\boldsymbol{\mu}$ used for shape deformation. We provide here a quick overview of these issues; deeper analyses can be found in [34, 23, 35, 36, 37].

Our starting point is a *truth* approximation $Y^{\mathcal{N}}(\boldsymbol{\mu})$ of problem (6), which can be obtained by standard Galerkin-Finite Element (FE) Method: find $Y^{\mathcal{N}}(\boldsymbol{\mu}) = (\mathbf{u}^{\mathcal{N}}(\boldsymbol{\mu}), p^{\mathcal{N}}(\boldsymbol{\mu})) \in \mathcal{Y}^{\mathcal{N}}$ s.t.

$$\mathcal{A}(Y^{\mathcal{N}}(\boldsymbol{\mu}), W; \boldsymbol{\mu}) + \mathcal{C}(Y^{\mathcal{N}}(\boldsymbol{\mu}), Y^{\mathcal{N}}(\boldsymbol{\mu}), W; \boldsymbol{\mu}) = \mathcal{F}(W; \boldsymbol{\mu}), \quad \forall W \in \mathcal{Y}^{\mathcal{N}}. \quad (23)$$

Here $\mathcal{Y}^{\mathcal{N}} = \mathcal{X}^{\mathcal{N}} \times \mathcal{Q}^{\mathcal{N}}$, being $\mathcal{X}^{\mathcal{N}} \subset \mathcal{X}$, $\mathcal{Q}^{\mathcal{N}} \subset \mathcal{Q}$ two sequences of FE approximation spaces of global dimension $\mathcal{N} = \mathcal{N}_{\mathcal{X}} + \mathcal{N}_{\mathcal{Q}}$; thus, \mathcal{N} will represent a measure of the computational complexity in the offline stage. Moreover, we assume that the difference $\|Y^{\mathcal{N}}(\boldsymbol{\mu}) - Y(\boldsymbol{\mu})\|_{\mathcal{Y}}$ is so small, so that $Y^{\mathcal{N}}(\boldsymbol{\mu})$ can be effectively considered as a “truth” approximation (at least for engineering purposes).

The matrix form corresponding to the FE discretization (23) is given by

$$\begin{bmatrix} A_{\mathcal{N}}(\boldsymbol{\mu}) + D_{\mathcal{N}}(\boldsymbol{\mu}) + C_{\mathcal{N}}(\mathbf{u}_{\mathcal{N}}(\boldsymbol{\mu}); \boldsymbol{\mu}) & B_{\mathcal{N}}^T(\boldsymbol{\mu}) \\ B_{\mathcal{N}}(\boldsymbol{\mu}) & 0 \end{bmatrix} \begin{bmatrix} \mathbf{u}_{\mathcal{N}}(\boldsymbol{\mu}) \\ \mathbf{p}_{\mathcal{N}}(\boldsymbol{\mu}) \end{bmatrix} = \begin{bmatrix} \mathbf{F}_{\mathcal{N}}^s(\boldsymbol{\mu}) + \mathbf{F}_{\mathcal{N}}^l(\boldsymbol{\mu}) \\ \mathbf{G}_{\mathcal{N}}^l(\boldsymbol{\mu}) \end{bmatrix} \quad (24)$$

where $\mathbf{u}_{\mathcal{N}}(\boldsymbol{\mu}) = (u_i^{\mathcal{N}}(\boldsymbol{\mu}))_{i=1}^{\mathcal{N}_{\mathcal{X}}}$ and $\mathbf{p}_{\mathcal{N}}(\boldsymbol{\mu}) = (p_k^{\mathcal{N}})_{k=1}^{\mathcal{N}_{\mathcal{Q}}}$ are the vectors corresponding to (the nodal values of) velocity and pressure. Here, $A_{\mathcal{N}}$ is the stiffness matrix, $B_{\mathcal{N}}$ the gradient matrix, while $C_{\mathcal{N}}(\mathbf{u})$ is the nonlinear transport matrix; $D_{\mathcal{N}}$ is the linearization of the nonlinear transport term, evaluated at $\mathbf{u}_D^{\mathcal{N}}$, FE interpolant of non-zero boundary conditions:

$$(A_{\mathcal{N}}(\boldsymbol{\mu}))_{ij} = a(\boldsymbol{\Phi}_j^{\mathcal{N}}, \boldsymbol{\Phi}_i^{\mathcal{N}}; \boldsymbol{\mu}), \quad (B_{\mathcal{N}}(\boldsymbol{\mu}))_{ki} = b(\varphi_k^{\mathcal{N}}, \boldsymbol{\Phi}_i^{\mathcal{N}}; \boldsymbol{\mu}),$$

$$(C_{\mathcal{N}}(\mathbf{w}; \boldsymbol{\mu}))_{ij} = c \left(\sum_{k=1}^{\mathcal{N}_{\mathcal{X}}} w_k^{\mathcal{N}} \boldsymbol{\Phi}_k^{\mathcal{N}}, \boldsymbol{\Phi}_j^{\mathcal{N}}, \boldsymbol{\Phi}_i^{\mathcal{N}}; \boldsymbol{\mu} \right),$$

$$(D_{\mathcal{N}}(\boldsymbol{\mu}))_{ij} = c(\mathbf{u}_D^{\mathcal{N}}, \boldsymbol{\Phi}_j^{\mathcal{N}}, \boldsymbol{\Phi}_i^{\mathcal{N}}; \boldsymbol{\mu}) + c(\boldsymbol{\Phi}_j^{\mathcal{N}}, \mathbf{u}_D^{\mathcal{N}}, \boldsymbol{\Phi}_i^{\mathcal{N}}; \boldsymbol{\mu}),$$

for $1 \leq i, j \leq \mathcal{N}_{\mathcal{X}}$, $1 \leq k \leq \mathcal{N}_{\mathcal{Q}}$, being $\{\boldsymbol{\Phi}_i^{\mathcal{N}}\}_{i=1}^{\mathcal{N}_{\mathcal{X}}}$, $\{\varphi_i^{\mathcal{N}}\}_{i=1}^{\mathcal{N}_{\mathcal{Q}}}$ two bases of $X^{\mathcal{N}}$ and $Q^{\mathcal{N}}$, respectively. On the other hand, $\mathbf{F}_{\mathcal{N}}^s$ is the source vector while $\mathbf{F}_{\mathcal{N}}^l$ and $\mathbf{G}_{\mathcal{N}}^l$ are the load vectors into which the lifting of nonhomogeneous inflow BCs has been incorporated. The nonlinear system (24) is solved by means of a fixed-point iteration, using as initial guess the Stokes solution obtained by dropping the convective terms.

The reduced basis (RB) method computes an approximation of $Y^{\mathcal{N}}(\boldsymbol{\mu}) = (\mathbf{u}^{\mathcal{N}}(\boldsymbol{\mu}), p^{\mathcal{N}}(\boldsymbol{\mu}))$ by using global approximation spaces made up of FE solutions corresponding to specific choices of the parameter values, or *snapshots* solutions. Denoting with $S_N = \{\boldsymbol{\mu}^1, \dots, \boldsymbol{\mu}^N\}$ a (relatively) small set of parameter values, we introduce the *reduced basis pressure space* $\mathcal{Q}_N^{\mathcal{N}} \subset \mathcal{Q}^{\mathcal{N}}$ as

$$\mathcal{Q}_N^{\mathcal{N}} = \text{span}\{\tilde{\zeta}_n = p^{\mathcal{N}}(\boldsymbol{\mu}^n), n = 1, \dots, N\};$$

the *reduced basis velocity space* $\mathcal{X}_N^{\mathcal{N}} \subset \mathcal{X}^{\mathcal{N}}$ can be built as

$$\mathcal{X}_N^{\mathcal{N}} = \text{span}\{\tilde{\sigma}_n := \mathbf{u}^{\mathcal{N}}(\boldsymbol{\mu}^n), T_p^{\boldsymbol{\mu}} \tilde{\zeta}_n, n = 1, \dots, N\},$$

where $T_p^\mu : \mathcal{Q}_N^\mathcal{N} \rightarrow \mathcal{X}_N^\mathcal{N}$ is the so-called *inner supremizer operator* acting on pressure term,

$$(T_p^\mu q, \mathbf{w})_X = b(q, \mathbf{w}; \boldsymbol{\mu}) \quad \forall \mathbf{w} \in \mathcal{X}_N^\mathcal{N}. \quad (25)$$

The enrichment of the velocity space by means of the *supremizer solutions* enables to fulfill an equivalent stability (LBB) condition also on the RB spaces; see [38, 39] for the details. Moreover, an automatic *greedy* procedure for the selection of the snapshots $\{\boldsymbol{\mu}^1, \dots, \boldsymbol{\mu}^N\}$ is used [14]. The RB approximation of the Navier-Stokes problem (6) is thus obtained through a Galerkin projection onto $\mathcal{Y}_N = \mathcal{X}_N^\mathcal{N} \times \mathcal{Q}_N^\mathcal{N}$: find $Y_N(\boldsymbol{\mu}) = (\mathbf{u}_N(\boldsymbol{\mu}), p_N(\boldsymbol{\mu})) \in \mathcal{Y}_N$ s.t.

$$\mathcal{A}(Y_N(\boldsymbol{\mu}), W; \boldsymbol{\mu}) + \mathcal{C}(Y_N(\boldsymbol{\mu}), Y_N(\boldsymbol{\mu}), W; \boldsymbol{\mu}) = \mathcal{F}(W; \boldsymbol{\mu}), \quad \forall W \in \mathcal{Y}_N. \quad (26)$$

By denoting $\{\zeta_j\}_{j=1}^{2N}$ and $\{\xi_l\}_{l=1}^N$ two orthonormal bases for $\mathcal{X}_N^\mathcal{N}$ and $\mathcal{Q}_N^\mathcal{N}$, respectively, the RB approximation can be written as the solution of the following nonlinear system, whose dimension is now $3N \times 3N$:

$$\begin{bmatrix} A_N(\boldsymbol{\mu}) + D_N(\boldsymbol{\mu}) + C_N(\mathbf{u}_N(\boldsymbol{\mu}); \boldsymbol{\mu}) & B_N^T(\boldsymbol{\mu}) \\ B_N(\boldsymbol{\mu}) & 0 \end{bmatrix} \begin{bmatrix} \underline{\mathbf{u}}_N(\boldsymbol{\mu}) \\ \underline{\mathbf{p}}_N(\boldsymbol{\mu}) \end{bmatrix} = \begin{bmatrix} \mathbf{F}_N^s(\boldsymbol{\mu}) + \mathbf{F}_N^l(\boldsymbol{\mu}) \\ \mathbf{G}_N^l(\boldsymbol{\mu}) \end{bmatrix} \quad (27)$$

where $\underline{\mathbf{u}}_N(\boldsymbol{\mu}) = (u_{Nj}(\boldsymbol{\mu}))_{j=1}^{2N}$ and $\underline{\mathbf{p}}_N(\boldsymbol{\mu}) = (p_{Nl}(\boldsymbol{\mu}))_{l=1}^N$ are the vectors corresponding to the RB velocity and pressure degrees of freedom, respectively. The matrices appearing in (27) are given by

$$(A_N(\boldsymbol{\mu}))_{mn} = a(\zeta^n, \zeta^m; \boldsymbol{\mu}), \quad (B_N(\boldsymbol{\mu}))_{nl} = b(\xi^l, \zeta^n; \boldsymbol{\mu}),$$

$$(C_N(\boldsymbol{\mu}; \mathbf{w}_N))_{mn} = c \left(\sum_{k=1}^{2N} w_{Nk} \zeta^k, \zeta^n, \zeta^m; \boldsymbol{\mu} \right),$$

$$(D_N(\boldsymbol{\mu}))_{mn} = c(\mathbf{u}_D^\mathcal{N}, \zeta^n, \zeta^m; \boldsymbol{\mu}) + c(\zeta^n, \mathbf{u}_D^\mathcal{N}, \zeta^m; \boldsymbol{\mu}),$$

for $1 \leq m, n \leq 2N$, $1 \leq l \leq N$, while the right-hand sides are given by $(\mathbf{F}_N^s)_n = \langle \mathbf{f}, \zeta^n \rangle$, $(\mathbf{F}_N^l(\boldsymbol{\mu}))_n = -a(\mathbf{u}_D^\mathcal{N}, \zeta^n; \boldsymbol{\mu}) - c(\mathbf{u}_D^\mathcal{N}, \mathbf{u}_D^\mathcal{N}, \zeta^n; \boldsymbol{\mu})$, for $1 \leq n \leq 2N$, and $(\mathbf{G}_N^l(\boldsymbol{\mu}))_l = -b(\xi^l, \mathbf{u}_D^\mathcal{N}; \boldsymbol{\mu})$, for $1 \leq l \leq N$.

All the FE computations needed for the construction of the RB spaces are performed on a fixed reference geometrical configuration (and related computational mesh); thus also the reduced solution of problem (26) lies on this reference configuration, and has to be mapped back to the original one. In the following subsections we sketch some important features related to the structure of system (27), its real-time implementation, as well as to the certification of the reduced basis solution with respect to the *truth* FE solution.

5.1 The key ingredient: Offline-Online decomposition

Real-time simulations are made possible because of a suitable *offline-online* computational decomposition: if the RB space generation can be fully decoupled from each subsequent evaluation, then each new *query* (for any parameter $\boldsymbol{\mu} \in \mathcal{D}$) is very cheap, as it involves the solution of a nonlinear system of order $N \ll \mathcal{N}$. The key property behind this decomposition is the *affine parametric dependence*, i.e. the possibility to write the differential forms (10)-(12), as well as the right-hand sides, upon isolating the parameter contribution as follows:

$$a(\mathbf{v}, \mathbf{w}; \boldsymbol{\mu}) = \sum_{q=1}^{Q_a} \Theta_a^q(\boldsymbol{\mu}) a^q(\mathbf{v}, \mathbf{w}), \quad b(p, \mathbf{w}; \boldsymbol{\mu}) = \sum_{q=1}^{Q_b} \Theta_b^q(\boldsymbol{\mu}) b^q(p, \mathbf{w}), \quad (28)$$

$$c(\mathbf{u}, \mathbf{v}, \mathbf{w}; \boldsymbol{\mu}) = \sum_{q=1}^{Q_c} \Theta_c^q(\boldsymbol{\mu}) c^q(\mathbf{u}, \mathbf{v}, \mathbf{w}), \quad \langle F^s(\boldsymbol{\mu}), \mathbf{w} \rangle = \sum_{r=1}^{Q_f} \Theta_f^r(\boldsymbol{\mu}) \langle F_r^s, \mathbf{w} \rangle \quad (29)$$

for suitable integers $Q_a, Q_b = Q_c, Q_f$, being q and r condensed indices for (i, j, k) and (i, k) , with

$$a^{q(i,j,k)}(\mathbf{u}, \mathbf{v}) = \int_{\Omega} \xi_k^{i,j}(\mathbf{x}) \frac{\partial \mathbf{u}}{\partial x_i} \frac{\partial \mathbf{v}}{\partial x_j} d\Omega, \quad b^{q(i,j,k)}(p, \mathbf{w}) = \int_{\Omega} \eta_k^{i,j}(\mathbf{x}) p \frac{\partial \mathbf{w}_i}{\partial x_j} d\Omega, \quad (30)$$

$$c^{q(i,j,k)}(\mathbf{u}, \mathbf{v}, \mathbf{w}) = \int_{\Omega} u_i \chi_k^{i,j}(\mathbf{x}) \frac{\partial v_k}{\partial x_j} w_k d\Omega, \quad \mathcal{F}_s^r(\mathbf{w}) = \int_{\Omega} \psi_k^i(\mathbf{x}) \mathbf{w}_i d\Omega. \quad (31)$$

This affine decomposition is straightforward when the mapping $T(\cdot; \boldsymbol{\mu})$ is affine. In the more general nonaffine case (such as for the RBF technique), we rely on an affine approximation obtained through an intermediate empirical interpolation procedure [40]; see [9] for a detailed description in the analogous case of a Stokes problem. In any case, we end up with the following (possibly approximated) affine expansion of the parametrized tensors:

$$\nu_{ij}(\mathbf{x}, \boldsymbol{\mu}) = \sum_{k=1}^{K_{ij}^a} \beta_k^{i,j}(\boldsymbol{\mu}) \xi_k^{i,j}(\mathbf{x}), \quad \chi_{ij}(\mathbf{x}, \boldsymbol{\mu}) = \sum_{k=1}^{K_{ij}^b} \gamma_k^{i,j}(\boldsymbol{\mu}) \eta_k^{i,j}(\mathbf{x}), \quad i, j = 1, 2 \quad (32)$$

which entails $\Theta_a^q(\boldsymbol{\mu}) = \beta_k^{i,j}(\boldsymbol{\mu})$, $\Theta_b^q(\boldsymbol{\mu}) = \Theta_c^q(\boldsymbol{\mu}) = \gamma_k^{i,j}(\boldsymbol{\mu})$ in (30)-(31). In the same way,

$$f_i(\mathbf{x}) |\det J_T(\mathbf{x}, \boldsymbol{\mu})| = \sum_{k=1}^{K_i^f} \delta_k^i(\boldsymbol{\mu}) \psi_k^i(\mathbf{x}), \quad \Theta_f^r(\boldsymbol{\mu}) = \delta_k^i(\boldsymbol{\mu}), \quad i = 1, 2, r = 1, \dots, Q_f.$$

In particular, from the affine parametric expansions (28)-(29), we can write the FE matrices isolating the parametric contribution as follows:

$$\begin{aligned}
A_{\mathcal{N}}(\boldsymbol{\mu}) &= \sum_{q=1}^{Q_a} \Theta_a^q(\boldsymbol{\mu}) A_{\mathcal{N}}^q, & A_{\mathcal{N}}^q &= a^q(\boldsymbol{\Phi}_j^{\mathcal{N}}, \boldsymbol{\Phi}_i^{\mathcal{N}}), \\
B_{\mathcal{N}}(\boldsymbol{\mu}) &= \sum_{q=1}^{Q_b} \Theta_b^q(\boldsymbol{\mu}) B_{\mathcal{N}}^q, & B_{\mathcal{N}}^q &= b^q(\varphi_k^{\mathcal{N}}, \boldsymbol{\Phi}_i^{\mathcal{N}}), \\
C_{\mathcal{N}}(\mathbf{w}_{\mathcal{N}}; \boldsymbol{\mu}) &= \sum_{q=1}^{Q_c} \Theta_c^q(\boldsymbol{\mu}) C_{\mathcal{N}}^q(\mathbf{w}_{\mathcal{N}}), & C_{\mathcal{N}}^q(\mathbf{w}_{\mathcal{N}}) &= c^q\left(\sum_{k=1}^{\mathcal{N}_X} w_k^{\mathcal{N}}(\boldsymbol{\mu}) \boldsymbol{\Phi}_k^{\mathcal{N}}, \boldsymbol{\Phi}_j^{\mathcal{N}}, \boldsymbol{\Phi}_i^{\mathcal{N}}\right),
\end{aligned} \tag{33}$$

so that the algebraic structures which do not depend on the parameters can be computed and stored just once, during the *offline* stage. Moreover, during this stage the RB spaces are constructed, obtaining at the end of this procedure two basis matrices, $\mathbb{Z}_{2N} \in \mathbb{R}^{\mathcal{N}_X \times 2N}$ and $\mathbb{Z}_N \in \mathbb{R}^{\mathcal{N}_Q \times N}$ for the velocity and the pressure RB space, respectively, given by

$$(\mathbb{Z}_N)_{il} = \xi_i^l, \quad (\mathbb{Z}_{2N})_{jm} = \zeta_j^m, \quad 1 \leq i \leq \mathcal{N}_Q, \quad 1 \leq j \leq \mathcal{N}_X,$$

for $1 \leq l \leq N$, $1 \leq m \leq 2N$, being the basis functions expressed as

$$\xi^l(\mathbf{x}) = \sum_{i=1}^{\mathcal{N}_Q} \xi_i^l \varphi_i^{\mathcal{N}}(\mathbf{x}), \quad \zeta^m(\mathbf{x}) = \sum_{j=1}^{\mathcal{N}_X} \zeta_j^m \boldsymbol{\Phi}_j^{\mathcal{N}}(\mathbf{x}), \quad 1 \leq l \leq N, \quad 1 \leq m \leq 2N.$$

The basis matrices \mathbb{Z}_{2N} and \mathbb{Z}_N connect the FE and the RB algebraic structures, by acting as a projection onto the reduced space. In fact, it is straightforward to show that

$$\begin{aligned}
A_N(\boldsymbol{\mu}) &= \sum_{q=1}^{Q_a} \Theta_a^q(\boldsymbol{\mu}) A_N^q, & A_N^q &= \mathbb{Z}_{2N}^T A_{\mathcal{N}}^q \mathbb{Z}_{2N}, & 1 \leq q \leq Q_a, \\
B_N(\boldsymbol{\mu}) &= \sum_{q=1}^{Q_b} \Theta_b^q(\boldsymbol{\mu}) B_N^q, & B_N^q &= \mathbb{Z}_N^T A_{\mathcal{N}}^q \mathbb{Z}_{2N}, & 1 \leq q \leq Q_b, \\
C_N(\zeta^l; \boldsymbol{\mu}) &= \sum_{q=1}^{Q_c} \Theta_c^q(\boldsymbol{\mu}) C_N^q(\zeta^l), & C_N^q(\zeta^l) &= \mathbb{Z}_{2N}^T C_{\mathcal{N}}^q(\zeta^l) \mathbb{Z}_{2N}, & 1 \leq q \leq Q_c,
\end{aligned} \tag{34}$$

so that, during the *offline* stage, all the structures required by the RB machinery can be simply assembled and stored. The construction of the structures corresponding to the right-hand-sides is very similar, while nonlinear terms require some additional work for constructing the $Q_c \times 2N$ matrices $C_{\mathcal{N}}^q(\zeta^l)$, for each of Q_c affine terms and each of $2N$ basis functions in the velocity space.

Once all this database is built, the solution of the RB problem (26) can be performed *online*, for each value of the parameters $\boldsymbol{\mu} \in \mathcal{D}$, implying only handling (very) small algebraic structures. This fact entails the possibility to obtain real-time solutions, at the cost of a very reduced nonlinear system, yet providing the accuracy of a full FE simulation. We remark that, thanks to (34), the reduced basis approximation $s_N(\boldsymbol{\mu}) = a(\mathbf{u}_N(\boldsymbol{\mu}), \mathbf{u}_N(\boldsymbol{\mu}); \boldsymbol{\mu})$ of the parametric output (13) can be simply evaluated as the matrix-vector product $s_N(\boldsymbol{\mu}) = \underline{\mathbf{u}}_N^T(\boldsymbol{\mu}) A_N(\boldsymbol{\mu}) \underline{\mathbf{u}}_N(\boldsymbol{\mu})$.

5.2 Reliability of the reduced model: some ingredients for a posteriori error estimation

In order to certificate the RB solution and provide an *a posteriori* error bound w.r.t. the FE truth solution, we can compute a residual-based estimator [22], specializing the fixed point argument in [33] to the RB framework. Let us denote $\tilde{r}(W; \boldsymbol{\mu}) = \mathcal{F}(W; \boldsymbol{\mu}) - \mathcal{G}(Y_N(\boldsymbol{\mu}), W; \boldsymbol{\mu})$ the residual of the Navier-Stokes equation and its dual norm

$$\|r(\cdot; \boldsymbol{\mu})\|_{\mathcal{Y}'} = \sup_{W \in \mathcal{Y}} \frac{\mathcal{G}(Y_N(\boldsymbol{\mu}), W; \boldsymbol{\mu})}{\|W\|_{\mathcal{Y}}}.$$

Moreover, let us introduce the following *adimensionalized* residual (or *proximity indicator*):

$$\tau_N(\boldsymbol{\mu}) = \frac{2\rho^2 \|r(\cdot; \boldsymbol{\mu})\|_{\mathcal{Y}'}}{\beta^{LB}(\boldsymbol{\mu})},$$

being ρ the Sobolev embedding constant (15) and $\beta_N^{LB}(\boldsymbol{\mu})$ a lower bound of the inf-sup stability factor

$$\beta_N(\boldsymbol{\mu}) = \beta_N(Y_N(\boldsymbol{\mu})) = \inf_{W \in \mathcal{Y}^N} \sup_{V \in \mathcal{Y}^N} \frac{d\mathcal{G}(Y_N(\boldsymbol{\mu}); \boldsymbol{\mu})(W, V)}{\|W^N\|_{\mathcal{Y}} \|V\|_{\mathcal{Y}}};$$

we require $0 < \beta_N^{LB}(\boldsymbol{\mu}) \leq \beta_N(Y_N(\boldsymbol{\mu}))$ for all $\boldsymbol{\mu} \in \mathcal{D}$. It is possible to show [23, 36] that, for $\tau_N(\boldsymbol{\mu}) < 1$, it exists a unique $Y^N(\boldsymbol{\mu}) \in \mathcal{B}_{\mathcal{Y}}(Y_N(\boldsymbol{\mu}); \beta_N(Y_N(\boldsymbol{\mu}))/\rho^2)$, being $\mathcal{B}_{\mathcal{Y}}(Z, r) = \{Y \in \mathcal{Y} : \|Y - Z\|_{\mathcal{Y}} \leq r\}$. Moreover, the following *a posteriori* error estimation is satisfied:

$$\|Y^N(\boldsymbol{\mu}) - Y_N(\boldsymbol{\mu})\|_{\mathcal{Y}} \leq \Delta_N(\boldsymbol{\mu}) := \frac{\beta_N(\boldsymbol{\mu})}{\rho^2} \left(1 - \sqrt{1 - \tau_N(\boldsymbol{\mu})}\right), \quad (35)$$

thus providing a rigorous error bound. It is important to remark that also the computations of (35) can be decoupled following the previous *offline-online* stratagem, in order to get an unexpensive evaluation of the error bound for each $\boldsymbol{\mu} \in \mathcal{D}$. Moreover, error bound (35) is employed also during the *Greedy* procedure for the RB space assembling^b: given a (sufficiently rich) finite training sample $\Xi_{train} \subset \mathcal{D}$ and the first parameter value $\boldsymbol{\mu}^1$, the remaining snapshots are chosen as $\boldsymbol{\mu}^n = \arg \max_{\boldsymbol{\mu} \in \Xi_{train}} \Delta_{n-1}(\boldsymbol{\mu})$ for $n = 2, \dots, N$ until an error tolerance ε_{tol}^{RB} is achieved, i.e. when $\Delta_N(\boldsymbol{\mu}) \leq \varepsilon_{tol}^{RB}$ for all $\boldsymbol{\mu} \in \Xi_{train}$.

6 Parametric descriptions of carotid bifurcations

Geometrical variations strongly influence flow features in arterial vessels, as well as the occurrence of lesions or pathologies [6, 41, 1, 3]. Typical portions of

^bIn this case, the surrogate $\beta_N(Y^N(\boldsymbol{\mu}))$ is used, since $Y_N(\boldsymbol{\mu})$ is of course not yet available [23].

cardiovascular network where they may develop are made up by curved vessels and bifurcations; an important segment where vessel diseases are often clinically observed is the human carotid artery. The common carotid artery (CCA) bifurcates in the lower neck into two branches, the internal and the external carotid arteries (ICA and ECA, respectively). Stenoses, that is the narrowing of the inner portion of an artery, manifest quite often in the ICA. It is well known that *(i)* the carotid shape yields a complex flow dynamics, *(ii)* its anatomy changes considerably among individuals, and *(iii)* the risk of strokes is directly related to carotid lesions severity. All that motivates the effort to characterize blood flows in this arterial tract. Our interest in developing a reduced model aims at:

- a) performing real-time simulations of blood flows in different carotid artery bifurcations, whose parametrized shapes may possibly be reconstructed from given data (see Sect. 4);
- b) investigating the effect of stenosed geometries on flow patterns by spanning a broad family of stenosed carotid bifurcations and identify severe stenosis configurations by measuring some indices of occlusion risk, such as the viscous energy dissipation or the vorticity.

This kind of analysis based on shape sensitivity, as well as shape optimization procedures [42], are also relevant for stenting procedures, where carotid endarterectomy may require the possibility to modify or reconstruct vessel geometry in order to optimize the flow characteristics therein.

On the one hand, several works have been focused on the numerical simulation of blood flows in one or few carotid artery bifurcations obtained directly by medical images, by using computational meshes reconstructed from patient data in a very precise way; on the other hand, many analyses [11, 43] have tried to investigate the relationship between intra-individuals variations of some risk factors and *geometric indices*. With this aim, some simplified shape parametrizations of the carotid bifurcation have been introduced, mainly based on sections areas or diameters, curvature of the branches, bifurcation angle and so on. In particular, we can mention (see Fig. 3):

- *Y model*: introduced in [44, 45] and derived from biplanar angiograms, it defines a carotid bifurcation as a function of the diameter D of the CCA, the diameters d_i and d_e of the ICA and ECA outflows, the diameters at the proximal section d_r , mid-sinus section d_s of the ICA, the distance l_s between proximal and midsinus sections, the angles α_i , α_e between the ICA (resp. ECA) and the CCA. A second version, using inwardly curved branches instead of straight branches, has been discussed in [6]; here a further parameter d_m representing the mid-sinus cross section is introduced, whereas d_s represents the diameter of the section where the maximum sinus width occurs. Because of the curvature, two angles between CCA and ICA are considered, at the origin of the branching ($\alpha_{i,o}$) and at its end ($\alpha_{i,e}$).

- *tuning fork model*: introduced in [46] and derived from carotid specimens obtained from autopsy, it represents the carotid bifurcation like a tuning fork. The geometrical parameters are in this case the five diameters D , d_i , d_e , d_r and d_s , the bifurcation angle $\alpha = \alpha_i + \alpha_e$, the distance l_s (as in the previous case), the distance between the two branches L , and two more lengths representing the sectional diameter d_u of the ICA at the end of the sinus and the distance G of this section from the divider tip. A parametric CAD model based on the tuning fork model has been introduced in [42, 11] and is based on the following parameters: the ICA and ECA angles θ_i , θ_e , the ICA and ECA diameters d_i , d_e , the position d of a Bézier curve control point P , which enables to control the maximum width of the sinus bulb. In this case, also the inner curve of the bifurcation is described by means of a Bézier curve.

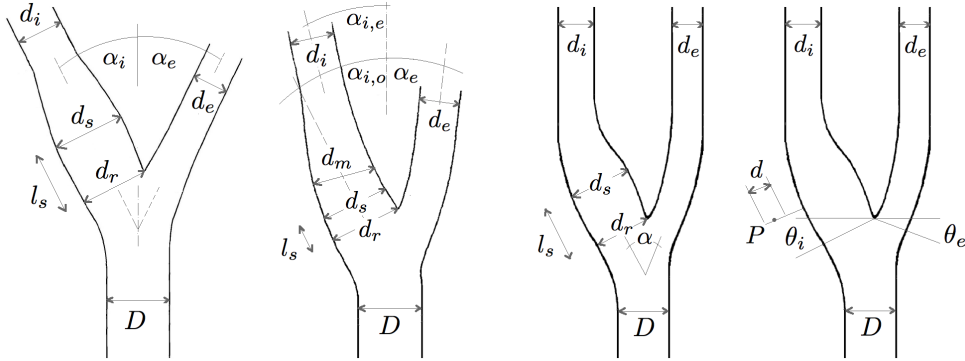


Figure 3: Parametrized models of the carotid artery bifurcation. From left to right: *Y model* with straight branches, *Y model* with curved branches, *tuning fork model* and parametric *tuning fork CAD model*.

With respect to these models, our approach to define a shape parametrization is based on a free choice of the *input* parameters, which are given by the displacements of some control points located on the boundary of a reference carotid configuration; hence, deformed shapes will be simply obtained by moving the control points. In this way, we get a versatile shape parametrization, since control points may be located either *a)* on peripheral positions on the branches for describing global deformations or *b)* on selected positions for modeling local modifications such as stenoses or restrictions. Moreover, since we are interested to perform real-time blood flow simulations, the shape parametrization has to be low-dimensional, i.e. defined with respect to a small number of *input* parameters μ_1, \dots, μ_p : in this way, the complexity of the parametrized problem, given by the number of terms in operators (28)-(29) and increasing with p , does not grow too much. For these reasons, a parametrization based on radial basis functions represents a good compromise: despite its low-dimensionality, its interpolatory nature enables to track some selected points on the shape boundary, and to obtain a global mapping by combining local deformations of control points.

With this aim, we introduce a reference configuration Ω , and select a set of control points and define a RBF mapping^c as in Sect. 3. This technique is applied here to a two-dimensional configuration, even if three-dimensional geometries can be handled using the same procedure. Hence, we compute the parametrized tensors appearing in the parametrized forms (7)–(9), get an affine approximation through the empirical interpolation method [40], assemble the $\boldsymbol{\mu}$ -independent FEM structures appearing in (33), then build the RB spaces, as well as the $\boldsymbol{\mu}$ -independent RB structures appearing in (34). This is the so-called *offline* stage. During the *online* stage, for each new carotid configuration we select the corresponding parameter value $\boldsymbol{\mu}$, evaluate the $\boldsymbol{\mu}$ -dependent $\Theta(\boldsymbol{\mu})$ functions and assemble the reduced system (34); its solution gives the RB approximation $(\mathbf{u}_N(\boldsymbol{\mu}), p(\boldsymbol{\mu}))$ of the flow field, from which we obtain the full velocity and pressure multiplying by the matrices \mathbb{Z}_{2N} and \mathbb{Z}_N , respectively. In case of a simulation performed on a reconstructed configuration, the parameter value $\boldsymbol{\mu}$ will be obtained by the solution of the parameter identification problem (22).

7 Numerical results

In this section we present some numerical results obtained within the reduced framework previously introduced. In particular, we illustrate some examples of real-time simulations on parametrized geometries reconstructed from surrogate data, as well as a first shape sensitivity analysis based on viscous energy dissipation on a family of parametrized stenosed configurations.

As reference configuration Ω we consider the geometry depicted in Fig. 4, obtained as a (longitudinal two-dimensional) projection of a carotid configuration extracted from medical images (MRI). No-slip conditions are imposed on Γ_w , while homogeneous (free-stress) Neumann conditions are imposed on the out-flow sections Γ_{out} ; on the inflow boundary Γ_{in} , we impose a Poiseuille parabolic profile; inflow peak velocity is $\tilde{v} = 30\text{cm s}^{-1}$, while arterial diameter is about 0.7cm . We model blood as an incompressible Newtonian flow, with dynamic viscosity is $\mu = 0.04\text{g cm}^{-1} \text{s}^{-1}$, density $\rho = 1\text{g cm}^{-3}$, thus yielding a kynematic viscosity $\nu = \mu/\rho = 0.04\text{cm}^2 \text{s}^{-1}$ and a Reynolds number $Re = \tilde{v}D/\nu$ of order 500.

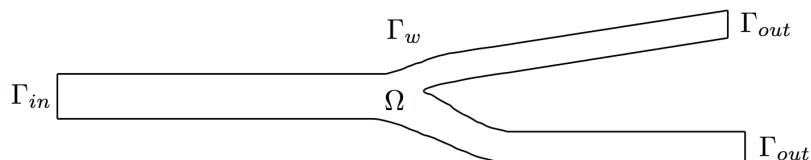


Figure 4: Reference configuration Ω used for all the FE and RB computations.

^cAll the details related to the shape parametrization, as well as to the numerical tests, will be discussed in the next section.

Two different cases have been investigated, dealing with either *(i)* global deformations of the carotid branches, or *(ii)* stenoses or restrictions near the carotid sinus. In both cases, the RB method for cutting down the computational burden – naturally arising in a *many-query* context – has been coupled to a RBF mapping for the reduction of geometric complexity. Computations^d for both cases are based upon a finite element approximation on $\mathbb{P}^2 - \mathbb{P}^1$ spaces of dimension $\mathcal{N}_X = 22,728$ and $\mathcal{N}_Q = 2,946$ respectively, obtained on a mesh of 5,473 triangular elements.

7.1 Reconstructed geometries

The first case focuses on the capability of our reduced framework to represent a family of different carotid bifurcations and provide real-time blood flow simulations. In particular, we build a RBF parametrization by using $k = 7$ control points and $p = 7$ input parameters μ_1, \dots, μ_p , given by the horizontal displacements of the control points (see Fig. 5), with $\mu_i \in [-0.2, 0.2] \forall i = 1, \dots, 7$. In this case, a cubic function $\sigma(h) = h^3$ has been chosen as radial kernel, in order to obtain global deformations by moving the control points located on the branches, far from the center. The affine approximation of the parametrized tensors (computed through the empirical interpolation method) entails a parametrized expression of the operators (28)-(29) made by $Q_a + Q_b + Q_c = 78$ terms^e.

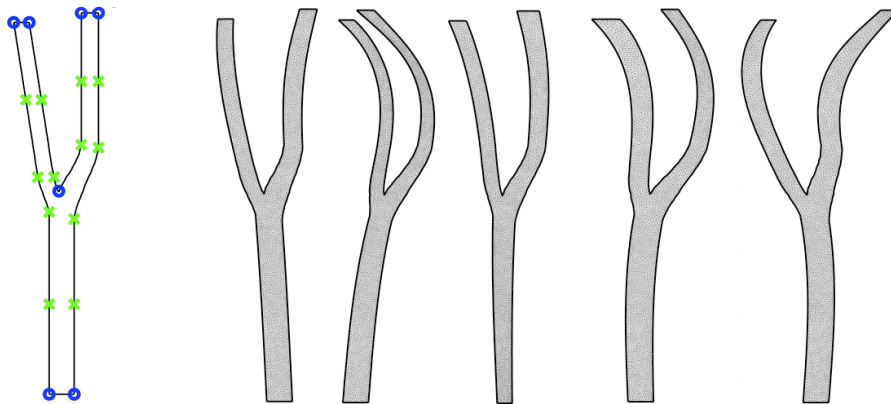


Figure 5: Case I. (Left) Reference configuration, control points (represented as blue dots) for the RBF map and registration points (represented as green crosses). (Right) Different carotid bifurcations obtained through the RBF map by displacing the control points.

The reduced spaces are built using a greedy procedure [14] based on the a posteriori error bound (35); by fixing a tolerance $\varepsilon_{tol}^{RB} = 2.5 \times 10^{-3}$, $N = 10$ snapshots

^dComputations have been executed on a desktop with 2×2 GHz Dual Core AMD Opteron (tm) processors 2214 HE and 16 GB of RAM.

^eThis complexity is still reasonable for a nonaffine parametrization after empirical interpolation treatment

are selected, resulting in spaces $\mathcal{X}_N^{\mathcal{N}}$, $\mathcal{Q}_N^{\mathcal{N}}$ of dimension 20 and 10, respectively. This yields the possibility to perform real-time blood flow simulations, each query $\boldsymbol{\mu} \rightarrow (\mathbf{u}_N(\boldsymbol{\mu}), p(\boldsymbol{\mu}))$ requiring $\mathcal{O}(1 s)$. In the case at hand, solving the nonlinear RB Navier-Stokes problem (27) require about 2.3 s, while the post-processing for the output ranges between $\mathcal{O}(10^{-3} s)$ (energy output) to 1.5 s (vorticity output).

Concerning the shape reconstruction, the parameter identification procedure (22) has been tested on 20 different configurations^f, by considering $N_r = 12$ registration points, divided in 6 couples, and placing two couples of points on each carotid branch. The minimization of the cost functional appearing in the problem (22) has been performed by means of an automatic version of the sequential quadratic programming (SQP) algorithm, which takes (average over the 20 run cases) 5.3 s. We remark that the choice of the weighting parameter $\beta > 0$ in (22) plays an important role (see Fig. 6); on one hand, values of β which are too small do not allow a good control of the shape at the extremities; on the other hand, penalizing too much the matching constraint on the control points and their images on the target shape results in large deformations on the central sections.

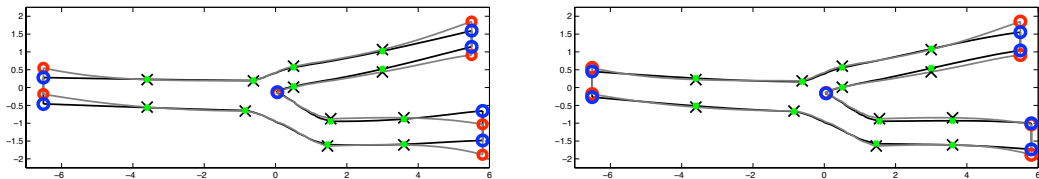


Figure 6: Case I. Shape reconstruction obtained through parameter identification with $\beta = 0.01$ (left) and $\beta = 0.5$ (right). The deformed and target configurations are represented in black and grey, respectively. Control points are depicted in blue (deformed positions) and red (target positions); registration points are depicted in green (deformed positions) and black (target positions).

7.2 Output evaluation for stenosed configurations

The second case deals with the parametrization of stenosed carotid configuration and a simple, preliminary shape sensitivity analysis based on reduced flow simulation. Our goal is to characterize the blood flow in a very large number (say, $\mathcal{O}(10^3)$) of configurations and investigate the correlation between some outputs of interest and a few geometrical features. Needless to say, the simulation in such a *many-query* context is made possible only by using a reduction strategy, following the framework set up in the previous sections.

^fIn each case, the (surrogate) target configuration to be reconstructed has been generated as a random free-form deformation of the reference geometry (built on a 8×6 lattice of control points) in order to avoid a scaling/translation preprocessing.

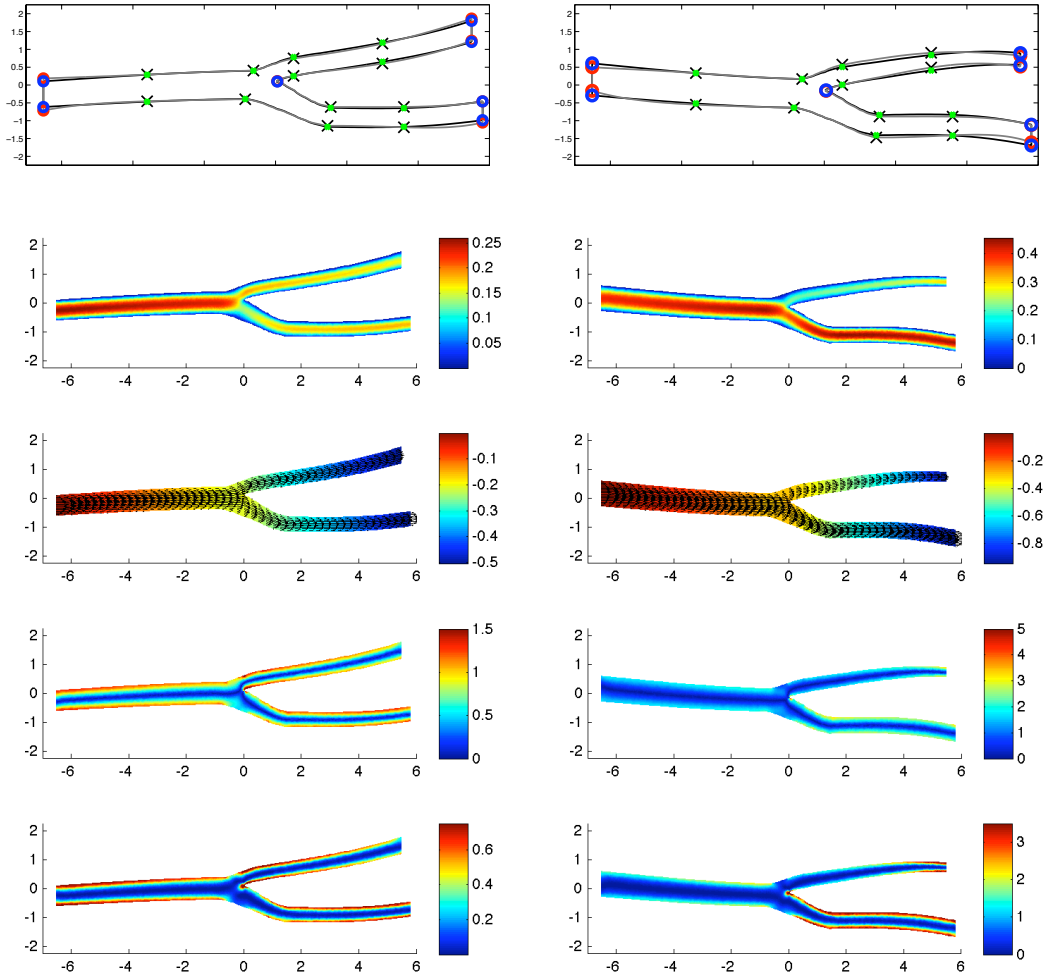


Figure 7: Case I. Reduced basis approximation of blood flows in two different reconstructed configurations. From top to bottom: reconstructed (in black) and target (in grey) configurations; velocity and pressure fields with velocity profiles; vorticity magnitude and velocity gradient squared magnitude, related with shear stresses.

In order to obtain a low-dimensional parametrization of stenoses located at the bifurcation and at the mid-sinus level of the ICA, we build a RBF parametrization on 4 control points located on these sections, with a Gaussian radial kernel $\sigma(h) = \exp(-h^2)$, so that the effect of the deformation is localized in these two segments. Nevertheless, in order to avoid (even small) global deformations, we fix the position of inflow/outflow sections, by considering as control points also the extrema of these three sections; since they are not free to move, the effective input parameters are indeed the horizontal displacements $\mu_i \in [-0.2, 0.2] \forall i = 1, \dots, 4$ of the central control points only (see Fig. 8). Concerning the parametrized formulation and the reduced basis approximation of the Navier-Stokes problem, operators (28)-(29) are made in this

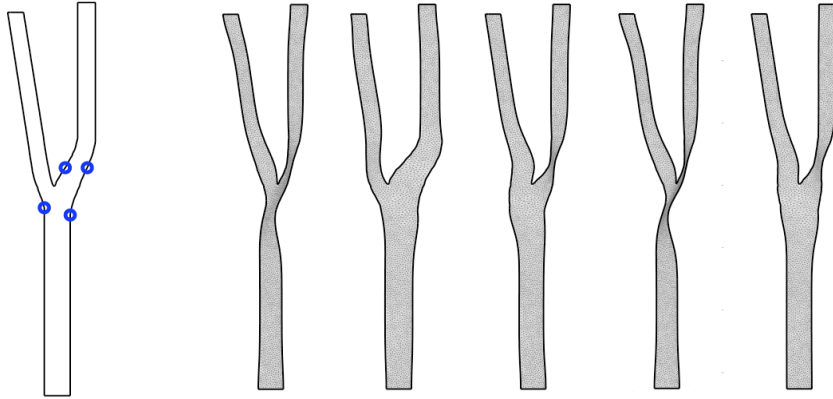


Figure 8: Case II. (Left) Reference configuration and control points (represented as blue dots) for the RBF map. (Right) Different carotid bifurcations obtained through the RBF map by displacing the control points.

case by $Q_a + Q_b + Q_c = 62$ terms[§]; by fixing a tolerance $\varepsilon_{tol}^{RB} = 2.5 \times 10^{-3}$, the greedy algorithm for the reduced spaces construction selects $N = 15$ snapshots, resulting in spaces \mathcal{X}_N^N , \mathcal{Q}_N^N of dimension 30 and 15, respectively. In Fig. 9 two representative examples of RB approximation of velocity and pressure fields are represented; the average time for a single flow simulation is around 2.5 s (all the details concerning computational times and reduction aspects are summarized in the next section). We underline that, thanks to the *offline-online* decomposition, we are able to obtain also a rigorous certification of the RB solution $Y_N(\boldsymbol{\mu}) = (\mathbf{u}_N(\boldsymbol{\mu}), p_N(\boldsymbol{\mu}))$ with respect to the “truth” finite element solution $Y^N(\boldsymbol{\mu}) = (\mathbf{u}^N(\boldsymbol{\mu}), p^N(\boldsymbol{\mu}))$. The *online* evaluation of the a posteriori error bound (35) require an average time of about 1.1 s. The average of the error bounds, computed on a sample of 1000 parameter values, is about 1.5×10^{-4} .

Thanks to the strong computational reduction, it is possible to perform a large number of simulations in a very reduced time with respect to more traditional discretization techniques, in order to investigate the effect of geometrical variations on flow features. In Fig. 10 we display the viscous energy dissipation obtained from the RB flow simulation in 1000 different configurations, as a function of the diameters d_c of the CCA at the bifurcation and d_b of the mid-sinus level of the ICA, respectively. As expected, flow disturbances caused by stenoses lead to higher values of the dissipated energy, the maximum occurring for the smallest diameters on both sections. The possibility to derive a (nonlinear) regression model – as well as a response surface model – from the computed output in order to explain quantitatively the relationship between dissipated energy and geometrical indices is foreseen and currently under investigation.

[§]As in the previous case, this complexity is still reasonable for a nonaffine parametrization after empirical interpolation treatment

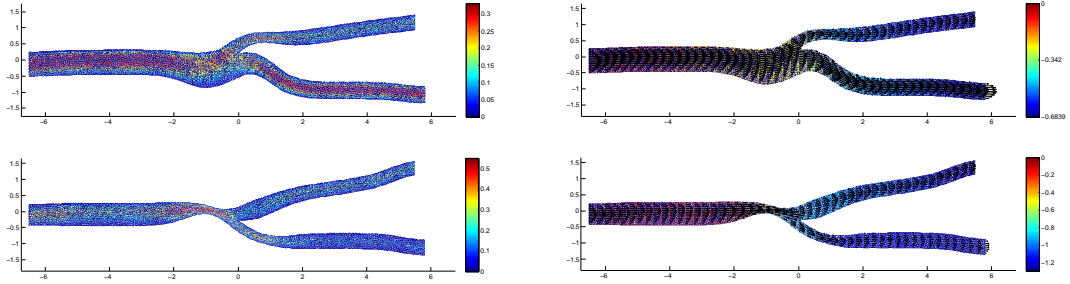


Figure 9: Case II. Reduced basis approximation of blood flows in two different reconstructed configurations: velocity (left) and pressure fields with velocity profiles (right).

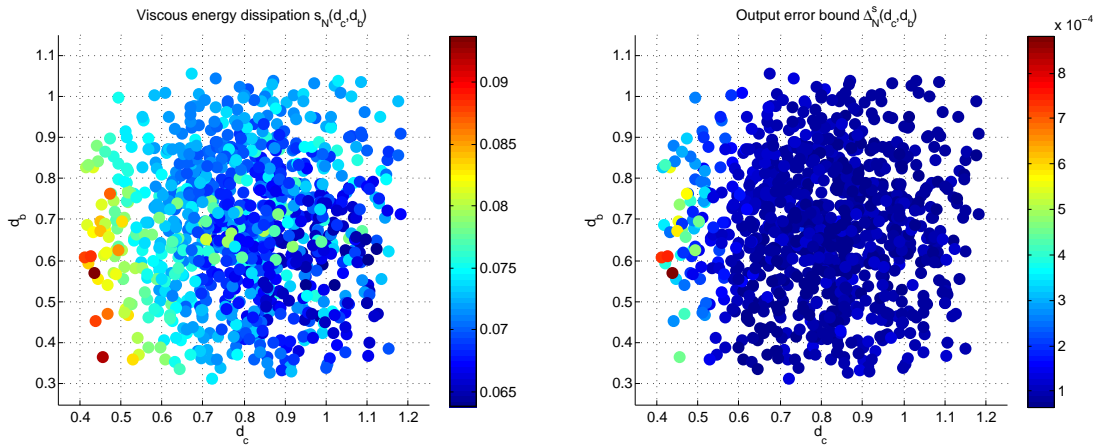


Figure 10: Case II. (Left) Viscous energy dissipation computed in 1000 configurations, as a function of the diameters $d_c = d_c(\mu_1, \mu_2)$ of the CCA at the bifurcation and $d_b = d_b(\mu_3, \mu_4)$ of the mid-sinus level of the ICA, respectively. (Right) A posteriori estimation $\Delta_N(\boldsymbol{\mu})$ of the error between the RB field solution and the corresponding FE solutions.

7.3 The complete reduction paradigm: computational features

We conclude this section by discussing some computational aspects related to the two cases presented above, and showing how reduced basis techniques allow a substantial reduction of computational work. All the details are reported in Table 1. Compared to the corresponding FE approximation, RB online evaluations of field variables enable a computational speedup, defined as $\mathcal{S}_j = t_{FE}/t_{RB}^{online}$, of about two orders of magnitude. This is made possible because of the reduction in linear system dimensions, and finally in the huge dimensional reduction – N vs. \mathcal{N} – between RB spaces and corresponding FE spaces. For the three cases considered, this ratio goes from 570 (first case) to 860 (second case).

By taking into account also the time spent for the offline construction and storage, we find the break-even point (given by $\mathcal{Q}_{BE} = t_{RB}^{offline}/t_{FE}$) is of $\mathcal{O}(10^2)$

in both cases, which can be considered acceptable whenever interested either in the real-time context, or in the limit of many queries with $\mathcal{O}(10^2)$ evaluations (at least).

Approximation data	Case 1	Case 2
Number of parameters p	7	4
Affine op. components $Q_a + Q_b + Q_c$	78	62
Affine rhs components Q_f	79	63
FE space dim. $\mathcal{N}_{\mathcal{X}} + \mathcal{N}_{\mathcal{Q}}$	25674	25674
RB space dim. $3N$	30	45
RB construction $t_{RB}^{offline}$	$\approx 12 h$	$\approx 11 h$
RB evaluation t_{RB}^{online}	2.3 s	2.5 s
FE evaluation t_{FE} (s)	210 s	240 s
Computational speedup \mathcal{S}_c	96	88
Break-even point Q_{BE}	205	165

Table 1: Numerical details for the test cases presented. RB spaces have been built by means of the greedy procedure, using a tolerance $\varepsilon_{tol}^{RB} = 2.5 \times 10^{-3}$ and a uniform RB greedy train sample of size $n_{train} = 1000$. A comparison of the computational times between the online RB evaluations and the corresponding FE simulations is reported. Here $t_{RB}^{offline}$ is the time of the offline RB construction and storage, t_{RB}^{online} is the time of an online RB computation, while t_{FE} is the time for a FE computation, once FE matrices are built.

8 Conclusions and remarks

In this work we have developed a reduced framework for real-time simulations in parametrized computational domains, which has then been applied to the blood flow simulation in parametrized carotid configurations. A radial basis function setting has been presented and applied to carotid bifurcations, for modeling either global and localized deformations. With respect to more classical carotid parametrizations, this technique has shown a better flexibility in shape representation keeping under control the number of parameters and allowing a better boundary control. A certified and accurate reduced basis approximation of the steady parametrized Navier-Stokes equations, built upon a finite element discretization, has been introduced and applied for the blood flow simulations in parametrized configurations. As shown in the considered cases, it allows a substantial reduction in computational efforts with respect to classical strategies based on full finite element simulations.

For more precise investigations, the extension of this framework to unsteady Navier-Stokes equations and more realistic three-dimensional configurations is foreseen and represents our current research activity.

Acknowledgements

We acknowledge the use of `MLife` library previously developed by Prof. Fausto Saleri as a basis for the numerical FE simulations presented in this paper. We also acknowledge, in the framework of the Progetto Roberto Rocca (MIT-Politecnico di Milano), Prof. A.T. Patera (MIT) for his insights, as well as all the people who have contributed with him to the `rbMIT` package used for RB computations presented in this work. In particular, we thank Dr. T. Lassila (EPFL) for his feedbacks and Ms. Laura Azzimonti (Politecnico di Milano) for helpful discussions and suggestions.

This work has been supported in part by the Swiss National Science Foundation (Project 200021-122136) and by the ERC-Mathcard Project (Project ERC-2008-AdG 227058).

References

- [1] Milner J, Moore J, Rutt B, Steinman D: **Hemodynamics of human carotid artery bifurcations: Computational studies with models constructed with magnetic resonance imaging of normal subjects.** *Journal of Vascular Surgery* 1998, **28**:143–156.
- [2] Formaggia L, Quarteroni A, Veneziani A: *Cardiovascular Mathematics. Modeling and simulation of the circulatory system.* Springer, Series MS&A, Vol. 1 2009.
- [3] Quarteroni A, Tuveri M, Veneziani A: **Computational vascular fluid dynamics: Problems, models and methods.** *Computing and Visualization in Science* 2000, **2**:163–197.
- [4] Ambrosi D, Quarteroni A, Rozza (Eds) G: *Modelling of Physiological Flows.* Springer, Series MS&A, Vol. 5 2011.
- [5] Loth F, Fischer P, Bassiouny H: **Blood Flow in End-to-Side Anastomoses.** *Annu. Rev. Fluid. Mech.* 2008, **40**:367–393.
- [6] Perktold K, Resch M: **Numerical flow studies in human carotid artery bifurcations: Basic discussion of the geometric factor in atherogenesis.** *Journal of Biomedical Engineering* 1990, **12**:111–123.
- [7] Mut F, Löhner R, Chien A, Tateshima S, Viñuela F, Putman C, Cebal J: **Computational hemodynamics framework for the analysis of cerebral aneurysms.** *International Journal for Numerical Methods in Biomedical Engineering* 2010.
- [8] Rozza G: **On optimization, control and shape design of an arterial bypass.** *Int. J. Numer. Meth. Fluids* 2005, **47**(10–11):1411–1419.
- [9] Manzoni A, Quarteroni A, Rozza G: **Shape optimization for viscous flows by reduced basis methods and free-form deformation.** *Submitted* 2010.
- [10] Agoshkov V, Quarteroni A, Rozza G: **Shape design in aorto-coronary bypass anastomoses using perturbation theory.** *SIAM J. Numer. Anal.* 2006, **44**:367–384.
- [11] Bressloff N: **Parametric geometry exploration of the human carotid artery bifurcation.** *Journal of Biomechanics* 2007, **40**:2483–2491.

- [12] Kolachalama V, Bressloff N, Nair P: **Mining data from hemodynamic simulations via Bayesian emulation.** *BioMedical Engineering OnLine* 2007, **6**:47.
- [13] McLeod K, Caiazzo A, Fernandez M, Mansi T, Vignon-Clementel I, Sermesant M, Pennec X, Boudjemline Y, Gerbeau J: **Atlas-based reduced models of blood flows for fast patient-specific simulations.** In *Statistical Atlases and Computational Models of the Heart. Proceedings from the First International Workshop, STACOM 2010, and Cardiac Electrophysical Simulation Challenge, CESC 2010, September 20, 2010, Beijing, China*, Springer, Series: Lecture Notes in Computer Science and Engineering, vol. 6364, O. Camara, M. Pop, K. Rhode, M. Sermesant, N. Smith, A. Young (Eds.) 2010:95–104.
- [14] Rozza G, Huynh P, Patera A: **Reduced basis approximation and a posteriori error estimation for affinely parametrized elliptic coercive partial differential equations.** *Arch. Comput. Methods Engrg.* 2008, **15**:229–275.
- [15] Quarteroni A, Rozza G, Manzoni A: **Certified reduced basis approximation for parametrized partial differential equations in industrial applications.** *Journal of Mathematics in Industry (in press)* 2011.
- [16] Holmes P, Lumley J, Berkooz G: *Turbulence, coherent structures, dynamical systems and symmetry.* Cambridge University Press, UK 1996.
- [17] Buhmann M: *Radial Basis Functions.* Cambridge University Press, UK 2003.
- [18] Patera A, Rozza G: *Reduced Basis Approximation and A Posteriori Error Estimation for Parametrized Partial Differential Equations.* Version 1.0, 2006, Copyright MIT, to appear in (tentative rubric) MIT Pappalardo Graduate Monographs in Mechanical Engineering. [Available at <http://augustine.mit.edu>].
- [19] Quarteroni A: *Numerical Models for Differential Problems.* Springer, Series MS&A, Vol. 2 2009.
- [20] Brezzi F, Rappaz J, Raviart P: **Finite Dimensional Approximation of Non-linear Problems. Part I: Branches of Nonsingular Solutions.** *Numerische Mathematik* 1980, **36**:1–25.
- [21] Girault V, Raviart P: *Finite Element Method for Navier-Stokes Equations.* Springer, Series in Computational Mathematics. Vol. 5 1986.
- [22] Nguyen NC, Veroy K, Patera AT: **Certified Real-Time Solution of Parametrized Partial Differential Equations.** In *Handbook of Materials Modeling.* Edited by Yip S, Springer 2005:1523–1558.
- [23] Veroy K, Patera A: **Certified real-time solution of the parametrized steady incompressible Navier-Stokes equations: rigorous reduced-basis a posteriori error bounds.** *International Journal for Numerical Methods in Fluids* 2005, **47**:773–788.
- [24] Quarteroni A, Valli A: *Numerical Approximation of Partial Differential Equations.* Springer-Verlag 1994.
- [25] Rozza G, Manzoni A: **Model order reduction by geometrical parametrization for shape optimization in computational fluid dynamics.** In *Proceedings of ECCOMAS CFD 2010, V European Conference on Computational Fluid Dynamics, Lisbon, Portugal, J.C.F. Pereira and A. Sequeira (Eds.)* 2010.

- [26] Sederberg T, Parry S: **Free-form deformation of solid geometric models.** *Comput. Graph.* 1986, **20(4)**:151–160.
- [27] Botsch M, Kobbelt I: **Freeform Shape Representations for Efficient Geometry Processing.** *International Conference on Shape Modeling and Applications* 2003.
- [28] Lassila T, Rozza G: **Parametric free-form shape design with PDE models and reduced basis method.** *Comput. Meth. Appl. Mech. Engr.* 2010, **199**:1583–1592.
- [29] Dryden I, Mardia K: *Statistical Shape Analysis.* Wiley 1998.
- [30] Morris AM, Allen CB, Rendall TCS: **CFD-based optimization of aerofoils using radial basis functions for domain element parameterization and mesh deformation.** *International Journal for Numerical Methods in Fluids* 2008, **58(8)**:827–860.
- [31] Montagnat J, Delingette H, Ayache N: **A review of deformable surfaces: topology, geometry and deformation.** *Image and Vision Computing* 2001, **19(14)**:1023 – 1040.
- [32] Bookstein FL: **Principal Warps: Thin-Plate Splines and the Decomposition of Deformations.** *IEEE Trans. Pattern Anal. Mach. Intell.* 1989, **11**:567–585.
- [33] Caloz G, Rappaz J: **Numerical Analysis for Nonlinear and Bifurcation Problems.** In *Handbook of Numerical Analysis, Vol. V*, Techniques of Scientific Computing (Part 2). Edited by Ciarlet P, Lions J, Elsevier Science B.V. 1997:487–637.
- [34] Peterson J: **The reduced basis method for incompressible viscous flow calculations.** *SIAM Journal on Scientific and Statistical Computing* 1989, **10(4)**:777–786.
- [35] Quarteroni A, Rozza G: **Numerical solution of parametrized Navier-Stokes equations by reduced basis methods.** *Numer. Methods Partial Differential Equations* 2007, **23(4)**:923–948.
- [36] Deparis S: **Reduced basis error bound computation of parameter-dependent Navier-Stokes equations by the natural norm approach.** *SIAM J. Numer. Anal.* 2008, **46(4)**:2039–2067.
- [37] Deparis S, Rozza G: **Reduced basis method for multi-parameter-dependent steady Navier-Stokes equations: Applications to natural convection in a cavity.** *Journal of Computational Physics* 2009, **228(12)**:4359–4378.
- [38] Rozza G, Veroy K: **On the stability of the reduced basis method for Stokes equations in parametrized domains.** *Comput. Meth. Appl. Mech. Engr.* 2007, **196(7)**:1244–1260.
- [39] Rozza G, Huynh P, Manzoni A: **Reduced basis approximation and error bounds for Stokes flows in parametrized geometries: roles of the inf-sup stability constants.** *Submitted* 2010.

- [40] Barrault M, Maday Y, Nguyen N, Patera A: **An ‘empirical interpolation’ method: application to efficient reduced-basis discretization of partial differential equations.** *C.R. Math. Acad. Sci. Paris, Series I* 2004, **339**(9):667–672.
- [41] Wells DR, Archie JP, Kleinstreuer C: **Effect of carotid artery geometry on the magnitude and distribution of wall shear stress gradients.** *Journal of Vascular Surgery* 1996, **23**(4):667–678.
- [42] Bressloff N, Forrester A, Banks J, Kolachalama V: **Shape optimization of the carotid artery bifurcation.** *Proceedings of ASMO-UK* 2004.
- [43] Lee S, Antiga L, Spence JD, Steinman DA: **Geometry of the Carotid Bifurcation Predicts Its Exposure to Disturbed Flow.** *Stroke* 2008, **39**(8):2341–2347.
- [44] Bharadvaj B, Mabon R, Giddens D: **Steady flow in a model of the human carotid bifurcation. Part I - Flow visualization.** *Journal of Biomechanics* 1982, **15**:349–362.
- [45] Ku D, Giddens D, Zarins C, Glagov S: **Pulsatile flow and atherosclerosis in the human carotid bifurcation. Positive correlation between plaque location and low oscillating shear stress.** *Arterioscler Thromb Vasc Biol* 1985, **5**(3):293–302.
- [46] Ding Z, Wang K, Li J, Cong X: **Flow field and oscillatory shear stress in a tuning-fork-shaped model of the average human carotid bifurcation.** *Journal of Biomechanics* 2001, **34**:1555–1562.