Contributions to scientific computing for data-simulation interaction in biomedical applications

Damiano Lombardi

SORBONNE UNIVERSITÉ

Inria

Contributions to scientific computing for data-simulation interaction in biomedical applications.

Mémoire d'Habilitation à Diriger des Recherches

Damiano Lombardi

Presented on July, the 1st, 2020 in front of the jury:

Prof. Yvon Maday, Sorbonne Université, President

Prof. Fabio Nobile, EPFL, Reviewer

Prof. Karen Veroy, TU Eindhoven, Examiner

Prof. Stefan Volkwein, Konstanz University, Reviewer

Prof. Esther Pueyo, Zaragoza University, Examiner

Dir. Miguel Fernandez, Inria Paris, Examiner

Dir. Jean-Frédéric Gerbeau, Inria DGDS, Examiner

Contents

List of Figures

Author's bibliography.

Ph.D. thesis: [22], *Inverse Problems for tumour growth modeling*, defended in Bordeaux on 09-09-2011.

Journal Articles:

- [4], Bakhta, A., Lombardi, D. (2018). *An a posteriori error estimator based on shifts for positive hermitian eigenvalue problems.* Comptes Rendus Mathematique, 356(6), 696-705.
- [30], Tixier, E., Raphel, F., Lombardi, D., Gerbeau, J. F. (2018). *Composite biomarkers derived from Micro-Electrode Array measurements and computer simulations improve the classification of drug-induced channel block.* Frontiers in physiology, 8, 1096.
- [18], Gerbeau, J. F., Lombardi, D., Tixier, E. (2018). *How to choose biomarkers in view of parameter estimation.* Mathematical biosciences, 303, 62-74.
- [3], Aletti, M., Lombardi, D. (2017). *A reduced-order representation of the Poincaré-Steklov operator: an application to coupled multi-physics problems.* International Journal for Numerical Methods in Engineering, 111(6), 581-600.
- [19], Gerbeau, J. F., Lombardi, D., Tixier, E. (2018). *A moment-matching method to study the variability of phenomena described by partial differential equations.* SIAM Journal on Scientific Computing, 40(3), B743-B765.
- [29], Tixier, E., Lombardi, D., Rodriguez, B., Gerbeau, J. F. (2017). *Modelling variability in cardiac electrophysiology: A moment-matching approach.* Journal of the Royal Society Interface, 14(133), 20170238.
- [13], Ehrlacher, V., Lombardi, D. (2017). *A dynamical adaptive tensor method for the Vlasov–Poisson system.* Journal of Computational Physics, 339, 285-306.
- [2], Aletti, M., Gerbeau, J. F., Lombardi, D. (2016). *A simplified fluid–structure model for arterial flow. application to retinal hemodynamics.* Computer Methods in Applied Mechanics and Engineering, 306, 77-94.
- [1], Aletti, M., Gerbeau, J. F., Lombardi, D. (2016). *Modeling autoregulation in three-dimensional simulations of retinal hemodynamics.* Journal for Modeling in Ophthalmology, 1(1), 88-115.
- [25], Damiano Lombardi, Sanjay Pant. *A non-parametric k-nearest neighbor entropy estimator.* Physical Reviev E, 2016, 10.1103/PhysRevE.93.013310.
- [2], Matteo Aletti, Jean-Frédéric Gerbeau, Damiano Lombardi. *A simplified fluid-structure model for arterial flow. Application to retinal hemodynamics.* Computational Methods in Applied Mechanics and Engineering, 2016, 306, pp.77-94.

- [27], Sanjay Pant, Damiano Lombardi. *An information-theoretic approach to assess practical identifiability of parametric dynamical systems.* Mathematical Biosciences, Elsevier, 2015, pp.66-79.

- [24], Damiano Lombard, Emmanuel Maitre, E. (2015). *Eulerian models and algorithms for unbalanced optimal transport.* ESAIM: Mathematical Modelling and Numerical Analysis, 49(6), 1717-1744.

- [17], Jean-Frédéric Gerbeau, Damiano Lombardi, Elisa Schenone. *Reduced order model in cardiac electrophysiology with approximated Lax pairs.* Advances in Computational Mathematics, Springer Verlag, 2014, pp.28.

- [16], J.-F. Gerbeau, D.Lombardi. *Approximated Lax pairs for the reduced order integration of nonlinear evolution equations*, Journal of Computational Physics, 265; 246:269, 03-2014

- [21], Angelo Iollo, Damiano Lombardi. *Advection modes by optimal mass transfer.* Physical Review E : Statistical, Nonlinear, and Soft Matter Physics, American Physical Society, 2014, 89, pp.022923.

- [10] Thierry Colin, Francois Cornelis, Olivier Saut, Patricio Cumsille, Damiano Lombardi, et al.. *In vivo mathematical modeling of tumor growth from imaging data : soon to come in the future?.* Diagnostic and Interventional Imaging, Elsevier, 2013, pp.571-574.

- [23], Damiano Lombardi. *Inverse problems in 1D hemodynamics on systemic networks: a sequential approach.* International Journal for Numerical Methods in Biomedical Engineering, John Wiley and Sons, 2013

- [8], Damiano Lombardi, Angelo Iollo, Thierry Colin, Olivier Saut. *Inverse Problems in tumor growth modeling by means of semiempirical eigenfunctions.* Mathematical Models and Methods in Applied Sciences, World Scientific Publishing, 2012, pp.1.

- [20], Damiano Lombardi, Angelo Iollo. *A Lagrangian Scheme for the Solution of the Optimal Mass Transfer Problem.* Journal of Computational Physics, Elsevier, 2011, 230, pp.3430:3442.

Book section:

- [5], Michel Bergmann, Thierry Colin, Angelo Iollo, Damiano Lombardi, Olivier Saut, et al.. *Reduced Order Models at work.* Quarteroni, Alfio. Modeling, Simulation and Applications, 9, Springer, 2013.
- [9], T.Colin, A.Iollo, D.Lombardi, O.Saut, F.Bonichon and J.Palussire: *Some models for the prediction of tumor growth: general framework and applications to metastases in the lung*, Computational Surgery, Springer 2012.

Scientific popularization:

- [7], Thierry Colin, Angelo Iollo, Damiano Lombardi, Olivier Saut. Prediction of the Evolution of Thyroidal Lung Nodules Using a Mathematical Model. ERCIM News, ERCIM, 2012, 82, pp.37-38.

In revision/preparation:

- [14], Ehrlacher, V., Lombardi, D., Mula, O., Vialard, F. X. (2019). *Nonlinear model reduction on metric spaces. Application to one-dimensional conservative PDEs in Wasserstein spaces.* arXiv preprint arXiv:1909.06626.
- [12], Ehrlacher, V., Grigori, L., Lombardi, D., Song, H. (2019). *Adaptive hierarchical subtensor partitioning for tensor compression.* HAL:https://hal.inria.fr/hal-02284456
- [26], Lombardi, D., Raphel, F. (2019). *A greedy dimension reduction method for classification problems.* HAL:https://hal.inria.fr/hal-02280502
- [28], Raphel, F., de Korte, T., Lombardi, D., Braam, S., Gerbeau, J. F. (2019). *A greedy classifier optimisation strategy to assess ion channel blocking activity and pro-arrhythmia in hiPSC-cardiomyocytes.* HAL:https://hal.inria.fr/hal-02276945
- [11], Alfonsi, A., Coyaud, R., Ehrlacher, V., Lombardi, D. (2019). *Approximation of Optimal Transport problems with marginal moments constraints.* arXiv preprint arXiv:1905.05663.
- [15], Galarce, F., Gerbeau, J. F., Lombardi, D., Mula, O. (2019). *State estimation with nonlinear reduced models. Application to the reconstruction of blood flows with Doppler ultrasound images.* arXiv preprint arXiv:1904.13367.
- Lombardi, D. *Fast state estimation in time dependent systems: tensors and Kalman filtering.*
- Lombardi, D., Maday, Y. and Uro, L. *A reduced basis approach for aided segmentation of facial muscles.*

- [6], Cances, E., Ehrlacher, V., Gontier, D., Levitt, A., Lombardi, D. (2018). *Numerical quadrature in the Brillouin zone for periodic Schrodinger operators.* arXiv preprint arXiv:1805.07144, in revision.

- Ratelade, J., Klug, N., Kaelle, M., Cruz, S. Debertrand, F., Domenga, V., Salman, R., Smith, C., Gerbeau, J.-F., Nelson, M., Joutel, A. *Reducing hypermuscularization of the transitional segment between arterioles and capillaries protects against spontaneous intracerebral hemorrhage*, in revision.

Introduction.

My research activity deals with scientific computing. In particular, the goal is to investigate numerical methods and algorithms to make data-simulation interaction feasible and efficient.

Two main axes can be identified: one, methodological, the other applied to the biomedical engineering context. My Ph.D. thesis work was devoted to the study of inverse problems for tumor growth modeling. The focus has, since my Post Doc, moved to the applications concerning the cardiovascular system. In general, the problems in engineering are primarily related to the ability of performing reliable estimations of quantities of interest by respecting the constraints of the scenario at hand. The predictions have to be formulated based on the data collected from the system, through an observation process. How to perform such an estimation is the object of Data Assimilation.

Several methodological issues arise when dealing with realistic Data Assimilation problems. They can be broadly divided into two classes: the often prohibitive computational cost and the need of accounting for the uncertainties that potentially impact the estimation. This thematic subdivision is reflected into the structure of the present manuscript, which is divided into three main chapters. Their content and the main contributions in each of the sub-topics are summarized hereafter.

Data Assimilation

In the first Chapter, several works in Data Assimilations are presented, in various contexts. In the current manuscript, only the study cases in which realistic data were used are presented. The Chapter ends with the identification and the synthetic description of the difficulties that these problems pointed out.

Inverse problems in tumour growth modeling. A model of avascular tumour growth has been set-up, as a trade-off between simplicity and ability to account for the main physical aspects of the proliferation of tumour cells inside a tissue. The macroscopic model based on mixture theory reduces to a set of parametric Partial Differential Equations. The goal consists in estimating the parameters and the initial conditions of this system in order to account, at best, for a given sequence of CT scans. This is achieved by using a variational formulation in which the discrepancy between the available measurements and the model output observation is minimised. A regularisation based on Proper Orthogonal Decomposition is used to reduce the computational cost and palliate the ill-conditioning of the system. The so-obtained calibrated model can then be used for forecast. Several realistic test cases were considered, in collaboration with Institut

Bergonié (an oncology institute in Bordeaux) in different regimes (fast growth, slow growth). These tests were helpful in highlighting the advantages and the limitations of such an approach. This case is an example of a Data Assimilation problem in which data are quite rich in space (even though this might strongly depend on the image resolution) but scarce in time. The main results are presented in Section 1.1.

Problems in haemodynamics. Three main problems are presented in Section 1.2. The first application consists in estimating the arterial stiffness by using, as much as possible, non-invasive measurements. This could be useful, for instance, when considering continuous monitoring of hypertensive patients. In the clinical practice, the stiffness is indirectly inferred real time by estimating the Pulse Wave Velocity (PWV) and using the Moens-Korteweg formula. The measurements that we considered are mainly pressure and flow signals in time, taken at some points of the vascular network (mainly in the periphery). It is an example of a Data Assimilation problem in which the measured quantities are quite well resolved in time but not in space. The nature of the measurements and the need of having a fast estimation make us consider sequential approaches. A 1D model of the 55 main arteries of the human body is introduced, and an Unscented Kalman Filter (UKF) is investigated to perform the joint parameter-state estimation. Several contributions are proposed: a comparison with PWV estimation of the aorta stiffness shows that sequential approaches are more precise; moreover, a first analysis of the sensitivity of the estimation to random pertubation of the system parameters is described, to get an assessment of the robustness. This contribution is presented in Section 1.2.1 and a more complete description can be found in [23].

The second problem studied is the mathematical modeling of the retinal haemodynamics. Retina is one of the best windows on micro-circulation and it is part of the central nervous system. In this the autoregulation phenomenon plays a key role. A fluid-structure interaction model has been proposed to account for the fluid-structure interaction in the large arterioles and the coupling between several compartments of the eye. The model described in Section 2.1.1 has been integrated with a model describing the smooth muscle cells and a simple control based on the local pressure. The goal is to be able to reproduce the behaviour of the system and fit measurements of type velocity/diameter on the retinal network. The model parameters have been calibrated by considering a simple optimisation based on several instances computed. The simulations described in Section 1.2.2 show that the model, albeit simplified, is able to reproduce, the main features of the retinal flows. This work is the object of [1].

The third problem presented, which is still an ongoing work, deals with the fast reconstruction of 3D flows in arteries, based on ultrasound Doppler measurements. They consists in a set of voxels in which the average (on the voxel) velocity in the direction of the beam is estimated. The method investigated is on optimal reconstruction, which is a way of regularising the state estimation problem by using a semi-empirical basis. This allows to significantly speed up the estimation. The contribution presented in Section 1.2.3 is related to the construction of an efficient basis to deal with a semi-realistic configuration and it is extensively presented in [15].

Problems in cardiac electrophysiology. The first contribution is the study of a population stochastic inverse problem. The prior distribution of the parameters of a model is sought to

fit statistics of the model output measured on a population. In Section 1.3.1 (and in the work presented in [29]) a numerical experiment is proposed, which has been constructed based on available realistic data on human atrial cells. We computed the probability density distribution of the parameters impacting the most the action potential in a model for atrial cells (Courtemanche-Ramirez), for two distinct sub-populations. The first population is composed of healthy individuals, the other by individuals affected by a disease. The results obtained are quite encouraging. In particular, the prior probability density distributions of the parameters are quite different in the two populations, their statistical moments show significant differences. These differences are large for the parameters which are supposed to be correlated to the pathology.

The second problem, presented in Section 1.3.2 and in [28], is motivated by applications in cardio-toxicity, one of the branches of the safety pharmacology. The main goal is to determine, based on data collected on a small in vitro tissue, whether a molecule, candidate to become a drug, have a negative and potentially dangerous impact on the cardiac activity. This is presented in Section 1.3.2. In particular, data coming from a Micro-Electrodes-Array are available both in control case and when molecules are added at different concentrations. This device records a number of electrograms which are called Field Potentials (FP). Contrary to the Action Potentials, they do not give direct information on the activity of a single cell, but they are representative of the overall activity of the small tissue living inside the micro-chip well. Henceforth they have to be related to the ionic activity of each individual cell. The method proposed, that worked with realistic data, is intermediate between an inverse problem and a machine learning approach. A population of *in silico* models is computed and it is used to augment the in vitro dataset. A classifier is then trained in order to answer questions about the eventual blockade of the three main ion channels (Sodium, potassium, calcium). The results obtained are encouraging.

Reduced-Order models and high-dimensional problems.

One of the main difficulties in performing Data Assimilation when the mathematical models are systems of parametric Partial Differential Equations is the computational cost, which makes realistic applications out of reach. The cost come from the fact that the objects playing a role in the assimilation process are high-dimensional. Several topics have been investigated, which are presented hereafter.

Reduced-Order models for Fluid-Structure interaction. This work is motivated by the simulation of the haemodynamics of the human eye. Two main contributions are proposed. The first one consists in a model to describe fluid-structure interaction in the cases in which the structure is thin and does not have a large resistance to bending. It has been developed primarily to model large arterioles. The physics of the system, in this case, makes it possible to simplify the equations for the structure and to treat it as a generalised Robin boundary condition for the fluid problem, which leads to a significant reduction of the computational cost. This is presented in Section 2.1.1 and a complete derivation can be found in [2].

The second problem arises when considering several sub-systems or compartments interacting together. In many scenarios, it may happen that we are interested in studying the behaviour

of one of the sub-systems, described for instance by a non-linear PDE model. The couplings oblige to consider also the other compartments and, in this work, we make the hypothesis that their behaviour is more simple and their dynamics is described by systems of linear time independent PDEs. In this case, the interaction between the sub-system of interest and the other compartments can be described by means of the Poincaré-Steklov operator. Despite the fact that this operator is defined only on the boundary between sub-systems, when discretised, it leads to full matrices, and hence, to large storage and computational cost (either to be computed and applied). To reduce the computational cost, a reduced basis with online update is proposed to build a low rank approximation of the Poincaré-Steklov operator. This contribution is presented in Section 2.1.2 and in [3].

Optimal Transport. It is a problem that finds application in a wide spectrum of fields, ranging from imaging to economics. Two main contributions are proposed: the first one is on the definition of numerical methods to approximate the solution of optimal transport, the second one, on a possible way of obtaining an unbalanced optimal transport.

Concerning the study of numerical methods, during my Ph.D. thesis I worked on a Lagrangian numerical approximation of the problem in the Benamou-Brenier formulation. The key observation is that the Lagrangian trajectories of the fluid particles are known, and they are straight lines. A particle method and a multi-level approach allowed to treat large 3D cases. Recently, a work has been proposed in which an approximation of the Kantorovich formulation of the multi-marginal optimal transport is defined. This formulation could lead to a numerical method suitable for a large number of marginal problems and with non-standard constraints (such as martingale constraints). These methods are presented in Section 2.2.1 and are the object of [20, 11].

The second contribution, studied during my Post-Doc, is related to a possible extension to the Benamou-Brenier formulation of optimal transport to account for unbalanced densities. Indeed, the classical optimal transport problem is well defined when the marginal densities have the same mass. A possible extension of the interpolation between two densities has been investigated by modifying the continuity constraint in the Benamou-Brenier formulation. The sources model considered are such that several properties of the McCann interpolation are preserved. In certain cases, it is possible to show an equivalence between these formulations and a balanced optimal transport with non-uniform metric; this observation leads also to the possibility of proving that the solution exists and it is unique. These contributions are presented in Section 2.2.2 and the details of the derivations are proposed in [24].

Reduced-Order models for advection-dominated solutions. In the cardio-vascular system, as in a considerable number of engineering applications, the solutions of the system under investigation are featured by traveling or progressive waves, advection, front propagation. All these phenomena are a challenge for the classical projection based Reduced-Order modeling methods. After a synthetic review of the works proposed in the literature, two contributions are presented in Section 2.3.

The first one is related to the use of distances different from the classical L^2 one. The first work in this respect, described in [21] has been proposed during my Ph.D. thesis. It consists in

using Optimal Transport to construct a set of mappings that transform the snapshots; they are transported into a same barycentric reference configuration. By doing so, the approximation obtained by classical Proper Orthogonal Decomposition is much more efficient. In a more recent work (that can be found in [14]) a different approach is proposed. We consider the snapshots as the elements of a Riemannian metric space. The exponential and logarithm maps on the manifold make it possible to define a reduced approximation that turns out to be more efficient than the one obtained by using simply the L^2 distance. The framework is particularised for the Wasserstein distance, as an example, an tested on 1D parametric PDEs. These works are discussed in Section 2.3.1.

The second axis consists in the study of a dynamical basis method, based on a numerical analogous of Lax-pairs. These make it possible to represent in a very elegant way some solutions of non-linear PDEs, and in particular PDEs whose solution is featured by the traveling interacting waves. In the proposed method, a linear self-adjoint, continuous, inverse compact operator is defined, depending upon the solution itself. We make the arbitrary choice of restricting to Schrödinger type of operators. The evolution of the eigenvalues and the eigenfunctions makes it possible to define a discretisation for the system. When the dynamics reduces to a linear transport, the basis follows exactly the dynamics of the system. The method has been tested in challenging scenarios and produces encouraging results. This method is presented in Section 2.3.2 and in [16, 17].

Tensor methods. When dealing with high-dimensional problems an a priori way of seeking an approximation is represented by the separation of variables principle. Tensor methods have a long history and are now a topic of great interest due to their application in a number of different fields. A first work on the use of tensor methods to solve a high-dimensional Partial Differential Equations (Vlasov-Poisson) is presented in Section 2.4.1 and in [13]. The main objective is to build a parsimonious discretisation starting from heterogeneous separated discretisations. Moreover, the tensor rank (related to the size of the approximation) is not fixed a priori; instead, it is determined in order to fulfill a prescribed error criterion. The Vlasov-Poisson test case is also useful to see if it is possible to construct, with tensors, a discretisation which is preserving the geometric structure of the problem: a symplectic time discretisation is considered, that helps in preserving, at a certain order, the Hamiltonian structure of the system. The results obtained are encouraging; however, the standard format used is not always suitable to represent the solution and the rank tend to significantly increase, making the method less efficient. This motivates the current methodological developments (which are the object of an ANR project, ADAPT), presented in Section 2.4.2. A first contribution is described (details can be found in [12]): a piece-wise adaptive tensor decomposition is devised to deal with the approximation of moderate order tensors. A Proposition is proved in which a sufficient condition on the functions regularity is shown, such that the decomposition converges to the function to be approximated. The numerical tests showed that the compression achieved is quite relevant. This can be seen as a first step towards the set up of parsimonious discretisations for high-dimensional PDEs.

Dealing with uncertainty.

Providing an estimation is inevitably associated to the ability of dealing with the uncertainties that affect the system description and the measurements. This is key to get reliability. The first studies proposed in the literature address the problem of Forward Uncertainty Quantification: given a parametric system, the parameters being distributed according to a known probability density distribution, the model outcome distribution (or some statistics) are sought. This problem is pertinent to study the system nature and to assess the robustness of a parameter estimation, for instance, after a posterior distribution is found as a result of a Bayesian inverse problem. Several other classes of problems were introduced and studied in the literature, related to the uncertainty. In this manuscript we do not consider the problem of Forward Uncertainty Quantification. Three contributions are presented, which are related to other problems relevant in the Data Assimilation context.

Identifiability. When a parameter, state, or joint parameter-state estimations are performed a simple yet fundamental mathematical question reads: is the estimation possible? This is the object of identifiability studies. A practical identifiability notion is introduced, based on Information theoretic quantities, which is well adapted to the Bayesian framework. Moreover, a notion useful to make practical sense of an expected gain in information is proposed. In this work, presented in Section 3.1, a classical entropy and mutual information estimator are used. However, when observables are the result of the measurement of a parametric model, the estimators tend to degrade their performances. This fact motivated the development of a modified estimator for the differential entropy, that turned out to be more adapted to the scenarios encountered in parameter estimation than the classical one. This is presented in Section 3.1. These contributions are detailed in [27, 25]

Backward Uncertainty Quantification. One of the problems which are relevant is the estimation of the parameter density distribution based on statistics of a model output on a population of individuals or a set of experiments. This stochastic inverse problem is referred to as Backward Uncertainty Quantification. Its two main applications are: the estimation of the densities of parameters to evaluate the natural variability of biophysical phenomena; the set up of priors to be used in Bayesian estimation. This problem turns out to be, as most inverse problems, potentially ill-conditioned and computationally expensive. An approximation is proposed, called OMM, that consists in finding the probability density distribution of the parameters such that the model outputs statistical moments match the observed ones in a population. An entropy regularisation is added. Moreover, in order to make the problem better conditioned, a strategy is proposed to select a subset of the degrees of freedom and make the problem feasible from a computational point of view. This work is presented in 3.2 and it is detailed in [19].

Composite biomarkers: towards learning-simulation interaction. In most of the applications in biomedical engineering, some quantities or markers are extracted from the observations, to make predictions. These are based on the insight of the community. A numerical method is proposed to build a correction, in a semi-empirical way, to the proposed biomarkers. The outcome of the method makes it possible to perform classification and regression tasks in an

efficient way. This highlight two main persepectives: the first one is the use of composite biomarkers as a non-linear preconditioner for inverse problems; the second one is a possible way of using machine learning tools and numerical simulations in a collaborative framework that defines hybrid strategies for data assimilation. These contributions are presented in Section 3.3 and in [18, 26].

Chapter 1

Data assimilation.

Context

The problems to be faced in applications can be broadly seen as finding a way to predict a system configuration, optimise its evolution and design its interactions with the environment it operates into. Aiming at providing a quantitative description of a system, we introduce a *model*[1]. The *state* is the set of quantities that are necessary and sufficient to describe its configuration in a non-ambiguous way. Having a perfect knowledge of the system state would make it possible to compute some *Quantities of Interest* (QoI), in relation to the application under scrutiny. The knowledge about the system state is acquired through measurements. In most of the realistic scenarios, however, it is impossible to completely measure the state or the QoI directly; instead, a subset of quantities, the *observables*, are measured. Otherwise stated, measurements are always partial and imperfect (corrupted by noise). The model can be seen as a way, systematic but affected by some errors, of integrating the *a posteriori* knowledge coming from the *data* with an *a priori* knowledge; it is a rigorous link between the observables and the QoI. In this respect, data assimilation methods are the mathematical tools to exploit at best this link. In the rest of the chapter, several contributions in data assimilation applied to semi-realistic applications are described. At the end of this chapter, a partial conclusion is presented, that motivates the methodological investigations presented in the subsequent chapters of this manuscript.

1.1 Tumour growth modeling

Cancer is a very diffuse and diverse disease affecting potentially all the tissues of the body. Its nature is intrinsically multi-scale: a tumour is generated by genetic mutations causing abnormalities that induce an evolutionary advantage, it starts growing and interacting with the tissue at a macroscopic scale, it can spread at the rest of the body (metastasis). In the work presented hereafter only solid tumours are considered. Several models were proposed in the literature to describe the different stages of the tumour growth. They can be broadly divided into two categories: agent based models, that try to model each cell, individually; models based on systems of

[1]This is rather general, the model can be derived on the basis of first principles or constructed in a more agnostic way.

parametric PDEs, which provide a description of the tissue evolution at a macroscopic scale. In the current clinical practice, the available data used to assess the presence and the progression of the pathology are of different kind: the most commonly used is medical imaging. When a solid tumour is discovered, it is often of more than 1 mm in size, meaning that it could be already in the vascular stage, interacting with the tissue and asking the body to build a vascular network to be fed. In [10, 8, 9, 5], an ensemble of contributions is proposed to perform data assimilation based on medical imaging. It is a set of works trying to understand if some applications can actually be performed and which are the limitations of such an approach. Several points are studied therein:

1. A PDE advection-reaction-diffusion model is designed to simulate the avascular growth of lung nodules with a moderate number of free parameters.

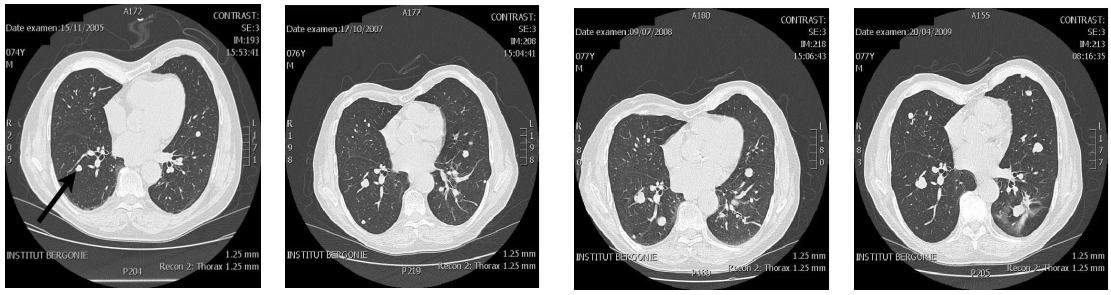
2. A Proper Orthogonal Decomposition regularised parameter estimation method is introduced to estimate the parameters starting from medical images.

3. A comparison to realistic data coming from patients, in different growth regimes is proposed.

All the details are provided in [22]. The model proposed reads as follows: let $\Omega \subset \mathbb{R}^d$ be an open bounded set. Let $x \in \Omega$ be a space point, the time $t \in [0, T]$; let $P(x,t), Q(x,t) \geq 0$ be the densities of proliferating and quiescent cells respectively. Let $\mathbf{v}$ be the velocity field, and $\Pi(x,t)$ a pressure; the oxygen concentration in the tissue is denoted by $C(x,t)$. The system is written:

$$\partial_t P + \nabla \cdot (\mathbf{v}P) = (2\gamma - 1)P, \tag{1.1}$$
$$\partial_t Q + \nabla \cdot (\mathbf{v}Q) = (1 - \gamma)P, \tag{1.2}$$
$$\nabla \cdot \mathbf{v} = \gamma P, \tag{1.3}$$
$$\mathbf{v} = -\kappa(P, Q)\nabla\Pi, \tag{1.4}$$
$$\kappa = \kappa_0 + \kappa_1(P + Q), \tag{1.5}$$
$$-\nabla \cdot (\eta(P, Q)\nabla C) = -\alpha PC - \beta C, \tag{1.6}$$
$$\eta = \eta_0 - \eta_1(P + Q). \tag{1.7}$$

The initial and boundary conditions are determined according to the setting under consideration. The parameter $\gamma \in \mathbb{R}^+$ is a proliferation rate, the permeability $\kappa \in \mathbb{R}^+$ as well as the oxygen diffusivity $\eta \in \mathbb{R}^+$ depend upon the variables P, Q. The system state is the set of variables (P, Q, Π, C). Considering that the measurement is performed by CT scan (whose gray scale is related to the tissue density) the observable is the variable $Y = P+Q$. The rationale behind this choice is simple: CT scan images gray scale depends upon the absorption properties of the tissues and since tumour cells (both proliferating and quiescent) have a larger density with respect to the healthy tissue, what is seen on the image is roughly proportional to the concentration Y. The objective is the following: starting from a sequence of available images, *i.e.* $Y(x, t_1), \dots Y(x, t_n)$, determine the unknown initial conditions for the distribution of proliferating and quiescent cells inside the tumour, and the parameters appearing in the definition of permeability and diffusivity, in order to account, at best, for the observations. Once the model has been calibrated, it makes it



(a) (b) (c) (d)

Figure 1.1: Sequences of images, testcase presented in Section 1.1.

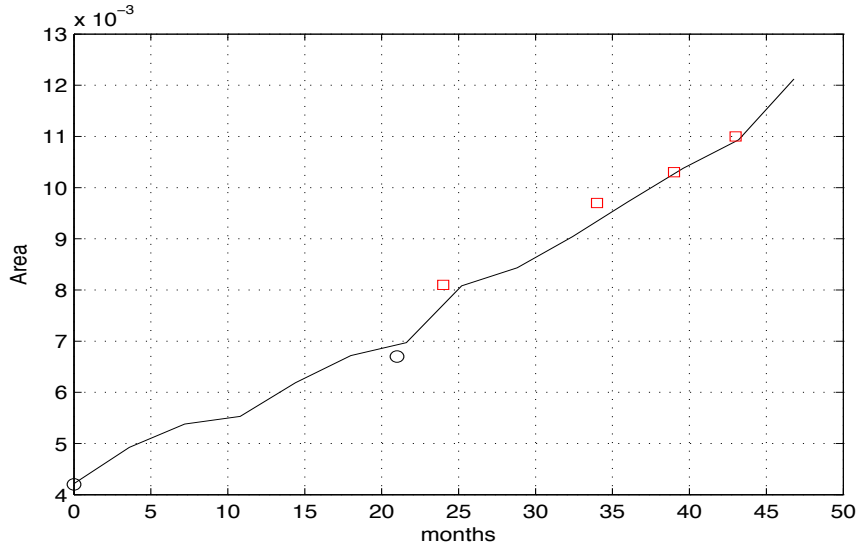


Figure 1.2: Evolution of a lung nodule, with slow dynamics, over 45 months. Test case presented in Section 1.1

possible to perform a forecast on the solid tumour evolution. The assumptions and the methods investigated are reported in details in [22]. We describe, here, a realistic testcase.

The (slow) evolution of a lung nodule (secondary tumour) is shown in Figure 1.1. A set of images on 45 months is available. The experiment consists in using the two first images of the sequence to performe the parameter estimation and use the rest of the sequence as a validation set. Using two images means restring to the least possible amount of information available.

In Figure 1.1, the result is shown in terms of area (in the image plane) as function of time. In the case of slow tumour growth, using two images can be enough to have a reasonable forecast. However, as observed in [22], there are cases in which more information is needed. In general, several methodological questions arise, namely about the parameters identifiability, the uncertainties affecting the system, the reliability of the forecast.

1.2 Problems in haemodynamics

Systemic circulation is a complex multi-physic, multiscale system and the cardiovascular haemodynamics can be described at different levels (a comprehensive presentation can be found in [FQV10]). In particular, a hierarchy of mathematical model is defined, that could be useful to address the description of the haemodynamics in different contexts.

1. *Circuit analogy*: these are systems of Ordinary Differential Equations (0D models), well suited to describe the global behavior of flow and pressure, in multiple points of the systemic vascular network, as function of time: $Q(t), P(t)$, where Q is the flow and P is the pressure. In these models the flow plays the role of the current and the pressure plays the role of the voltage.

2. *Hyperbolic network*: 1D space-time Partial Differential Equations models, adapted to describe pressure waves through a reduced fluid-structure interaction model. The unknowns are flow and pressure along the axial coordinate of the blood vessels $Q(x,t), p(x,t)$. The resulting equations for a single blood vessel are an hyperbolic system in subcritical regime. The couplings between different vessels is performed by considering the equations of the characteristics, mass and momentum conservation.

3. *3D Fluid-Structure Interaction* (FSI): 3D incompressible Navier-Stokes equations to describe the blood flow, coupled with non-linear elasticity to describe the vessel wall dynamics. It is the most realistic description of the mechanical behaviour of the blood vessels. The classes of models presented above can be derived by the 3D models by making some simplifying assumptions.

The use of one class of model depends upon the questions under investigation. In this Sections, some contributions are presented in which these three classes of models are used (often in combination) to describe different subsystems of the arterial circulation, and perform data assimilation.

1.2.1 A sequential approach for the systemic circulation

Due to its prohibitive computational cost, the 3D FSI description of the whole arterial network is, at present, out of reach. In some clinical applications, as, for instance, the continuous monitoring of hypertensive patients, it is important to estimate the mechanical state of the blood vessels (especially close to the heart), by exploiting non-invasive or micro-invasive measurements taken at the periphery of the vascular network. This inevitably involves the need to describe the whole circulation or, at least, a large portion of it. To this end, 0D and 1D model need to be used. In this work, we use the most simple model proposed in the literature: let, for each vessel, x be the axial coordinate, the unknowns are the cross sectional area of the blood vessels $A(x,t)$ and

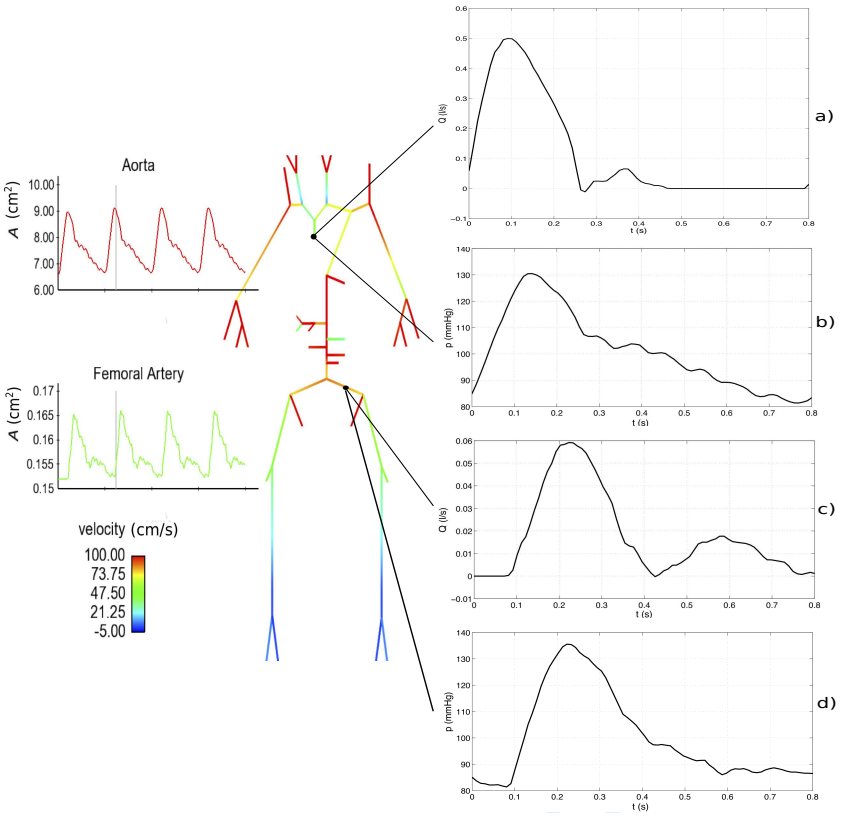


Figure 1.3: Example of a simulation of the 55 main arteries of the human body, Section 1.2.1.

ve average in the section blood velocity $u(x,t)$. The resulting system of equations reads:

$$\partial_t A + \partial_x(Au) = 0, \tag{1.8}$$

$$\partial_t u + \partial_x\left(\frac{u^2}{2} + \frac{p}{\rho}\right) = -\kappa\frac{u}{A}, \tag{1.9}$$

$$p = p_e + \beta\left(A^{1/2} - A_0^{1/2}\right). \tag{1.10}$$

In this system p_e represents the external pressure, A_0 the cross-sectional area at rest, $\rho = 1050\text{kg/m}^3$ is the blood density, $\kappa = 9\cdot 10^{-6}\text{m}^2/\text{s}$ accounts for momentum diffusivity and β is related to vessels elasticity. All the details are commented in [23]. The discretisation of this model is perfomed by using a Taylor-Galerkin finite element method, with a number of degrees of freedom $\mathcal{N} \approx 8\cdot 10^3$. At the bifurcations, mass and momentum continuity, and equations for the characteristics, provide the coupling conditions between adjacent vessels. The outlet boundary conditions are classical 3–elements Windkessel models. Overall, the model has approximately 130 free parameters. In Figure 1.3, an example of a simulation is provided, of the 55 arteries of the human body.

The available data are typically time sampled flow and pressure signals, taken in several peripheral points of the network. The nature of the data and the need to provide a fast estimation

motivated us to study sequential data assimilation methods. The contribution is threefold:

1. An Unscented Kalman filter is set up and applied to the 1D model of the 55 main arteries of the human body.

2. Several semi-realistic test cases are proposed, in which the stiffness of the different segments of the aorta is estimated.

3. The sensitivity of the estimation with respect to the parametric uncertainty is studied.

The scenarios considered are semi-realistic, data used are both synthetic and experimental (Only few experimental signals are available). The specific settings are provided in [23]. The method used to perform data assimilation is one of the non-linear extensions of the Kalman filter, the UKF.

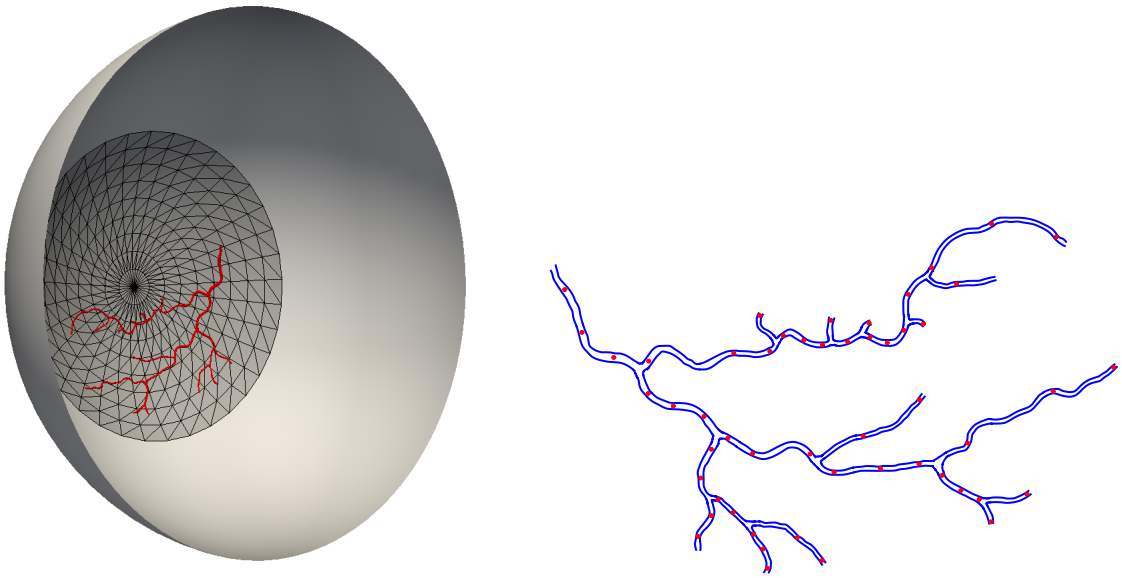
One of the outcomes of the proposed model is to assess the precision of the techniques which are currently used to estimate the stiffness of the blood vessels. These are currently based, in the clinical practice, on the Pulse Wave Velocity (PWV), which is a measure of the average speed of the pressure waves in the vascular network. In particular, a standard test consists in recording the pressure signals in points of the femoral and carotid arteries, and estimate an average speed by evaluating the wave shift. This is one of the criteria used to assess the severity of the hypertension in patients. The estimations performed by using the 1D model are more precise if compared to the estimations obtained by measuring the Pulse Wave Velocity and computing the stiffness by means of the Moens-Korteweg formula (which relies on too restrictive geometric assumptions). Furthermore, one of the experimental observations is that, in a same patient, there are significant oscillation of the measured value of PWV during the day. The model used provides some insight and an interpretation of these oscillations in relation to the daily routine.

The use of 1D model could henceforth improve the stiffness estimation with respect to the currently used technique. However, after evaluating the sensitivity to random perturbations, the conclusion of this study is that, in general, a small number of measurements at the periphery, even if quite well resolved in time, are not enough to estimate the aorta stiffness in a robust way.

If the measurements available are rich enough (a concept to be well defined and quantified), for example by fusing multiple sources of information together, sequential approaches seem to be well suited to the estimation of hidden quantities in a time compatible with the clinical practice. The main difficulty to be circumvented is, therefore, to understand how many measurements and where they should be taken in order to reliably estimate the QoI. A step towards this direction is proposed in Section 3.1.

1.2.2 Retinal haemodynamics

Retina is part of the central nervous system, and an exceptional window on the micro-circulation: it can be easily imaged by standard optical techniques, such as fundus cameras. In [1], a model describing the mechanical behaviour of large arterioles (diameter $d > 40$ μm) is proposed. The fluid-structure interaction model used for this study is presented in Section 2.1.1 and in [2]. The main features of the model of retinal circulation are the following:



(a) (b)

Figure 1.4: Geometry of large arterioles, for the simulations discussed in Section 1.2.2: a) global view of the mesh; b) map of the control points for the comparison with the experiment detailed in [RRG$^+$86].

1. A 3D simplified fluid-structure interaction is built on a realistic geometry segmented by fundus camera.
2. A feedback control is added aiming at modeling autoregulation.
3. The actuation of the control is rendered through the addition of a model of Smooth Muscle Cells.

The geometry of the retinal network used is presented in Figure 1.2.2.a), and it is obtained by segmenting a retinal image acquired by means of a standard 40° Field Of View fundus camera. In Figure 1.2.2.a) the points in which the velocity (and the flow) are monitored is shown. The direct numerical simulations are obtained by discretising the model presented in in [1] by means of P1-P1 SUPG stabilised finite elements. The number of degrees of freedom is about $6 \cdot 10^5$. The parameters of the model are discussed in [1] and they are tuned in order to be representative of a physiological range. In this first work, the experiment presented in [RRG$^+$86] is used as a validation. In their work, the authors studied experimentally the distribution in the network of the flow as a function of the diameter of the arterioles.

This comparison to real data is a first form of weak validation for the models used and show how data assimilation can be a valuable tool for mathematical modeling. The main sources of error, commented in [1], are due to the optical distortion in the periphery of the field of view (leading to an overestimation of the vessel diameters for the distal portion of the network) and on the fact that, when the diameter is small (normally $d \lesssim 40$ μm), non-Newtonian effects could start playing a role in the blood dynamics.

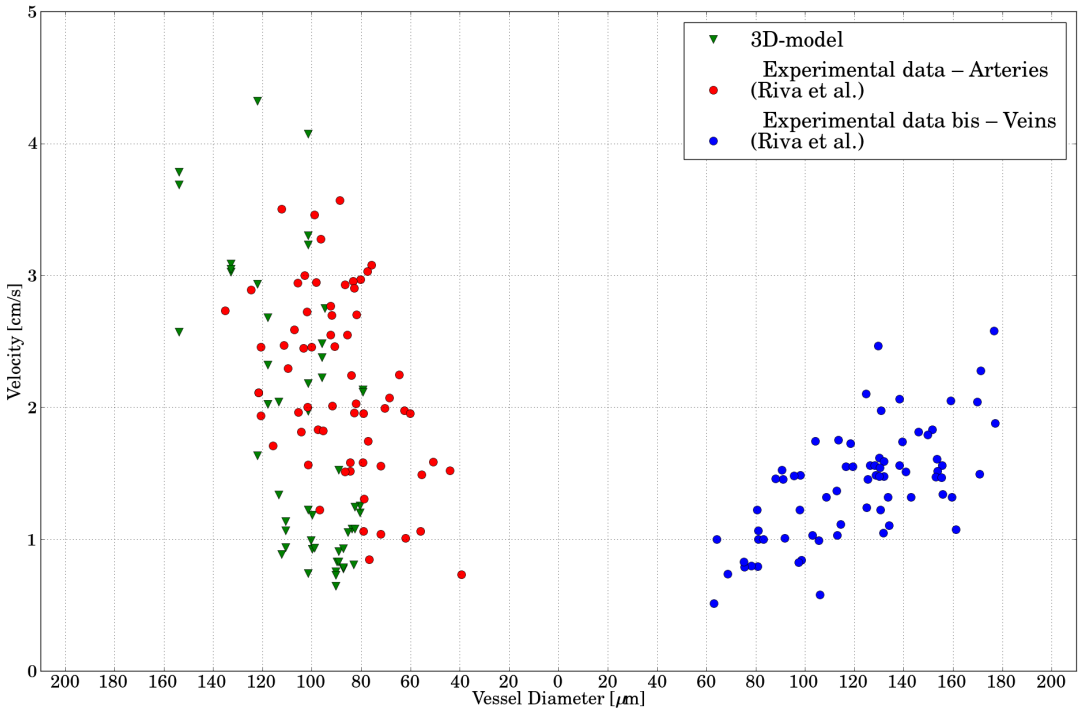


Figure 1.5: Comparison to the experimental data, Section 1.2.2: the green markers represent the values extracted in the control points for the 3D model, to be compared to the red markers obtained experimentally.

This model has been recently adapted in an ongoing work, to provide a mathematical and physical interpretation of some hypotheses on the occurrence of micro-hemorrhages in the vascular network of the central nervous system.

1.2.3 State estimation from ultrasound measurements

Doppler ultrasound imaging is one of the most commonly used techniques to measure the blood velocity in a non-invasive way.

The reconstruct of a 3D blood velocity field on a human carotid artery from Doppler ultrasound images is proposed. The images are synthetically generated (even if based on a realistic geometry) and the use of data from real patients is deferred to a future work. The main goal of the contribution is twofold:

1. A reduced-order optimal reconstruction method is introduced, that can be seen as a non-linear extension to the Parametrised Background Data Weak (PBDW) method. The main novelty is a partition of the solution set $\mathcal{M} = \cup_{k=1}^{K} \mathcal{M}^{(k)}$ which we exploit to build reduced spaces $V_n^{(k)}$ for each $\mathcal{M}^{(k)}$.

2. Second, a non trivial application is proposed, the reconstruction performance between the classical PBDW and its non-linear extensions are proposed. These are the first step towards realistic applications

At every time $t \in [0, T]$, we are given a Doppler ultrasound image that contains information on the blood velocity on a subdomain of the carotid. From the image, we extract the observations $\ell_i(u)$ that we will use to build a complete time-dependent 3D reconstruction of the blood velocity in the whole carotid Ω.

Depending on the technology of the ultrasound device, there are two different types of velocity images. In most cases, ultrasound machines give a scalar mapping which is the projection of the velocity along the direction n of the ultrasound probe. This mapping is called color flow image (CFI). In more modern devices, it is possible to get a 2D vector flow image corresponding to the projection of the velocity into the plane. This mapping is called vector flow image (VFI).

In the following, we work with an idealized version of CFI images. For each time t, a given image is a local average in space of the velocity projected into the direction in which the ultrasound probe is steered. More specifically, we consider a partition of $\Omega = \cup_{i=1}^{m} \Omega_i$ into m disjoint subdomains (voxels) Ω_i. Then, from each CFI image we collect

$$\ell_i(u) = \int_{\Omega_i} u \cdot n \; d\Omega_i, \quad 1 \leq i \leq m, \tag{1.11}$$

where n is a unitary vector giving the direction of the ultrasound beam. From (1.11), it follows that the Riesz representers of the ℓ_i in V are simply

$$\omega_i = \chi_i n,$$

where χ_i denotes the characteristic function of the set Ω_i. Thus the measurement space is

$$W_m = W_m^{(CFI)} = \text{span}\left\{\omega_i\right\}_{i=1}^{m}.$$

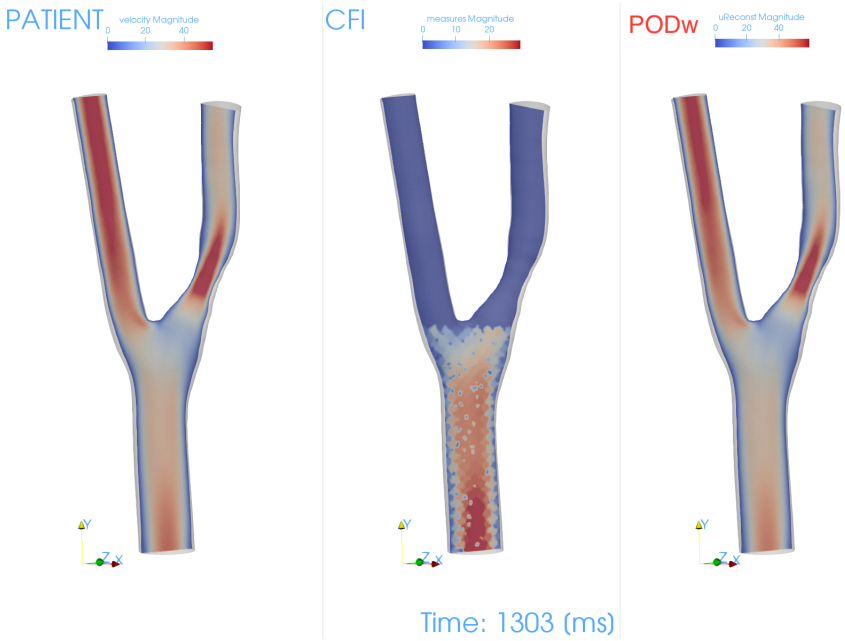


Figure 1.6: Result for the experiment commented in Section 1.2.3: velocity magnitude for the target solution (left), the CFI data (center), reconstruction (right)

Since the voxels Ω_i are disjoint from each other, the functions $\{\omega_i\}_{i=1}^m$ are orthogonal and therefore having a CFI image is equivalent to having

$$\omega = \mathbb{P}_{W_m} u = \sum_{i=1}^{m} \langle \omega_i, u \rangle \omega_i = \sum_{i=1}^{m} \ell_i(u) \omega_i. \tag{1.12}$$

The PBDW method consists in a regularised least square method that manage to perform a trade off between the model solutions and the data (both a priori and a posteriori knowledge are affected by some errors). In its first version (noiseless measurements) it can be cast as an optimisation problem:

$$u^* = \arg \inf_{u | \langle \omega_i, u \rangle = \ell_i, \ \forall i} \| \mathbb{P}_{V_n}^{\perp} u \|_{\mathcal{V}}^2, \tag{1.13}$$

where $\mathcal{V}$ is the Hilbert space the solution u belongs to.

The model used is the system of incompressible Navier-Stokes equations, in which the boundary conditions are imposed by using classical three elements Windkessel models. The details about the formulation are proposed in [15]. The solution set to be used is built by making the parameters vary in a physiological range. The solution set is partitioned based upon two parameters that can be easily assessed in the online phase (the cardiac rhythm and the normalised time in the cardiac cycle). The optimal partitioning of the solution set is a problem which is far from trivial. In the present work, a partitioning based on the empirical evaluation of the reconstruction performances on the database is proposed.

A realistic common carotid bifurcation geometry is considered. As it is shown in Figure

1.6 the data (center) are acquired only in the first portion of the carotid artery. The parametric Windkessel governing the outflow boundary conditions are used to mimic arterial blockage situations (that would correspond to an altered flow split with respect to the non-pathological configuration). The goal is to be able to predict these situations by using only the measurements in the first half of the geometry. The result shown in Figure 1.6 is a preliminary result in a semi-realistic setting. In particular, the reconstruction of the velocity field is quite accurate (errors are of the order of 1% in L^2 norm), which is promising in view of improving and deploying the proposed strategy in more realistic settings. More details about the results of the numerical experiments are proposed in [15].

1.3 Data assimilation in cardiac electrophysiology.

Cardiac toxicity is part of the Safety Pharmacology, and consists in studying the negative effects of drugs and chemical compounds on the cardiac function. The Comprehensive in vitro Proarrhythmia Assay (CiPA) is an initiative for a new paradigm in safety pharmacology to redefine the non-clinical evaluation of Torsade de Pointes (TdP) [CVJS16, MDS$^+$18, YKI$^+$18].

A way of measuring the impact of a molecule of a given ionic channel is to consider the Patch Clamp, that consists in putting micro-electrodes across the cellular membrane and measuring the activity of the single cells (The transmembrane potential recorded as a function of time is called Action Potential, AP). This method is precise but low-throughput and hence time consuming. The need to have a high-throughput screening of molecules candidate to become a drug motivated the development of protocols to exploit devices such as the Micro-Electrodes Array (MEA). In this device, a layer of cells (cardio-myocites) is put into a well, in close contact to electrodes, that record the electrical activity. The observation of the electro-graph (the signal is called Field Potential, FP in what follows) conveys an information on the overall electrical activity and how it is eventually altered by a certain molecule. In the following, two cases are presented, with realistic data sets: in the first one, a stochastic inverse problem is solved, in which the probability density distribution of the model parameters is estimated in such a way that the statistics of the model outputs match the statistics on the experimental measurements; in the second one, a classification problem on MEA signals is described.

1.3.1 Stochastic inverse problems on action potentials.

The variability observed in action potential (AP) measurements is the consequence of many different sources of randomness. In this contribution we focus on parameter randomness which, in the context of AP modeling, corresponds to the natural variability of the cardiomyocyte electrical properties such as its capacitance, ionic channel conductances and gate time constants. Due to the large number of free parameters in AP models, these parameters are in practice unidentifiable [DL04], *i.e.* different combinations of these parameters can lead to the same AP. Therefore, we choose to restrict our analysis to a subset of ionic channel maximal current densities which are referred to as conductances in the following.

AP measurements may result from heterogeneity within a population of cells (inter-subject variability) [SBOW$^+$14] or from dynamic variations within a single cell (intra-subject variability) [JCB$^+$15, PDB$^+$16].

Investigating the variability of AP models parameters has several motivations. For instance, it can be used to predict the response of cardiomyocytes to certain drugs [BBOVA$^+$13], or it can provide insight into cell modifications at the origin of common heart diseases such as atrial fibrillation [WHC$^+$04, SBOW$^+$14] or ventricular arrythmia [GBRQ14].

There are two main strategies to estimate the parameters variability given a set of AP measurements. First, one could fit the AP model to each measurement individually and therefore obtain a set of parameters from which useful statistics may be computed. The problem of fitting an individual AP has been addressed many times and using a large variety of methods [DL04, SVNL05, LFNR16]. However, the computational cost of such a strategy may be prohibitive, especially for large datasets.

The second strategy belongs to the so-called population of models approach. The experimental set is considered as a whole and the parameters statistics are estimated by solving a statistical inverse problem. Several techniques were developed to solve such problems [Kou09, GLV14] and their application to electrophysiology has recently generated much interest [MT11, BBOVA$^+$13, SBOW$^+$14, DCP$^+$16]. The method developed in this contribution (called Observable Moment Matching, OMM) aims at estimating the parameters (thought as random variables) PDF in such a way that the model outcome statistics match the observed ones. The method is explained in detail in Section 3.2 and detailed in [19]. The OMM method is applied to the estimation of the PDF of key conductances from AP measurements. Several experiments are proposed in [29]. We report, hereafter, the results on experimental signals collected from human atrial cells.

The data consist in a set of published AP biomarkers recordings that are readily available online[2]. The AP are recorded from human atrial cardiomyocytes coming from different subjects [SBOW$^+$14] . The data set is divided into two groups: one counting 254 Sinus Rythm (SR) patients and another one counting 215 chronic Atrial Fibrillation (AF) patients. Both groups exhibit a strong inter-subject variability in addition to the inter-group variability. The available biomarkers are: APD90, APD50, APD20, APA, RMP, $\mathrm{dV/dt_{max}}$, V20. The ionic model used is the Courtemanche-Ramirez model, in which 11 conductances were parametrised. All the details are available in [29].

The results of the estimation of the PDFs of the parameters are shown in Figure 1.7.

To each group is associated a most representative individual whose biomarkers values are the closest to the median ones of its group. The calibration step is very informative as it allows for a first comparison between the two groups, or more precisely between the two representatives of each group. The calibration leads to high differences for g_{K1} (+220%), g_{to} (-100%), g_{CaL} (-63%) and g_{Kur} (-60%) which are qualitatively similar to those reported in [SBOW$^+$14].

Beyond these inter-group variations captured in the calibration step, the inter-group variability is revealed by the study of the estimated PDFs (see Figure 1.7 B). The results highlight the distribution differences of g_{to} and g_{Kr} between the two groups: in the SR group, these two conductances feature a normal-like distribution whereas in the AF group those distributions are skewed and with a larger variance. The distribution of g_{Na} are similar between the two groups which suggests that it does not play an important role in the AF mechanisms. g_{K1} also features a much higher mean value and higher variance in the AF group. A posteriori distributions of the biomarkers of interest may be computed from the estimated PDF. When compared to the actual

[2]Data are available here *http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0105897*

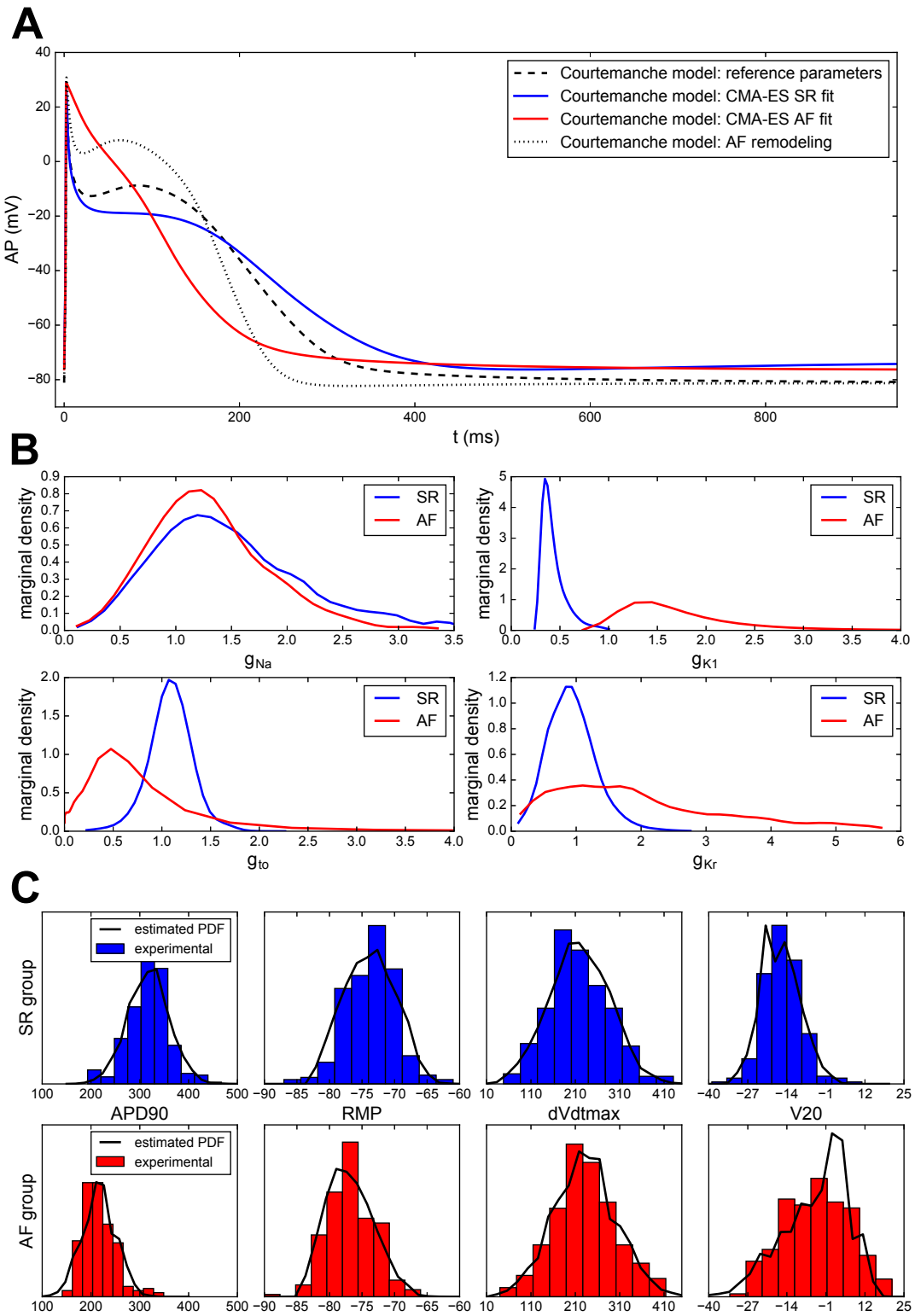


Figure 1.7: (A) CMA-ES parameter calibration of the Courtemanche model prior to the inverse procedure. APs obtained for the most representative samples of the SR (blue) and AF (red) groups, reference parameters (dashed) and after AF remodeling (dotted). (B) Courtemanche conductances estimated marginal densities for the SR (blue) and AF (red) groups. Conductances are normalized by the literature values. (C) Normalized histograms of the four experimental biomarkers of interest for both SR (blue) and AF (red) groups. The black solid lines correspond to the PDF of each biomarker estimated by OMM.

distributions (approximated by histograms of the experimental biomarkers), it shows that the OMM method succeeded in matching the variability in the measurements.

A point to be discussed is the use of biomarkers versus time traces. This is often imposed by the type of experimental data available. Ranges of biomarkers using standard protocols are easily accessed by experimentalists, and raw action potential data are not always available. It is therefore important to evaluate the use of both biomarker ranges and action potential traces. The set of available biomarkers is often dictated by experimental constraints. It is however possible, when there are many available biomarkers, to conduct a preliminary study to determine which biomarkers should be taken into account in order to recover certain parameters of interest. This issue is discussed in Section 3.3.

1.3.2 Classifying the electrical activity based on MEA signals.

In this Section we describe how to classify the action of 12 compounds on the ionic channel activity based on *in vitro* data derived from Micro Electrodes Array (MEA) recordings of spontaneous beating hiPSC-CMs (Pluricyte® Cardiomyocytes) cultured on 96 well MEA plates (8 electrodes per well, Axion Biosystems), as described in [28].

The contribution of this work is twofold:

1. The *in vitro* dataset is complemented by an *in silico* dataset, obtained by simulating the experimental device in a number of meaningful scenarios.

2. The classifier used to produce the result is optimised to deliver optimal classification performances.

In vitro data used for this part are FP traces recorded from a hiPSC-CM monolayer (Pluricyte® Cardiomyocytes, Ncardia) plated on a 96 well MEA plate (8 electrodes per well) Axion Biosystems [3].

The 12 CiPA compounds listed in Table 1.1 were tested on Pluricyte® Cardiomyocytes and FP traces were recorded before and 30 minutes post compound addition. MEA results of 5 compounds were used for the training and MEA results of 7 "blind" compounds for the validation.

Each compound was tested at 4 concentrations, 1 concentration per well and in 5 replicates ($n = 5$ per concentration). For this study only FP traces were recorded and used for the training and classification, no calcium transient measurements were performed. The final experimental sample size was 75 for the training set and 85 for the validation set (some wells were removed from the analysis due to quiescence or noisy signal observations).

To perform this application, several observations are crucial:

1. The number of available signals is limited; moreover, they do not cover the spectrum of potentially meaningful scenarios.

2. Several variability sources affect the observable: one is related to the different behaviour of the ionic channels; the other is related to the experimental conditions, such as fluctuations in the temperature, heterogeneity of the medium, variability in the physical parameters.

[3] Axion Biosystems device: Classic MEA 96 M768-KAP-96

Compound	IC50 (μM)			Concentration (μM)				Cmax (μM)	T/V	Label
	hERG	Cav1.2	Nav1.5	#1	#2	#3	#4			
Loratadine	6.1	11.4	28.9	0.001	0.003	0.0095	0.03	0.00046	V	K, Ca
Ibutilide	0.018	62.5	42.5	0.0001	0.001	0.01	0.1	0.1	T	K
Droperidol	0.06	7.6	22.7	0.03169	0.10014	0.31646	1.0	0.02	T	K
Mexiletine	62.2	125	38	0.1	1.0	10	100	2.5	T	Na, K
Dofetilide	0.03	26.7	162.1	0.0003	0.001	0.0032	0.01	0.002	V	K
Diltiazem	13.2	0.76	22.4	0.01	0.1	1.0	10	0.13	T	Ca
Chlorpromazine	1.5	3.4	3.0	0.0951	0.3004	0.9494	3	0.0345	V	K, Ca, Na
Clozapine	2.3	3.6	15.1	0.0951	0.3004	0.9494	3	0.07	T	K, Ca
Clarithromycine	32.9	>30	NA	0.1	1	10	100	1.2	V	K
Cisapride	0.02	11.8	337	0.0032	0.01	0.0316	0.1	0.0026	V	K
Bepridil	0.16	1.0	2.3	0.01	0.1	1	10	0.03	V	K, Ca, Na
Azimilide	<1	17.8	19	0.01	0.1	1	10	0.07	V	K, Na, Ca

Table 1.1: Section 1.3.2: experimental data. In the first column, the compounds name is written, followed by the properties, such as IC50 and the concentrations used. The column T/V indicates if the data associated to the compounds are used in the Training or in the Validation set, and the label denotes the ionic channel impacted by the molecule.

To account for these, the idea is to integrate, to the *in vitro* dataset, an *in silico* dataset, obtained by simulating the experimental setup. All the details are available in [28]. The model used is a 2D bidomain model, introduced in [RBZ$^+$17], discretised on a mesh accounting for the device geometry. To reproduce the heterogeneity of the cells distribution, a space stochastic process is used (constructing by making use of the Karhunen-Loeve decomposition). A set of scenarios is simulated by considering a distribution of:

- The ionic channel conductivity (for sodium, potassium and calcium).
- The source position (the place where the electric stimulation starts).
- The heterogeneity field.

The database of the observable is generated and added to the experimental set. In these set, however, the size of the discretised signals is large, and the number of the available samples is low. In view of performing classification tasks, this regime is particularly critic and unfavorable. A goal-oriented dimension reduction is performed to construct the classifier input in such a way that the classification score is maximised.

Several results are presented in [28]. Hereafter, the ternary classification of the molecules is proposed. A classical LDA classifier was used to assess whether a compound blocks the three main ionic channels. The result presented is synthetic, trying to provide an indication irrespective of the concentration of the compound. A finer analysis is proposed in [28], in which the classification is performed at each concentration. The classification results are presented in Figure 1.8, for the molecules in the validation set.

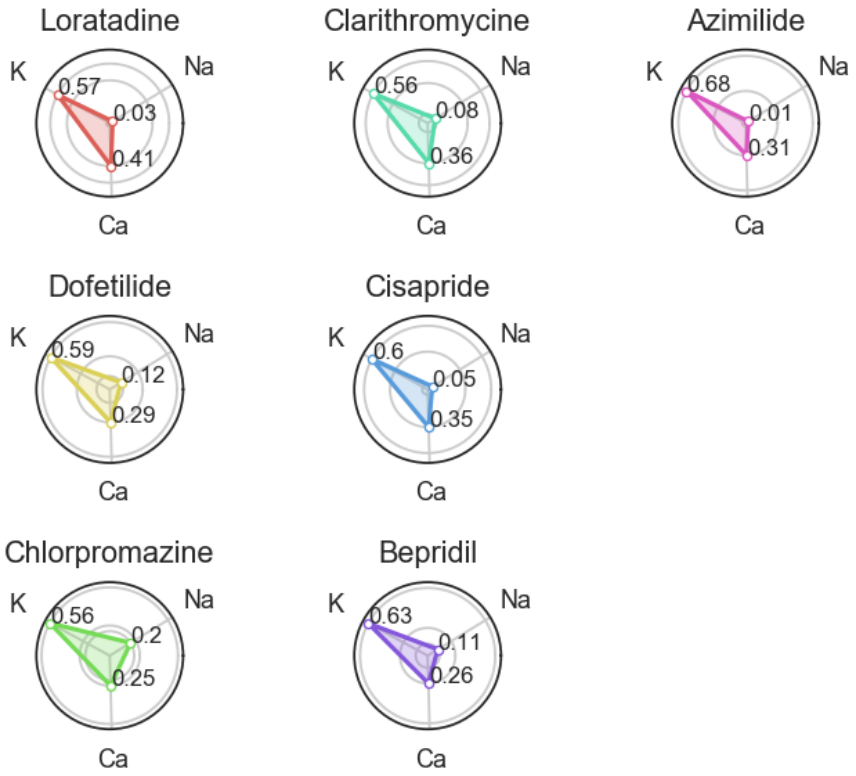


Figure 1.8: Experimental data ternary classification results, Section 1.3.2.

As expected, the probability returned by the classifier decreases when the IC50 value increases (for example the probability for Loratadine to be a calcium channel blocker is 0.41 with $IC50 = 11.4\mu M$ (see Table 1.1) and the probability for Dofetilide to be a calcium channel blocker is 0.29 with $IC50 = 26.7\mu M$ (see Table 1.1)).

Overall, these first results are encouraging, 11 molecules out of 12 were correctly classified. This work raises several methodological questions, which are partly addressed in Section 3.3.

1.4 Conclusions

Often, when dealing with complex systems, as the ones encountered in biomedical engineering, the ability to formulate models that accurately describe the system configurations, and simulate it, is not followed by the ability to actually use this quantitative description in a realistic scenario. Several methodological issues arise:

1. System identifiability is an open problem for PDEs. It boils down to a basic question: given the observables, can we infer in a reliable way the QoI? A possible way of defining a notion of practical identifiability and getting an assessment is proposed in Section 3.1.

2. Accounting for the inherent uncertainties is essential to seek robustness. Moreover, assessing the variability in a population is a key aspect in many realistic applications. How this can be solved in some applications is the object of the contributions detailed in the Chapter 3.

3. Dealing with models described by parametric PDEs (UQ, optimisation, forecast) entails a large computational cost, which turns out to be, in most of the cases, prohibitive or not compatible to the applications practical constraints. Some contributions to the approximation of high-dimensional problems solutions and Reduced-Order models is presented in Chapter 2.

Chapter 2

Reduced-order modeling and high-dimensional problems.

Context

In this chapter, some contributions to the study of Reduced-Order models and High-dimensional problems are presented. High-dimensional problems are ubiquitous in science and arise naturally when taking the uncertainty of phenomena into account (UQ), when dealing with mesoscopic descriptions of systems (Kinetic theory), or when considering optimal mass rearrangement (Optimal Transport). In applications, the solution of parametric systems of PDEs need to be performed in a time compatible with industrial needs. All these problems have a common inherent difficulty to be overcome when seeking a numerical approximation.

When considering generic functions belonging to certain functional spaces, say for instance $\Omega \subset \mathbb{R}^d$, $W^{k,p}(\Omega)$, the standard Sobolev spaces, the error when approximating a function goes roughly as $\propto n^{-\frac{k}{d}}$, where d is the space dimension, k is the regularity of the functions involved (and for most of the applications is $k = 1, 2$), n is the number of terms used for the approximation (a parameter that can be related to the rank) and d is the domain dimension. This simple relation illustrates well the phenomenon of the *curse of dimensionality*, and the scaling with the dimension is rather general and applies also to other spaces, such as the Besov and Hölder-Zygmund function spaces ([Pin12]). An important point is that, in general, we are not interested in approximating generic elements in the unit ball of the space. Instead, we would like to improve the approximation of functions belonging to a given subset of the unit ball of the space, solutions to a specific problem in a certain parameter range. This may result in breaking the isotropy of the target set to be approximated and in the potential emergence of sparsity in certain representations, depending on the nature of the problem at hand. The objective is precisely to investigate how to exploit this in order to construct parsimonious approximations.

2.1 Physical based ROMs for FSI

When dealing with the simulation of the cardio-vascular system, fluid-structure interaction plays a key role. Simulating large portions of the vascular network, or taking the interactions between

the vessels and the surrounding tissue into account can be computationally intensive. In this part, two contributions are presented: in the first one, a simplified fluid-structure interaction model is derived, based on physical arguments, to describe the mechanical behaviour of the arterioles; in the second one, a model reduction method is proposed to account for the interaction between a system under investigation and other (secondary) systems. These works were motivated by the modeling of the haemodynamics in the eye, in the context of the ITN Revammad (2013-2016).

2.1.1 A simplified fluid-structure interaction model

To model the cardiovascular system mechanical behaviour, complex nonlinear models that include large displacements and deformations have to be considered. This is, for example, the case for valve [LS13, vAdB04, vAv06, AGPT09] or aorta simulations [BCZH06, CRD$^+$11, MXA$^+$12]. It is well-known that these simulations are very demanding, in spite of the progress achieved in recent years ([FLV15, BČG$^+$14, BF14]).

In this contribution we consider those situations where the physical behaviour of the system make it possible to simplify the models and solve them efficiently. The idea is to replace the full fluid-structure problem by a fluid problem with non-standard boundary conditions at the fluid-structure interface, accounting for the solid mechanics. Various approaches have been recently proposed in this direction [FVCJ$^+$06, NV08, CDQ14, Pir14].

In [NV08], the authors started from a Koiter linear shell model and neglected the flexural terms. After discretization, the resulting fluid-structure equations are reduced to a fluid problem with Robin boundary conditions. In this approach, the fluid domain is moving and an ALE method is devised. In [Pir14], a further simplification is proposed: the fluid domain is fixed by introducing a zero order transpiration boundary condition; moreover the curvature of the artery is assumed to be constant. With these simplifications, the authors were able to perform a comprehensive mathematical analysis of the problem [RGMP14]. In [FVCJ$^+$06], A. Figueroa et al. also assumed that the computational domain was fixed and used a zero order transpiration boundary condition. The structural model was derived assuming homogeneity throughout the thickness. Compared to the two previous approaches, this one requires adding new degrees of freedom to the fluid problem. This drawback is, however, counterbalanced by the fact that the resulting model is more stable on real geometries featuring variations of curvature, according to [CDQ14] where an extensive comparison was proposed.

The main features of the simplified fluid-structure model proposed in this contribution are the following:

1. It mainly consists of a fluid problem on a fixed domain, with generalised Robin boundary conditions, which makes it insensitive to the added-mass effect.

2. The boundary is fixed and first order transpiration conditions are imposed on the velocity.

3. It takes into account in a simplified manner the presence of fibers in the solid; it is therefore less sensitive to strong variations or inaccuracies in the curvatures and, as a consequence, it remains robust in the presence of flat regions in the surface.

The equations for the mechanics of the wall were derived based on [Cia00]. The hypotheses are the following:

- The displacement of the structure is parallel to the normal field in the reference configuration.
- The bending terms are negligible.
- The material is linear, isotropic and homogeneous.

As a consequence, only the membrane part of the Koiter model is considered, the shell deformation is described by the change of metric tensor and the stress is linear in the deformation. Let $(\xi_\alpha, \xi_\beta) \in \omega \subset \mathbb{R}^2$: the reference configuration of the structure is obtained by a smooth mapping:

$$\boldsymbol{\phi} : \left\{ \begin{array}{ccc} \omega & \rightarrow & \mathbb{R}^3 \\ (\xi_\alpha, \xi_\beta) & \mapsto & (x_1, x_2, x_3) \end{array} \right.$$

The displacement of the structure is a field $\boldsymbol{\eta}(\xi_\alpha, \xi_\beta) \in \mathbb{R}^3$, the change of metric tensor (for the details we refer to [Cia00]) is denoted by g, the linear constitutive laws are expressed via the elastic tensor $\mathcal{E}$. We denote a the determinant of the first fundamental form and h_κ the thickness of the shell. The equilibrium configuration for the shell is the stationary point of the energy functional:

$$\psi^\kappa(\boldsymbol{\eta}) = \frac{1}{2} \int_\omega \mathcal{E}^{\alpha\beta\sigma\tau} g_{\sigma\tau}(\boldsymbol{\eta}) g_{\alpha\beta}(\boldsymbol{\eta}) \; h_\kappa \sqrt{a} \, d\boldsymbol{\xi} - \int_\omega \mathbf{f} \cdot \boldsymbol{\eta} \; h_\kappa \sqrt{a} \, d\boldsymbol{\xi}, \tag{2.1}$$

where $\mathbf{f}(\xi_\alpha, \xi_\beta) \in \mathbb{R}^3$ are the external forces. The equations for a generic fiber layer are detailed. The main hypotheses are the following:

- The elastic energies of the shell and the fiber layer sum up.
- From a kinematical point of view, the fibers are perfectly attached to the shell.
- The fibers are characterized by an affine stress-strain constitutive law.

The second hypothesis implies that the deformation of the fibers equals the deformation of the underlying shell structure in the direction of the fibers. Let $\mathbf{w} \in \mathbb{T}(\boldsymbol{\Phi})$ be a vector of the tangent bundle to the surface (the shell reference configuration), $\varrho_{\mathbf{w}}$ be the fraction of the total number of fibers aligned with the direction $\mathbf{w}$ and h_f the thickness of the fibers layer. The elastic energy of the fibers aligned in the direction $\mathbf{w}$ is expressed in the form:

$$\psi^{\mathbf{w}}(\boldsymbol{\eta}) = \frac{1}{2} \int_\omega \varrho_{\mathbf{w}} \left[k_0 + k_1 \varepsilon_{1D}(\boldsymbol{\eta}) \right] \varepsilon_{1D}(\boldsymbol{\eta}) \; h_f \sqrt{a} \, d\boldsymbol{\xi} + \int_\omega r_{\mathbf{w}} \; h_f \sqrt{a} \, d\boldsymbol{\xi}, \tag{2.2}$$

where $r_{\mathbf{w}}$ represents the potential energy of a force acting on the fibers aligned with the direction $\mathbf{w}$ of the tangent space. The energy of the structure is simply obtained by: $\Psi^s = \Psi^\kappa + \Psi^{\mathbf{w}} + \Psi^{\mathbf{w}^\perp}$, by virtue of the first hypothesis.

The fluid is described by the incompressible Navier-Stokes equations and the coupling conditions are the kinematic and the dynamic continuities of the velocity and the stress respectively.

The details of the derivation are reported in [2]. When using the hypothesis of normal displacement, it holds: $\boldsymbol{\eta} = \eta \mathbf{n}$. The system reads:

$$\begin{cases} \langle \partial_t \mathbf{u}, \mathbf{v} \rangle + c(\mathbf{u}; \mathbf{u}, \mathbf{v}) + a(\mathbf{u}, \mathbf{v}) + b(p, \mathbf{v}) = 0 & \text{in } \Omega, t > 0 \\ \langle \nabla \cdot \mathbf{u}, q \rangle = 0 & \text{in } \Omega, t > 0 \\ \rho_s h_s \langle \partial_{tt}^2 \eta, \chi \rangle_\omega + \Psi^s(\eta, \chi) + \langle p^{ref}, \chi \rangle_\omega = \langle p + \eta \nabla p \cdot \mathbf{n}, \chi \rangle_\omega & \text{on } \Gamma \\ \langle \partial_t \eta, \chi \rangle_\omega = \langle \mathbf{u} \cdot \mathbf{n} + \eta \nabla \mathbf{u} \mathbf{n} \cdot \mathbf{n}, \chi \rangle_\omega & \text{on } \Gamma \\ \langle (I - \mathbf{n} \otimes \mathbf{n})(\mathbf{u} + \eta \nabla \mathbf{u} \mathbf{n}), \mathbf{w} \rangle_\omega = 0 & \text{on } \Gamma. \end{cases} \tag{2.3}$$

The forms a, b, c read:

$$\begin{aligned} &a : V \times V \to \mathbb{R}, \quad a(\mathbf{u}, \mathbf{v}) = \nu^f (\nabla \mathbf{u} + \nabla \mathbf{u}^T, \nabla \mathbf{v})_\Omega && \forall (\mathbf{u}, \mathbf{v}) \in V \times V \\ &b : M \times V \to \mathbb{R}, \quad b(p, \mathbf{v}) = -(p, \nabla \cdot \mathbf{v})_\Omega && \forall (p, \mathbf{v}) \in M \times V \\ &c(\mathbf{w}) : V \times V \to \mathbb{R}, \quad c(\mathbf{w}; \mathbf{u}, \mathbf{v}) = (\mathbf{w} \cdot \nabla \mathbf{u}, \mathbf{v})_\Omega && \forall (\mathbf{u}, \mathbf{v}) \in V \times V \end{aligned} \tag{2.4}$$

The system of equations is discretised by means of P1-P1 SUPG (Streamline Upwind Petrov-Galerkin) stabilized finite elements. In order to impose the boundary conditions, a P1 reinterpolation of the normal field $\mathbf{n}$ on the boundary of the fluid domain is needed, to avoid numerical issues, as reported in [CDQ14]. A vector generalised Robin boundary condition is derived, that improved the numerical stability. All the details about the method can be found in [2]. This contribution is then complemented by a model of smooth muscle cells and of the feedback characterising the autoregulation phenomenon. These were detailed in [1]. The proposed method made it possible to perform a simulation on a realistic 3D retinal vascular network geometry composed by 24 vessels.

2.1.2 Reduced-order Steklov operator for coupled problems

This work deals with the study of a Reduced-Order Method to approximate the solution of coupled multi-physics systems. In particular, we investigate the case in which one system of interest, described by a possibly non-linear Partial Differential Equation (PDE) interacts with one (or more) other systems through its boundaries. Numerous applications in science and engineering are characterized by different compartments in interaction, think for instance to thermal-fluid-structural or electro-mechanical-fluid couplings. In several cases, one is not interested in approximating the solutions of all the systems, but only to the solution of a main system, denoted by $\mathcal{P}_1$. The objective is to be able to compute precisely the solution of this system (by using a full-order classical method) but to reduce the computational costs associated to the solution of the systems in interaction with it (denoted as $\mathcal{P}_2$). This results in a significant speed up of the problem simulation, in the case in which the size of $\mathcal{P}_2$ is larger than that of $\mathcal{P}_1$. Indeed, classically, there are two ways to deal with coupled systems: a monolithic approach in which all the systems are simultaneously solved, or a Domain Decomposition method (see [QV99] for a complete review of the method). In the latter a fixed point iteration is adopted, in which all the systems are separately solved and share the boundary data. For sake of simplicity, we made the assumption that there is only one system to be reduced and that it is described by linear PDEs. In this case the interaction of the linear system with $\mathcal{P}_1$ may be described by the

Poincaré-Steklov operator. With a slight abuse of notation we call Poincaré-Steklov operator the one associated to a generic linear PDE, even if historically this name refers to the case in which the secondary system is described by a Laplace equation (in [AL85] a first analysis of the problem is presented).

The need to set up efficient solvers and to decouple the solution of the problems in interaction is related to the ability to solve the problem at the interface. The use of the Poincaré-Steklov operator as a preconditioner in fluid-structure interaction iterations was investigated in [DDFQ06]. An efficient non-linear coupling strategy was devised in [CPW14] to set up an uncertainty quantification method applied to networks of coupled systems. Unfortunately, the problem at the interface is in general not sparse and ill-conditioned (as detailed in [QV99]). To tackle this issue several strategies were proposed in the literature. They can be broadly divided into two classes: local and spectral approximations. A local approximation of the Poincaré-Steklov operator consists in solving one or more external problems ($\mathcal{P}_2$) in a strip localised around the interface. Such a method was proposed for example in [PS05], for applications in hydrology. A similar procedure, based on a two-scale method, was presented in [GAG11]. A different strategy consists of approximating the leading part of the action of the Poincaré-Steklov operator through a spectral decomposition. Such an approach was proposed in [Nat95, Nat97] in the case of elliptic problems and a multiscale version was proposed in [CZAL13] for applications in heterogeneous media. An approximation of the Poincaré-Steklov operator via a Padé expansion was detailed in [LP10] for the study of the vibrations in fluid-structure couplings. In [FN14] the Poincaré-Steklov operator is computed in the context of the wave propagation in elastodynamics by considering a family of smooth functions at the interface and by solving the problem $\mathcal{P}_2$ by taking these functions as inputs. In the recent work [BRD15] a compressed sensing approach is proposed to retrieve the discretised Poincaré-Steklov operator for coupled Helmholtz problems. The method consists of probing, randomly, the matrix associated to the Poincaré-Steklov operator, by selecting inputs from a kernel space and performing a direct full-order simulation for a small number of them.

In the present contribution, a low rank decomposition of the Poincaré-Steklov operator is computed by a Reduced-Order Modeling method. In the literature, similar works were recently proposed, based on a Reduced-Basis framework [HKP13, EP14, MRH15, IQR16]. In these, multi-domain systems are considered for coupled linear steady problems. The framework proposed in these works deals with parametrized systems whereas in the present work we focus on the acceleration of single-scenario simulations. In the present case, a parametrization can be considered for the problem $\mathcal{P}_1$.

The proposed strategy can be divided into two phases: an *offline* phase and an *online* one. In the offline phase, a deterministic sampling of the functional space of the input for the $\mathcal{P}_2$ at the interface is considered and the output is saved, at the interface. This is similar to what was proposed in [HKP13], in which harmonic functions at the interface are used and the offline phase is somewhat independent from the coupling. The output of $\mathcal{P}_2$ at the interface is used to get a Low Rank Decomposition of the Poincaré-Steklov operator. In order to make the method more robust in cases in which the inputs coming from $\mathcal{P}_1$ are outside the space spanned by the sampled functions used to construct the database, an *online* update of the reduced Steklov representation is performed, in the spirit of the methods proposed in [PW15, AZW15].

The advantages of the proposed framework are the following:

1. It is a straightforward method, allowing to speed up coupled multi-physics time dependent systems in a domain decomposition approach.

2. The offline phase is completely independent of the nature of $\mathcal{P}_1$ and of the coupling.

Its main limitations concern the assumptions made on $\mathcal{P}_2$: in the present work, $\mathcal{P}_2$, once discretized, has to be autonomous and linear. Although these assumptions are quite restrictive, they are fulfilled by a wide range of applications, for which a system of interest interacts with surrounding media or compartments whose dynamics is linear.

In the offline phase, a number of simulations is performed to construct a database of meaningful solutions, to be exploited in the online phase. Contrary to most of the classical methods of model reduction, we made the choice of simulating only $\mathcal{P}_2$ in the offline phase, and not the whole system. From the point of view of the memory usage, there is no need to save the whole solution of $\mathcal{P}_2$, but only a restriction of a linear operator applied to its solution on the interface Γ. We need to choose a set of basis functions to represent the input datum on $\mathcal{P}_2$. Such basis has to be defined on a generic Riemannian manifold, to be orthonormal and complete. In view of these desired properties, a reasonable choice is to a priori take the first N_ℓ eigenfunctions of the Laplace-Beltrami operator defined on the surface, with $N_\ell \ll N_\Gamma$, the number of degrees of freedom at the interface. The advantages of choosing the eigenfunctions of the Laplace-Beltrami operator are the following:

- The basis is a complete basis of $V := L^2(\Gamma)$.
- It is hierarchical.
- It automatically accounts for symmetries in the geometry.
- The extraction of the basis amounts to solve a sparse eigenvalue problem defined on Γ.
- On particular (but meaningful) geometrical settings, the basis coincides with the eigenfunctions of the Poicaré-Steklov operator (this was proved in the Appendix of [3]).

Let the eigenfunctions of the Laplace-Beltrami operator be denoted as follows: $v_i \in V$, $i = 1, \dots, N_\ell$ is such that $-\Delta_\Gamma v_i = \mu_i v_i$.

The problem in weak form reads:

$$\langle \nabla_\Gamma v_i, \nabla_\Gamma \omega \rangle_\Gamma = \mu_i \langle v_i, \omega \rangle_\Gamma, \ \forall\ \omega \in H^1(\Gamma), \tag{2.5}$$

where ∇_Γ denotes the surface gradient and $\langle u, v \rangle_\Gamma = \int_\Gamma uv \ d\Gamma$ is the inner product on the interface Γ.

Once the basis has been extracted by solving a sparse eigenvalue problem, the problem $\mathcal{P}_2$ is solved, for every input function v_i. Only the image of a linear operator applied to the solution is stored on the boundary, *i.e.*: $y_i = \mathcal{S} v_i = \mathcal{T}\ell(u_2^{(i)})$, where $u_2^{(i)} = \mathcal{L}^{-1}(v_i)$.

In the online phase, the action of the Poincaré-Steklov operator on a generic input datum d can be approximated as follows:

$$\mathcal{S} d \approx \mathcal{S}_0 + \sum_j^{N_\ell} \langle d, v_j \rangle_\Gamma \ y_j + \sum_k^{N_o^{(n)}} \langle d, w_k \rangle_\Gamma \ z_k, \tag{2.6}$$

where $\mathcal{S}_0$ accounts for eventual non-homogeneous boundary conditions for $\mathcal{P}_2$ on $\partial\Omega_2/\Gamma$, the second term corresponds to the contribution of the projection of the datum d in the space spanned by the Laplace-Beltrami eigenfunctions and the last term is an *online update* of the basis, such that $\langle w_k, y_j\rangle_\Gamma = 0$. The online update of the basis is similar, in the spirit, to what is proposed in [PW15, AZW15]. In particular, when an error criterion is not fulfilled, the basis is updated by adding elements to it, coming, in the present approach, from a full-order problem simulation. The main difference with respect to the cited works concerns the way the update is performed. In the present approach we decided not to use a thin SVD update, but simply to increase the basis size and perform an orthogonalisation through a Modified Gram Schmidt (MGS). The details about the basis enrichment are reported in [3]. An analysis of the method in the case of Laplace equations leads to:

1. *Proposition.* *Let $\partial\Omega_2$ be C^2 and let $S : H^{-1/2}(\Gamma) \to H^{1/2}(\Gamma)$ be the Neumann-to-Dirichlet (N2D) map, S^{-1} the Dirichlet-to-Neumann (D2N) map; let w be the datum and $\hat{w}$ its projection onto the subset of the first N_ℓ Laplace-Beltrami eigenfunctions. Then:*

1. *if $w \in L_2(\Gamma)$, then $\lim_{N_\ell\to\infty} \|S(w-\hat{w})\|_{L^2(\Gamma)} = 0$.*
2. *if $w \in H^2(\Gamma)$, then: $\|S(w-\hat{w})\|_{L^2(\Gamma)} \leq C_1 N_\ell^{-2/d_\Gamma} |w|_{H^2(\Gamma)}$.*
3. *if $w \in H^2(\Gamma)$, then: $\|S^{-1}(w-\hat{w})\|_{L^2(\Gamma)} \leq C_2 N_\ell^{-1/2d_\Gamma} |w|_{H^2(\Gamma)}$.*

The method was tested on non-trivial fluid-structure interaction problems in 3D settings.

2.2 Optimal Transport

Optimal transport is an example of a high-dimensional problem that finds applications in numerous and very diverse fields. It was first defined as a mathematical problem in [Mon81]. The relaxation of the Monge problem was proposed by Kantorovich in [Kan42]. The recent re-discovery of Optimal Transport is related to the work presented in [Bre87, Bre91], in which a generalisation of the Helmholtz and De Rham decompositions of vector fields is proposed. This entailed the understanding of the numerous links between the Monge-Kantorovich problem and other problems in mathematical physics, science and engineering. In [BB00], the Benamou-Brenier formulation is presented, that provides a fluid-mechanics interpretation of the transport. This naturally defines the Eulerian equivalent to the McCann interpolation of densities (the geometry of optimal transportation was investigated in [GM96]). The key mathematical properties of optimal transport solutions can be found in [Caf00, Fig07]. A comprehensive overview on the topic is available in [Vil03, Vil08, AG13, San15].

The notation is introduced. For sake of simplicity, the classical formulation is proposed, first, followed by the Benamou-Brenier formulation and the Kantorovich relaxation. Let $d \in \mathbb{N}^*$ and $(\Omega_0, \Omega_1) \subseteq \mathbb{R}^d$ be a countable intersection of open sets. Let $\xi \in \Omega_0, x \in \Omega_1$ and $\varrho_0(\xi), \varrho_1(x) > 0$ be two densities such as:

$$\int_{\Omega_0} \varrho_0 \, d\xi = \int_{\Omega_1} \varrho_1 \, dx = 1. \tag{2.7}$$

The goal is to find a mapping $X(\xi)$:

$$X : \left\{ \begin{array}{ccc} \Omega_0 & \rightarrow & \Omega_1 \\ \xi & \mapsto & x \end{array} \right.$$

The mass balance equation expressed in Eq.(2.7), in the case in which X is a one-to-one mapping leads to the Jacobi equation:

$$\varrho_0(\xi) = \varrho_1(X(\xi)) \det(\nabla_\xi X). \tag{2.8}$$

In 1D optimal transport, by imposing boundary conditions, this equation is per se sufficient to solve for the mapping. For $d > 1$, it is not, so that another condition on the mapping needs to be introduced. Among the mappings satisfying the Jacobi equation, we choose the one which is minimising a cost, that, in the case of the L^2 Monge-Kantorovich problem reads:

$$\mathcal{W}_2^2 = \int_{\Omega_0} \varrho_0 |X(\xi) - \xi|^2 \, d\xi, \tag{2.9}$$

and the cost is the 2–Wasserstein distance between ϱ_0 and ϱ_1. In the Benamou-Brenier formulation, the Eulerian version of the optimal transport is defined. In particular, a flow is sought, solution of the following:

$$[\varrho(x,t)^*, \mathbf{m}(x,t)^*] = \arg \inf_{\varrho, \mathbf{m}} \sup_{\phi} \int_0^1 \int_\Omega \frac{|\mathbf{m}|^2}{2\varrho} + \phi \left(\partial_t \varrho + \nabla \cdot \mathbf{m} \right) \, dxdt, \tag{2.10}$$

$$\varrho(x,0) = \varrho_0, \tag{2.11}$$

$$\varrho(x,1) = \varrho_1. \tag{2.12}$$

In the Kantorovich formulation, we consider the product space $\Omega_0 \times \Omega_1$ and look for a transport plan $\Pi(\xi, x) \in \mathcal{P}(\Omega_0 \times \Omega_1)$ such that:

$$\Pi(\xi, x) \in \mathcal{K}, \tag{2.13}$$

$$\mathcal{K} = \left\{ \Pi(\xi,x) | \int_{\Omega_1} d\Pi(\xi,x) = \varrho_0(\xi), \int_{\Omega_0} d\Pi(\xi,x) = \varrho_1(x) \right\} \tag{2.14}$$

and the optimal transport plan satisfies:

$$\Pi^* = \arg \inf_{\Pi \in \mathcal{K}} \int_{\Omega_0 \times \Omega_1} c(\xi, x) d\Pi(\xi, x), \tag{2.15}$$

where c is the cost, which is a l.s.c. function, and, that, for most of the problems of interest is a c-convex function. For the L^2 Monge-Kantorovich problem $c(\xi, x) = |x - \xi|^2$.

2.2.1 Numerical methods for optimal transport

Several difficulties must be overcome when solving an optimal transport problem:

1. In the original formulation a static mapping is sought, solution of a (highly) non-linear system of equations in $\Omega_0 \subseteq \mathbb{R}^d$. The advantage is that a function in a low dimensional space have to be approximated, all the difficulty consisting in taking care in an effective way of the non-linearity.

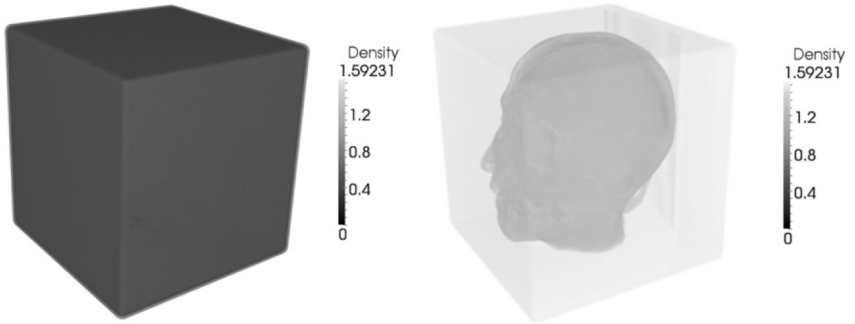


Figure 2.1: Example of 3D two marginals optimal transport, Section 2.2.1: at the left ϱ_0, at the right ϱ_1.

2. In the Benamou-Brenier scheme, the numerical approximation is carried out in two steps: the solution of a $d+1$ Poisson problem and a point-wise projection on a paraboloid, for which an analytic expression can be used. The disadvantage is in the increase in the dimension, the advantage is that the problem is decoupled into a costly linear problem and an inexpensive non-linear update step.

3. In the Kantorovich formulation the problem is a linear optimisation, and, once discretised, it is equivalent to a linear programming problem.

In [20] a first contribution is proposed to deal with the numerical approximation of optimal transport in its fluid-mechanical formulation. In particular, a Lagrangian discretisation is proposed, that exploits the knowledge of the trajectories geometry in the Benamou-Brenier formulation. Indeed, the trajectories are straight lines, as it can be deduced by the Hamilton-Jacobi-Bellman equation as well as by the McCann interpolation formula. A set of Lagrangian particles is considered, that approximates well the density ϱ_0. The initial velocity of these particles is estimated via an optimisation problem, such that the final density ϱ_1 is approximated at best and the kinetic energy of the transport is minimised.

$$\mathbf{v}_* = \arg\inf_{\mathbf{v}} \int_\Omega \frac{1}{2}\varrho_0(x)|\mathbf{v}|^2 \, dx + \int_\Omega \lambda \left(\varrho_0(x) - \varrho_1(x+\mathbf{v}) \det(I + \nabla \mathbf{v})\right) \, dx. \tag{2.16}$$

Several 2D and 3D test cases are proposed in [20], by using an augmented Lagrangian formulation, solved by means of a Uzawa-2 method. The solution of the problem reduces hence to a non-linear optimisation. The disadvantage is that it is non-convex, due to the formulation in $\mathbf{v}$ instead of $\mathbf{m}$ and to the discretisation choice; the main advantage consists in its simplicity (it exploits the knowledge of the Lagrangian trajectories geometry) that, coupled to a multi-level approach, made it possible to perform large 3D tests (As, for example, the one proposed in Figure 2.2.1).

In the more recent [11], a possible discretisation of the multi-marginal Kantorovich formulation is considered. The starting point is an approximated formulation of the Kantorovich

problem, that we called Moment Constrained Optimal Transport (MCOT). It reads:

$$\mathcal{C}_k^{(\xi)} = \mu_k - \int_{\Omega_0} v_k(\xi)\mu(\xi)\ d\xi,\ \text{for } k = 1, \dots, K_\xi,\ v_k \in \mathcal{V}_\xi,$$
$$\mathcal{C}_k^{(x)} = \nu_k - \int_{\Omega_1} w_k(x)\nu(x)\ dx, \text{for } k = 1, \dots, K_x,\ v_k \in \mathcal{V}_x,$$
$$\mathcal{K} = \left\{\Pi \in \mathcal{P}(\Omega_0 \times \Omega_1) \mid \mathcal{C}_k^{(\xi)} = 0, \mathcal{C}_k^{(x)} = 0\right\},$$
$$\Pi_* = \arg\inf_{\Pi\in\mathcal{K}} \int_{\Omega_0}\int_{\Omega_1} c(\xi, x)\Pi\ d\xi dx. \tag{2.17}$$

A basis of the marginal spaces is chosen (the marginal spaces are normally of dimension $d \leq 3$ and hence a classical discretisation can be adopted). The transport plan is discretised by a sum of Dirac deltas (we consider an approximation of the plan Π by the empirical distribution). The Tchakalov theorem, originally introduced to deal with quadrature of polynomials, makes it possible to establish a relationship between the number of basis functions in the marginal spaces and the number of Dirac deltas used to approximate the plan. The advantage of the proposed formulation is that it is suitable for large dimensional problems (The number of degrees of freedom is linear in the dimension); the disadvantage is that the different discretisations make the resulting optimisation problem non-linear: a robust and efficient numerical method to deal with it is still under investigation.

2.2.2 Unbalanced optimal transport

The optimal transport problem is defined for balanced densities, *i.e.* densities that have the same mass. When dealing with realistic data, this is rarely the case, due to noise, or to the fact that the observables are not balanced densities. This motivates the study of extensions of optimal transport allowing for mass sources and sinks. Several works are proposed in the literature to define such an extension. Among them, we cite [Ben03] in which the L^2 distance is used in order to deal with the change of mass (particularly effective when the difference in mass is due to noise); in [PR14] a first generalisation of the Wasserstein distance is proposed, in the same spirit of what has been done more recently in [LMS18, CPSV18].

In the work [24], an extension of the Benamou-Brenier formulation is proposed to deal with unbalanced densities. Several properties are required:

1. In the limit case of balanced densities, the Benamou-Brenier formulation is retrieved.
2. The McCann interpolation obtained preserves the time reversal symmetry as in the original optimal transport problem.
3. The McCann interpolation between the densities has finite speed.

The considered modification of the Benamou-Brenier formulation has the following form:

$$[\varrho(x,t)^*, \mathbf{m}(x,t)^*] = \arg\inf_{\varrho,\mathbf{m}} \sup_{\phi} \int_0^1 \int_\Omega \frac{|\mathbf{m}|^2}{2\varrho} + \phi\left(\partial_t\varrho + \nabla\cdot\mathbf{m} + S\right)\ dxdt, \tag{2.18}$$
$$\varrho(x,0) = \varrho_0, \tag{2.19}$$
$$\varrho(x,1) = \varrho_1. \tag{2.20}$$

Where S is a source term modifying the mass continuity constraint. To respect the symmetry constraint of the McCann interpolation, the source needs to satisfy

$$S(x,t,\varrho,\mathbf{m},\varrho_0,\varrho_1) = -S(x,1-t,\varrho,-\mathbf{m},\varrho_0,\varrho_1).$$

Moreover, the source vanishes when the initial and final densities have the same mass. A generic model of source is introduced of the form $S = \varrho D_t \Gamma$, where Γ is a scalar field and D_t denotes the total derivative. A proposition is proved in [24] to show that such a model of source, after a change of variable, can be interpreted as a balanced optimal transport with isotropic non-uniform metric. Moreover, several possible cases were analysed, namely, a pure exponential growth $S = c\varrho$ or a more complex source $S = \mathbf{m} \cdot \nabla\Gamma$, where Γ is the solution of a Poisson problem depending on the initial and final densities. For these models, existence and uniqueness of the solution are proved:

2. *Proposition.* *Let the unbalanced transport problem be defined, in Benamou-Brenier formulation as:*

$$[\varrho_*, \phi_*] = \arg\inf_{\varrho,\mathbf{m}} \sup_{\phi} \int_0^1 \int_\Omega \frac{|\mathbf{m}|^2}{2\varrho} + \phi \left(\partial_t \varrho + \nabla \cdot \mathbf{m} + \log\left(\frac{\int_\Omega \varrho_0 \ dx}{\int_\Omega \varrho_1 \ dx} \right) \varrho \right) \ dxdt.$$

The solution exists and it is unique.

3. *Proposition.* *Let $\delta = \varrho_1 - \varrho_0$ and the unbalanced transport problem be defined, in Benamou-Brenier formulation as:*

$$[\varrho_*, \phi_*] = \arg\inf_{\varrho,\mathbf{m}} \sup_{\phi} \int_0^1 \int_\Omega \frac{|\mathbf{m}|^2}{2\varrho} + \phi \left(\partial_t \varrho + \nabla \cdot \mathbf{m} + \mathbf{m} \cdot \nabla\Gamma_* \right) \ dxdt.$$

$$[\Gamma_*, \mu_*] = \arg\inf_{\Gamma} \sup_{\mu} \int_\Omega \frac{|\nabla\Gamma|^2}{2} - \frac{|\int_\Omega \delta \ dx|}{\int_\Omega \delta \ dx} \delta\Gamma \ dx - \mu \int_\Omega \delta e^\Gamma \ dx.$$

The solution exists and it is unique.

The proof of these results follows from the equivalence of these problems to balanced transport in a Riemmanian manifold, and to the fact that the geodesics have finite speed.

A nonlinear model is also introduced, to deal with applications such as morphogenesis. The algorithm proposed in [24] is a modification of the classical Benamou-Brenier algorithm, and enjoys the same properties. Several numerical test-cases were proposed to assess the performances of the method.

In the work [24] the focus was on the density interpolation. The source models proposed respect all the characteristics of the classical optimal transport interpolation and simply reduce to it when the densities are balanced. The optimisation problems defined, however, do not make it possible to always define a distance between the densities.

2.3 Advection dominated problems

An extensive literature on model reduction is available. We refer to [BHL93, QR14, QMN15, BMS05, BCOW17] for an overview of the classical model reduction methods.

Most of them are commonly based on the construction of a (possibly low-dimensional) linear subspace, which is exploited to build a parsimonious discretisation, often by projection, of the system of equations. The properties of such an approach are well described, from an approximation theory point of view, by the notion of Kolmogorov widths (the reader can refer to [Pin12] for a general introduction and to [CD15] for an investigation of the behaviour of the widths under holomorphic mappings). When dealing with sets of parametrised solutions featured by advection, transport, progressive waves, the standard methods often have insufficient performances.

Several works try to circumvent the poor representation given by a linear space by building local spaces. In [Car15] a strategy inspired by the mesh $h-$refinement is proposed. Each mode (obtained by Proper Orthogonal Decomposition, POD in the following) can be split (along the coordinate of the dofs) into a set of disjoint support modes. The so obtained space is more suitable to represent local features in the solution. In [AZF12] a construction of linear subspaces is presented. During the offline stage a k-means clustering is performed to group together similar snapshots of the solution (where the similarity is defined by a distance which is pertinent to the problem). The subspace is then chosen online in an efficient way, that ensures that the overhead of this operation does not jeopardize the gain provided by the use of a lower dimensional subspace. In [AH16] the authors investigate the way of building the subspaces to be used for the approximations. In particular, they show that by using the true projection error in order to cluster the snapshots of the system solution, the local bases can be built in a more efficient way. A similar approach is presented in [PBWB14]. In this work the authors present a method based on machine learning (k-means and classifiers to be used online) to build local subspaces and overcome some drawbacks of the Discrete Empirical Interpolation Method (DEIM) approach. In the two recent preprints [LC18, GB18], the subspaces to be used for model reduction are identified by using autoencoders and neural networks.

Other works focused on the stabilisation of the solution of the Reduced-order model, which can be particularly challenging when advection is dominating. In [MMQ16] an online stabilization is proposed. A more recent work [TBR18] applied a similar stabilization (weighted RB) to advection dominated problems with random inputs. A different form of stabilization is proposed in [AAC16], in which a dictionary of snapshots is introduced and a L^1 minimisation problem is used in order to identify the solution, leading to better (more stable) performances than the classical L^2 one. This is an example of the use of a distance different from the L^2 one to improve the performances of model reduction.

Several classes of methods were investigated that exploit the geometrical nature of the system to build non-linear approximations. Several contributions in the literature were proposed in which the snapshots are transformed by making a Lie group acting on them. In [MS18] the authors introduced a method to deal with parametric dynamical systems whose dynamics is invariant under continuous group action. In particular, they defined the method of slices to get rid of the continuous symmetry. Once the snapshots are all in the template configuration a classical reduction is performed[1]. In [OR13] the method of freezing is introduced. This is based on making a Lie group acting on the snapshots so that a reduced basis approach can be applied in this frozen-in-time configuration, resulting in a better decay of the approximation

[1]On the KdV equations these symmetries were studied analytically in [GW92]

widths. Works that try to exploit the geometry and to better adapt to the time dependence of the solution are the ones proposing a dynamical basis discretisation. We cite in particular [FL18] where the Dynamically Orthogonal (DO) approximation is explored. It is the canonical reduced order model for which the corresponding vector field is the orthogonal projection of the original system dynamics onto the tangent spaces of the solution manifold. A mathematical analysis of the approach is done in [MNZ15] and it uses ideas introduced in [KL10]. In the context of Hamiltonian systems, [AH17, HP18] introduce a reduced-order framework that preserves the geometrical nature of the dynamical system. The authors propose a modification of the model reduction frameworks that are commonly used in order to respect the symplectic geometrical structure of the phase space, leading to more precise spaces. In [Wel15, Wel17], a transformed snapshot interpolation method is introduced. The aim is to find a set of non-linear transformations such that the approximation of the transformed snapshots by a linear combination of modes is efficient. The main object of the paper is to investigate how these non-linear transformations can be identified by using a database of solutions. The transformation can be computed by solving an optimisation problem. In order to make the problem better conditioned and feasible, transitivity of the subsequent application of such transformations is used. The method is tested in several 1D and 2D configurations arising in compressible gas-dynamics. In [NB18] the set of the snapshots is transported into a reference configuration by a displacement field which is identified through a polynomial expansion. The targeted application is compressible fluid-dynamics, for which a change of the parameters can lead to a significant change in the position of shocks and rarefactions. The rationale behind the method is that, despite the solution is far from being regular, the displacement field transporting the snapshots into a reference configuration is, and a polynomial expansion can accurately identify it. A traditional reduction paradigm is applied to the transported snapshots. In [CCMA17] a set of non-linear transformations is introduced to deal with hyperbolic problems. The focus is on fluid-dynamics problems presenting a shock whose position may vary according to the parameters. This is well illustrated by an application on transonic flows around airfoils. In a recent work [CMS19] the authors propose a general framework to look for reference configurations (not necessarily one) and for transformations depending upon few parameters in order to deal with advection dominated problems.

In what follows two contributions are presented: in the first one, a non-linear mapping is introduced, built by considering the Wassertein distance between the snapshots; in the second, a dynamical basis technique is defined by considering a numerical approximation of the Lax pairs notion.

2.3.1 Using distances different than L^2

The classical model reduction strategies are based on the definition of possibly small dimensional linear subspaces of a Hilbert space (identified in a semi-empirical manner) and of Galerkin or Petrov-Galerkin approximations. The modes, which are the basis functions of the subspace used for the approximation, are often built by linear combination of pre-computed solutions. In [21] a contribution is proposed in which a non-linear transformation is built, based by optimal transport theory, in order to map, at best, the solution snapshots in a reference configuration. The rationale is the following: in the reference configuration, all the snapshots are very similar (in

an L^2 sense), and hence, we can expect that the remapped snapshot set has a better Kolmogorov widths decay than the snapshots set itself. In the reference configuration, a classical strategy can be used. The object of the contribution is how to build a parametric non-linear transformation based on optimal transport. The use of the Wasserstein distance has been proposed in several works. We use some techniques introduced in [MVL11], in which the Wasserstein distance was used to analyse time series and dynamical system behaviour. Let $u_1, \dots, u_{n_s}$ be a set of snapshots, supposed to be densities (for sake of simplicity).

The first step of the method consists in proposing an approximation structure. Let $\xi \in \Omega_0$ denote a reference configuration, to be defined. Let $x = X(\xi, t) \in \Omega$ denote a one-to-one mapping from the reference to the actual configuration, in which the snapshots are given. Let $\xi = Y(x, t)$ be the inverse mapping, $Y(x, t)$ are the backward characteristics. The approximation $\tilde{u}$ of the solution is:

$$\tilde{u}(x, t) = (u_0(Y(x, t)) + R(Y(x, t))) \det(\nabla_x Y(x, t)),$$

where u_0 is a barycentric mode, that plays the role of a template, and R is the residual of the transport, that can be interpreted as the discrepancy between the snapshot and the barycentric mode mapped by the backward characteristics. Two elements need to be defined, namely the mapping $Y(x, t)$ and the barycentric mode $u_0(\xi)$.

To this end, in [21] we first construct the direct mapping $x = X(\xi, t)$ and we try to approximate it by a modal decomposition. The key idea is to consider the Wasserstein distance between all the possible pairs of snapshots and use a Multi-Dimensional Scaling. Let X_{ij} be the optimal transport map between u_i and u_j and $D \in \mathbb{R}^{n_s \times n_s}$ be the matrix containing the pairwise distances squared between the snapshots: $D_{ij} = \mathcal{W}_2^2(u_i, u_j)$. Let us introduce vectors $\chi_i \in \mathbb{R}^{n_s}$, that represent the coordinates of the snapshots in a Euclidean space. The objective is to find the vectors, such that $D_{ij} \approx \|\chi_i - \chi_j\|^2_{\ell^{2,n_s}}$. Let $B_{ij} = \langle \chi_j, \chi_i \rangle_{\ell^{2,n_s}}$. It holds:

$$\begin{aligned} J &= I - \frac{1}{n_s} \mathbb{1}\mathbb{1}^T, \\ B &= -\frac{1}{2} JDJ. \end{aligned}$$

Aiming at finding the minimal number of coordinates per snapshot useful to approximate the distances, an eigenvalue decomposition of the matrix B is performed, which is then projected onto the cone of positive definite matrices:

$$\begin{aligned} B &= U \Lambda U^T, \\ \Lambda^+ &= \frac{\Lambda + |\Lambda|}{2}, \\ B^+ &= U \Lambda^+ U^T. \end{aligned}$$

The spectrum Λ^+ provides an information on the approximation of the vectors χ and hence it provides the number of coordinates which are needed, say $m \in \mathbb{N}^*$. In the Euclidean space, the coordinates of the snapshots are approximated by $\tilde{\chi}_i \in \mathbb{R}^m$ and the representation of the template $u_0(\xi)$ is simply the origin $\chi_0 = 0$, that can be expressed as a linear combination of m linearly independent vectors $\chi_{\bar{i}} - \chi_i,\ i \neq \bar{i}$. Let $\beta_j \in \mathbb{R}$ denote the coefficients of the linear

combination; the barycentric density can be obtained by:

$$X_{\bar{i}0} = \sum_{i=1}^{m} \beta_i X_{\bar{i}i},$$
$$u_0(\xi) = \frac{u_{\bar{i}}}{\det(\nabla_\xi X_{\bar{i}0})}.$$

The advection modes are simply defined as the optimal transport maps between u_0 and the m snapshots used to compute it through their representative vectors ξ_i. Once the mappings are computed, the residual for a generic snapshot i is obtained as follows:

$$R_i(\xi) = u_i(X_{0i}) \det(\nabla_\xi X_{0i}) - u_0(\xi).$$

The set of all the residual is reduced by means of a POD. The reduction is performed based on two sets of modes: the ones describing the mappings between the template and the set of snapshots and the one describing the residual fields.

Several examples are proposed in [21] to assess the performances of the method.

In the more recent [14], we make the first steps towards the model reduction of non-linear PDEs in generic metric spaces. The heuristics are based on the observation, on 1D PDEs solutions, that the approximation in a Wasserstein distance sense is much more efficient that the approximation in a L^2 sense. Generalising, and otherwise stated, the widths that could be defined for the approximation of the solution set, based on some distance (non-necessarily a norm induced by a scalar product) could decay much faster than the classical L^2 one. In view of applying this, the notion of exponential and logarithm maps on a manifold are used. The idea which is developed and investigated is the following: the solution $u(x,t;\vartheta)$ can be seen as a weighted baricenter with respect to the distance chosen of the set of snapshots.

One of the main difficulties related to non-linear model reduction (the proposed methods as well as all the other methods based on mappings) is to perform computations in the low-dimensional representation. The main difficulty is the non-linearity that often introduces overheads and stability issues in terms of computation.

2.3.2 Dynamical low rank approximations by Approximated Lax Pairs

When dealing with time dependent systems, one of the possible ways to improve the basis ability to parsimoniously describe the system solution, is to consider a dynamical basis, as proposed for instance in [FL18].

In general, let the solution to be approximated be $u(x,t) \in \mathcal{V}$, where $\mathcal{V}$ is a Hilbert space equipped with the scalar product $\langle u, v \rangle$; let φ denote a basis of of $\mathcal{V}$. A dynamical basis approximation reads:

$$u(x,t) \approx u_n = \sum_{i=1}^{n} a_i(t)\varphi_i(x,t). \tag{2.21}$$

Let u be a solution of an equation of the form $\partial_t u = f(u;x,t)$. When the approximation u_n is inserted into the Equation, we have:

$$\sum_{i=1}^{n} \dot{a}_i \varphi_i + a_i \partial_t \varphi_i = f. \tag{2.22}$$

Several methods are proposed in the literature to specify the basis dynamics $\partial_t\varphi$: in the dynamical POD method, this is done by imposing that $\langle\partial_t\varphi,\varphi\rangle = 0$. In the contribution described in this section the basis dynamics is specified by a different criterion. In particular, a construction inspired by Lax pairs is proposed. Lax pairs are a non-linear analysis construction that make it possible to express in an elegant form the solution of infinite dimensional integrable systems. The notation is synthetically introduced here, to comment upon the heuristics of the method. More details are available in [16]. Let $\mathcal{L}$ be a linear, self-adjoint, inverse compact operator, depending implicitly or explicitly on time. Let φ be the basis that diagonalises it:

$$\mathcal{L}\varphi_i = \lambda_i\varphi_i,\ \forall i \in \mathbb{N}^*.$$

It follows from the assumptions that φ is a complete orthonormal basis of $\mathcal{V}$, for all time instants. When this equation is derived with respect to time, we get:

$$\partial_t\mathcal{L}\varphi_i + (\mathcal{L} - \lambda_i\mathcal{I})\,\partial_t\varphi_i = \partial_t\lambda_i\varphi_i.$$

The fact that the basis is, $\forall t$ orthonormal and complete entails that there exist a skew-adjoint operator such that: $\partial_t\varphi_i = \mathcal{M}\varphi_i$. Henceforth, it holds:

$$\partial_t\mathcal{L}\varphi_i + (\mathcal{L}\mathcal{M} - \mathcal{M}\mathcal{L})\,\varphi_i = \partial_t\lambda_i\varphi_i.$$

Since this holds true for all the elements of the complete basis, the equation can be written directly for the operators:

$$\partial_t\mathcal{L} + [\mathcal{L},\mathcal{M}] = \partial_t\lambda\mathcal{I}, \tag{2.23}$$

were $[A,B]$ denotes the commutator between the operators A, B. For integrable systems, the operator $\mathcal{L}$ depends on $u(x,t)$, and hence implicitly on time. The remarkable property in integrable systems is that $\partial_t\lambda = 0$, the spectrum is the set of the infinite first integrals of the system and the solution u can be reconstructed, at all times, by using these values and the basis φ. Let us observe that, for integrable systems, the projection of the solution onto the moving basis is constant in time. Leveraging this observation to reduced-order models, we have that, for integrable systems, the reduced-order equation to be integrated would be of the form $\dot{a} = 0$. In the approximated Lax pairs method, we make an arbitrary choice for the operator defining the basis: $\mathcal{L} = -\Delta - \chi u\mathcal{I}$, where $\chi \in \mathbb{R}^+$. This means that $\partial_t\mathcal{L} = -\chi f\mathcal{I}$. In general we cannot conclude that the spectrum is constant. A bilinear form is introduced, $b(\varphi_i,\varphi_j) : \mathcal{V}\times\mathcal{V} \to \mathbb{R}$, aiming at defining the time variation of the eigenvalues as well as the representation of the operator $\mathcal{M}$:

$$\langle\partial_t\mathcal{L}\varphi_i,\varphi_j\rangle + \langle[\mathcal{L},\mathcal{M}]\varphi_i,\varphi_j\rangle = \partial_t\lambda\langle\varphi_i,\varphi_j\rangle.$$

By exploiting the orthonormality of the basis at each time, the self-adjoint character of $\mathcal{L}$ as well as the skew-symmetry of the commutator, the following holds:

$$\begin{aligned}\partial_t\lambda_i &= -\chi\langle f,\varphi_i^2\rangle,\ \forall i \in \mathbb{N}^*,\\ (\lambda_j - \lambda_i)\langle\mathcal{M}\varphi_i,\varphi_j\rangle - \chi\langle f,\varphi_i\varphi_j\rangle &= 0,\ i,j \in \mathbb{N}^*,\ i \neq j.\end{aligned}$$

The discretisation of this system of equations leads to the ALP method presented in [16]. In this exploratory work we investigates the possibility of defining the basis at initial time and hence

no *offline* phase is needed. This has the advantage of reducing the overall cost and propose a basis which is adapted, in a way, to the scenario at hand. Another advantage is that, roughly speaking, the basis is following the dynamics, and in some situations $\|\dot{a}\|_{\ell^2}$ decreases significantly with respect to using a fix basis: this implies a more stable or an easier resolution of the reduced equations. When the solution $u(x,t)$ of a problem is decomposed on the time evolving basis, the following holds:

$$\sum_{i=1}^{n} \dot{a}_i \varphi_i + a_i \mathcal{M}\varphi_i = f,$$
$$\dot{a} + Ma = \hat{f},$$

where the second equation has been obtained by projecting the first equation on the basis, $M_{ji} = \langle \partial_t \varphi_j, \varphi_i \rangle$, $\hat{f}_i = \langle f, \varphi_i \rangle$. In this equation we can see that the term Ma, which is normally not present in the static basis discretisations, account for the effects of the moving basis on the representation of the solution on the basis.

The disadvantages of the ALP method are related to the fact that the operator $\mathcal{L}$ is chosen a priori, and hence there is no guarantee that the initial datum can be approximated by a small number of modes. Moreover, making the basis evolve under the action of an operator $\mathcal{M}$ which is approximated *online* increases the computational cost. The method is tested on challenging advection dominated scenarios in [16, 17]. The positive finding is that the method is versatile and more robust in handling situations like localised obstacles or sources, which are difficult to be sampled extensively in methods based on a more classical offline/online decomposition.

We consider here a test which illustrates well the ability of the basis to follow the system dynamics. The domain in the unit square $\Omega = (0,1)^2$ and $x \in \Omega$. The monodomain equations with the FitzHugh-Nagumo model are considered. Let $A_m, \sigma_m, \eta_{0,1}, a, s_0 \in \mathbb{R}^+$ be scalar parameters. Let $\Omega_s \subset \Omega$ be a bounded set strictly contained in the domain. The system reads:

$$\begin{aligned} A_m \left(\partial_t u + I_{ion}(u,w)\right) - \nabla \cdot (\sigma \nabla u) &= A_m I_{app}, \\ \partial_t w - \eta_0(\eta_1 u - w) &= 0, \\ I_{ion}(u,v) = s(x)u(u-a)(u-1) + w, \\ s(x) = s_0 \mathbb{1}_{\Omega/\Omega_s} + s_0 \frac{36x_2 - 7}{20} \mathbb{1}_{\Omega_s}, \\ \sigma = \sigma_m I. \end{aligned}$$

The test consists in considering two cases: the nominal scenario in which $\Omega_s = \emptyset$, and the conductivity $s = s_0$, constant throughout the domain, and a scenario in which Ω_s is non-empty. In the latter case, the conductivity is penalised in a region of the domain, and this make the depolarisation wave behave potentially in a very different way. The main goal is to highlight the difference between a classical approach of model reduction and the proposed dynamical approach in handling a scenario different with respect to the nominal one. This is a case in which there is virtually an infinite number of parameters describing Ω_s, and hence it could be difficult to build an exhaustive *offline* phase.

A POD basis is constructed by using the time snapshots of the evolution of the system in the nominal scenario, and a ROM is obtained by projecting the monodomain equations on the

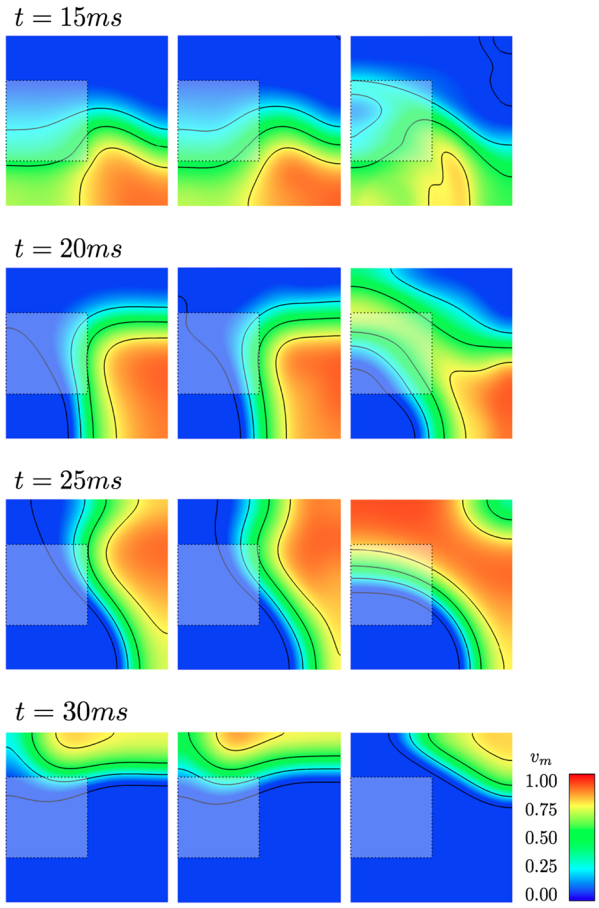


Figure 2.2: Simulation of the Monodomain equation, Section 2.3.2: left column, FEM solution; in the center, the solution obtained by ALP; right column, POD solution.

basis. The non-linearity appearing in the FitzHugh-Nagumo model is a cubic polynomial, so that it can be dealt with by 4–th order projection tensors. An EIM approach has also been tested. The ALP method is initialised by simply computing the eigenvalues and eigenfunctions of the Schroedinger operator with the potential defined as the initial datum $u(x, 0)$:

$$-\nabla \cdot (\sigma \nabla \varphi_i(x, 0)) - \chi u(x, 0)\varphi_i(x, 0) = \lambda_i \varphi_i(x, 0).$$

The basis evolve under the action of the operator $\mathcal{M}$, numerically approximated, and the coefficients of the solution are approximated by solving the equation 2.24.

The results are shown in Figure 2.2. On the left, the reference solution obtained by a standard FEM method is shown. The set Ω_s is a square localised on the left of the domain, depicted in grey in the figure. At the center, we show the approximation obtained by the ALP method and, at the right, the approximation obtained by the POD ROM method. The penalised conductivity region Ω_s make the depolarisation turn and exit the domain in the upper left corner, whereas in the nominal case it exits the domain in the upper right corner. The solution obtained by ALP, in the center, captures all the features of the dynamics of the depolarisation wave; on the contrary, the POD-ROM is not able to retrieve the correct solution.

2.4 Tensor methods

Tensor methods have a long history; in the last decade they had a significant development because of their pertinence in providing a compact representation of large, high-dimensional data. Nowadays, tensor methods are an active field of research in different communities such as applied mathematics and numerical analysis, machine learning, computational physics. As a result of all the efforts, several tensor decompositions and methods are currently available (comprehensive reviews are [KB09, Kho12, Hac12, GKT13]). Contrary to matrices, for tensors of order higher or equal to three there does not exist, in general, a best low rank approximation, the problem being ill posed ([DSL08]). The canonical format (CP, [KB09]) is a natural way to represent a function of multiple variables in a separated form, but for this format there does not exist a best low rank approximation and it is numerically unstable when a large number of variables is involved. In order to circumvent the ill posedeness and the numerical instabilities other formats were proposed. The Tucker format consists in determining a core tensor and in expressing a tensor as product of matrices (called factor matrices) with the core. This is more stable than the CP format. However, it is costly for large systems and high-order tensors, since the core size still scales exponentially with the tensor order (the dimension of the problem). The Tensor Train (TT) format ([Ose11]) is a good trade-off between optimality and stability and consists in computing the tensor entries as a cascade of matrices, and the TT-cores couple two indices at a time. By doing so, the storage complexity is linear with respect to the dimension and scales as a fixed power of the rank (the maximum number of terms needed). It is probably the format which is more often used in the applications. A format which is exploiting tensorization (the *blessing of dimensionality*) and then reduction is the Quantics-TT (QTT) format, described in [Kho11], that exhibits the phenomenon of super-compression.

The High Order Singular Value Decomposition (HOSVD) is one of the possible extensions of the Singular Value Decomposition (SVD) to multi-linear algebra, and it can be considered

a restriction of the Tucker format in which orthogonality and all-orthogonality properties are ensured. It requires, though, the computation of several SVDs for the different possible tensor unfoldings, implying often a prohibitive computational cost when large tensors are considered.

To improve the performances when dealing with high-order tensors, hierarchical formats have been proposed, based on Tucker decomposition and the TT format can be considered as a special case of hierarchical format. In these hierarchical formats, tensor indices (corresponding to coordinate variables when considering a function discretisation) are grouped together and arranged to form a tree. In the same spirit and aiming at generalising this idea, tensor networks have being proposed, in which indices are not arranged in a tree but in a network that has a more general topology. Among these formats we cite PEPS (Projected Entangled-Pair States), MERA (Multi-scale Entanglement Renormalisation Ansatz, that contains cycles), HCL (Honey-Comb Lattice).

In general, in most of the methods, the rank is fixed *a priori*, with the exception of the Density Matrix Renormalization Group (DMRG) and Proper Generalized Decomposition (PGD), explained in [Nou10], and, to some extent, HOSVD. When dealing with the solution of problems, the ability to adapt the rank as function of a prescribed tolerance is particularly important to improve the performances. The dynamical evolution of tensor approximations (with fixed rank) is studied in [LRSV13, LOV15], in which the equations for the dynamics of the factors of TT and Tucker tensors is derived. This can be related to the Dirac-Frenkel variational principle, thus enjoying some optimality properties.

A fundamental building block for the problem solution approximation is linear tensor systems: they are the generalisation to multilinear algebra of linear systems, and consists in finding a tensor which is transformed into a given tensor under the action of a linear tensor operator. Following [GKT13], the methods proposed can be divided into two classes: iterative methods (Richardson, Krylov subspace methods) with truncation in several forms and optimisation by greedy methods. Few works were proposed in the litterature to parallelise tensors computations. Among them we cite: [KS08], in which parallel methods are setup for data mining purposes, [RSSK14], in which memory efficient algorithms are investigated. At present, there are no studies that show definite conclusions on how tensor methods can be systematically parallelised for generic tensor formats.

Tensors in applications are the representation of functions defined over domains in cartesian product. This representation, in view of solving high-dimensional equations was studied in [HK07], where, in particular, the complexity and the storage of collocation approaches with a prescribed accuracy is detailed. A general work concerning the approximation of high-dimensional PDEs by a greedy method is proposed in [LBLM09]. In Kinetic theory, the Boltzmann equation, with a special focus on the collision term approximation, was studied in [Kho07]. A first work on the use of TT approximation to simulate the Vlasov-Poisson system was done in [Kor15a], in which every direction of the phase space is considered separately and the equations in semi-lagrangian form are discretised. Three classes of methods were proposed and compared in [CVK16], in order to approximate the solution of probability density functions evolution. The first class is a separate expansion that can be seen as a greedy approach to determine a tensor in CP format, the second class is a BBGKY expansion and the third class, suitable for higher dimension, is based on ANOVA expansion.

In the field of Uncertainty Quantification (and stochastic equations), a work in the Galerkin

approximation for elliptic PDEs was proposed in [KS11]. Functions of random Gaussian fields and their approximation through tensor methods was studied in [KKNT15]. The application of TT format in UQ is investigated in [BEKM16], in which the idea of spectral approximation in a tensor train format is proposed, with an heuristic on how to order the variables in order to get better compression performances. The TT format was used also to compute response surfaces for parametric models in [DKLM14]. A method to simulate large parametric systems, that combines ANOVA and the TT format were studied in [ZYO$^+$15]. In [BG15] a hierarchical tensor is adaptively construct by exploiting sparse data in order to approximate the quantities of interests of parametric PDEs, with a method which can be seen as an extension to higher-order tensors of cross approximation for matrices. A sparse tensor discretisation based on generalised polynomial chaos was proposed in [SG11]. Hierarchical formats and tensor networks to solve PDEs were investigated in [BSU16].

2.4.1 A first application to Kinetic Theory

Kinetic theory is a quantitative mesoscopic description of many-particle (or, in general, many-body, many-agents) systems. It is a first example of a class of high-dimensional problems.

In this first contribution, we focused on the Vlasov-Poisson equation, which is a model for collisionless plasmas, in electrostatic regime.

Let $d \in \mathbb{N}^*$ denote the spatial dimension of the problem and $\Omega_x, \Omega_v \subseteq \mathbb{R}^d$. The Vlasov-Poisson system for negative electric charges reads:

$$\begin{aligned} \partial_t f + v \cdot \nabla_x f - E \cdot \nabla_v f = 0, \qquad & \text{in } (0,+\infty) \times \Omega_x \times \Omega_v, \\ -\Delta_x \varphi = 1 - \int_{\Omega_v} f \, dv, \qquad & \text{in } (0,+\infty) \times \Omega_x, \\ E = -\nabla_x \varphi, \qquad & \text{in } (0,+\infty) \times \Omega_x, \\ f(0,x,v) = f_0(x,v), \qquad & \text{in } \Omega_x \times \Omega_v, \end{aligned} \tag{2.24}$$

with appropriate boundary conditions on $\Omega_x \times \Omega_v$, where

$$f : \left\{ \begin{array}{ccc} (0,+\infty) \times \Omega_x \times \Omega_v & \to & \mathbb{R} \\ (t,x,v) & \mapsto & f(t,x,v) \end{array} \right.$$

is the particle distribution function in the phase space, $f_0 \geq 0$ is the initial particle distribution function, $E(t,x)$ the electric field and $\varphi(t,x)$ the electric potential. The particle density $\rho(t,x)$ is given by $\rho(t,x) = \int_{\Omega_v} f(t,x,v)\, dv$, and hence, the equation for the electric potential reads $-\Delta_x \varphi = 1 - \rho$.

The global existence of positive (weak or strong) solutions has been studied in several works [Ars75, BD85, DD91, LP91, Gla96, Hwa04, ACF14].

For instance, in [LP91], when $\Omega_x = \Omega_v = \mathbb{R}^3$, the existence of a strong non-negative solution $f \in \mathcal{C}_c(\mathbb{R}_+; L^1(\mathbb{R}^3 \times \mathbb{R}^3)) \cap L^\infty(\mathbb{R}_+ \times \mathbb{R}^3 \times \mathbb{R}^3)$ is proved provided that the initial condition $f_0 \in L^1 \cap L^\infty(\mathbb{R}^3 \times \mathbb{R}^3)$ satisfies the additional condition: for some $m_0 > 3$,

$$\int_{\mathbb{R}^3 \times \mathbb{R}^3} f_0(x,v) \mid v \mid^{m_0} \, dx\, dv < +\infty.$$

The Vlasov-Poisson system admits a reduced hamiltonian formulation [MW82, Mor05]. The Hamiltonian for the Vlasov-Poisson system reads:

$$\mathcal{H} = \int_{\Omega} \frac{1}{2} f \mid v \mid^2 \; dx \; dv - \int_{\Omega_x} \frac{1}{2} \varphi \rho \; dx. \tag{2.25}$$

The first term in the Hamiltonian is the kinetic energy of the particles, and the second term accounts for the electro-static energy. As commented in [MW82], the Vlasov-Poisson equations can be derived by introducing a reduced Poisson bracket:

$$\{a, b\} := \nabla_x a \cdot \nabla_v b - \nabla_v a \cdot \nabla_x b. \tag{2.26}$$

The evolution equation for the system can thus be written as:

$$\partial_t f = - \{f, h\}, \tag{2.27}$$

where $h := \frac{1}{2} \mid v \mid^2 - \varphi$ is the reduced Hamiltonian.

Several methods were proposed in the literature to approximate the solution of the Vlasov-Poisson system. Among them we cite particle methods (Particle-In-Cell [GFA+16, Bra16, CH14], Particle-In-Cloud [WSJY16]), semi-lagrangian approaches [CLS09, CDM13, Kor15b, PM08, CMS10] and full-deterministic Eulerian methods [FS03, MRS14, XOMN10]. We decided to discretise the equations in an Eulerian formulation.

The contribution is threefold:

1. Exploit tensor methods to deal with the problem high-dimensional character. Allow for arbitrary and possibly different discretisations in space and momentum.

2. Build a rank adaptive tensor method, that makes the rank evolve to fulfill a criterion on the residual norm.

3. Construct a time advancing scheme that preserves, at best, the Hamiltonian structure of the Vlasov-Poisson system.

For any measurable functions $r : \Omega_x \to \mathbb{R}$ and $s : \Omega_v \to \mathbb{R}$, we define the tensor product function $r \otimes s : \Omega_x \times \Omega_v \to \mathbb{R}$ as

$$r \otimes s : \left\{ \begin{array}{ccc} \Omega_x \times \Omega_v & \to & \mathbb{R} \\ (x, v) & \mapsto & r(x) s(v). \end{array} \right.$$

In the sequel, such a function is referred to as a *pure tensor-product* function. A linear combination of n pure tensor-product functions (for some $n \in \mathbb{N}^*$) is called a *rank-n* tensor product function.

We also introduce here the notion of tensorized operators. Let H_x (respectively H_v) be a Hilbert space of real-valued functions defined on Ω_x (respectively on Ω_v) and H a Hilbert space of functions defined on $\Omega_x \times \Omega_v$ so that $H_x \otimes H_v \subseteq H$. An operator A acting on functions depending on both x and v variables is a *tensorized operator* if it can be written as

$$A = \sum_{\lambda=1}^{L} A_x^{\lambda} \otimes A_v^{\lambda},$$

for some $L \in \mathbb{N}^*$, where for all $1 \leq \lambda \leq L$, A_x^λ (respectively A_v^λ) is an operator on H_x (respectively on H_v). Let us remind the reader that for all operators A_x on H_x, A_v on H_v, and $(r, s) \in H_x \times H_v$,

$$(A_x \otimes A_v)(r \otimes s) = (A_x r) \otimes (A_v s).$$

The structure of the approximation of the solution of the Vlasov-Poisson system is *a priori* chosen as:

$$f(x, v, t) \approx \sum_{k=1}^{n(t)} r_k(x, t) s_k(v, t) = \sum_{k=1}^{n(t)} r_k(\cdot, t) \otimes s_k(\cdot, t), \tag{2.28}$$

where $n(t) \in \mathbb{N}^*$ is the rank, depending possibly upon time. When this expression is inserted into the evolution equation written in a hamiltonian form, it reads:

$$\partial_t f = -\{f, h\} \approx \sum_{k=1}^{n(t)} -\{r_k, h\} s_k - r_k \{s_k, h\}, \tag{2.29}$$

which means that a tensorised decomposition induces a natural splitting, and that the equation $\partial_t f$ can be seen as a tensor operator. A Symplectic scheme (Strömer-Verlet, [HLW06]) is proposed: set $f^{(0)} := f_0$ and for all $m \in \mathbb{N}$, define

$$\begin{aligned} \left(I - \frac{\Delta t}{2} E^{(m)} \cdot \nabla_v\right) f^{(m+1/3)} &= \left(I - \frac{\Delta t}{2} v \cdot \nabla_x\right) f^{(m)}, \\ \left(I + \frac{\Delta t}{2} v \cdot \nabla_x\right) f^{(m+2/3)} &= f^{(m+1/3)}, \\ f^{(m+1)} &= \left(I + \frac{\Delta t}{2} E^{(m+2/3)} \cdot \nabla_v\right) f^{(m+2/3)}. \end{aligned} \tag{2.30}$$

The function $f^{(m)}$ then gives an approximation of the solution f to (2.24) at time $t_m := m\Delta t$. Observe from the semi-discretised in time formulation that, at each stage of the time advancing scheme, tensorised operators act on the solution approximation f. A tensor-based method is introduced to solve the following problem: find $f \in H$ solution of

$$(I + \Delta t P) f = g, \tag{2.31}$$

where

- g is a finite-rank tensor product element of H;
- $\Delta t \geq 0$ is a small constant;
- I is an operator on H of the form $I = I_x \otimes I_v$ where I_x (respectively I_v) is a symmetric continous coercive operator on H_x (respectively H_v);
- P is an arbitrary *tensorized* operator on H (not necessarily symmetric).

We still assume that we start from an initial guess for f given by an element $g_0 \in H$. Assume now that f is the solution of a problem of the form

$$\forall g \in H, \quad \langle f, g \rangle_H + \widetilde{\mathbf{a}}(f, g) = \mathbf{b}(g), \tag{2.32}$$

where $\mathbf{b}$ is a continuous linear form on H and $\widetilde{\mathbf{a}} : H \times H \to \mathbb{R}$ is a continuous bilinear form which is not symmetric nor coercive in general. There exists a unique solution of this problem for instance when $\|\widetilde{\mathbf{a}}\|_{\mathcal{L}(H\times H;\mathbb{R})} < 1$.

A natural idea to solve (2.32) when the operator is a *small* perturbation of the identity operator on H is to consider the following fixed-point PGD algorithm:

Algorithm 1 Fix-point PGD

1: **Initialise:** Set $n = 0$ and $f_0 = g_0$.
2: **Iterate** $n > 0$:
3: Compute $(r_{n+1}, s_{n+1}) \in H_x \times H_v$:

$$(r_{n+1}, s_{n+1}) \in \underset{(r,s)\in H_x\times H_v}{\operatorname{argmin}} \mathcal{E}_n(r \otimes s), \tag{2.33}$$

where for all $g \in H$, $\mathcal{E}_n(g) = \frac{1}{2}\langle f_n + g, f_n + g\rangle_H - \mathbf{b}(f_n + g) - \widetilde{\mathbf{a}}(f_n, f_n + g)$
4: Define: $f_{n+1} := f_n + r_{n+1} \otimes s_{n+1}$
5: Define: $n \leftarrow n + 1$.

The following Proposition holds, that states that the iteration of the algorithm are well defined, and highlight a CFL-like condition for the time advancing scheme:

4. *Proposition.* *All the iterations of the Fixed-point PGD algorithm are well-defined. Moreover, let us assume that* $\widetilde{A} = \sum_{\mu=1}^{M} \widetilde{A}_x^\mu \otimes \widetilde{A}_v^\mu$ *where for all* $1 \le \mu \le M$, $\widetilde{A}_x^\mu \in \mathcal{L}(H_x; H_x)$ *and* $\widetilde{A}_v^\mu \in \mathcal{L}(H_v; H_v)$. *Let* $\kappa := \max_{1\le\mu\le M} \left\| \widetilde{A}_x^\mu \otimes \widetilde{A}_v^\mu \right\|_{\mathcal{L}(H;H)}$. *Assume that at least one of these two assumptions is satisfied:*

(A1) $H = H_x \otimes H_v$ *and* $3M\kappa < 1$*;*

(A2) $5M\kappa < 1$.

Then, there is a unique solution f *to the problem* (2.32) *and the sequence* $(f_n)_{n\in\mathbb{N}}$ *strongly converges in* H *to* f.

The method we propose for the resolution of the Vlasov-Poisson system is summarized hereafter, after having introduced suitable space and velocity discretisations. All the details are reported in [13]. Let $\epsilon > 0$ be a chosen tolerance threshold.

Algorithm 2 Verlet-PGD-ϵ

Initialisation: Set $f^{(0)} = f_0$.
while $m \leq M$ **do**
  Define: $P^{(m+1/3)} = \frac{\Delta t}{2} \sum_{\alpha=1}^{d} F_x(E_\alpha^{(m)}) \otimes D_{\alpha,v}$
  Define: $g^{(m+1/3)} = \left(I_x \otimes I_v - \frac{\Delta t}{2} V_{\alpha,v} \otimes D_{\alpha,x}\right) f^{(m)}$.
  Compute: $\overline{f}^{(m+1/3)} = \mathrm{PGD}_{FP}(P^{(m+1/3)}, g^{(m+1/3)}, f^{(m)}, \epsilon, \epsilon)$.
  Recompress: $f^{(m+1/3)} = \mathrm{POD}\left(\overline{f}^{(m+1/3)}, \epsilon, \epsilon\right)$.
  Define: $Q^{(m+1)} = -\frac{\Delta t}{2} \sum_{\alpha=1}^{d} F_x(E_\alpha^{(m+2/3)}) \otimes D_{\alpha,v}$.

  Define: $P^{(m+2/3)} = \frac{\Delta t}{2} \sum_{\alpha=1}^{d} D_{\alpha,x} \otimes V_{\alpha,v}$.
  Compute: $\overline{f}^{(m+2/3)} = \mathrm{PGD}_{FP}(P^{(m+1/3)}, f^{(m+1/3)}, f^{(m+1/3)}, \epsilon, \epsilon)$.
  Recompress: $f^{(m+2/3)} = \mathrm{POD}\left(\overline{f}^{(m+2/3)}, \epsilon, \epsilon\right)$.

  Define: $Q^{(m+1)} = -\frac{\Delta t}{2} \sum_{\alpha=1}^{d} F_x(E_\alpha^{(m+2/3)}) \otimes D_{\alpha,v}$.
  Compute: $f^{(m+1)} = POD((I_x \otimes I_v + Q^{(m+1)}) f^{(m+2/3)}, \epsilon, \epsilon)$.

Output: $f^{(0)}, \ldots f^{(M)}$.

Several numerical tests were performed. We report here the most significant findings, on a 3D-3D Landau damping test case. The initial condition is chosen to be

$$f_0(x,v) = \frac{0.5}{\sqrt{2\pi}^3} \left(1 - \beta_1 \sin(\omega x_1) - \beta_2 \sin(\omega x_2) - \beta_3 \sin(x_3)\right) \exp(-0.5(v_1^2 + v_2^2 + v_3^2)),$$

with $\omega = 0.2$, $\beta_1 = \beta_3 = 10^{-3}$, $\beta_3 = 5.10^{-3}$. Two sets of space-velocity discretisation parameters have been chosen: $(N_x, N_v) = (32^3, 32^3), (64^3, 64^3)$. Let us point out that the last test involves $64^6 \approx 69.10^9$ degrees of freedom and that the simulations were run on a single core. Figure 2.3 illustrates the evolution of the ranks with respect to time. Figure 2.4 represents the fluctuations of the density $1 - \rho(t,x) = 1 - \int_v f(t,x,v)\, dv$ at the following times $t = 0, 0.33, 0.67, 1.0$.

The tests performed in the context of the Vlasov-Poisson systems drove us to the following partial conclusions:

1. Adaptive and dynamical approximations are an interesting tool to build parsimonious discretisations; fixing the rank *a priori* could lead to severe drawbacks.

2. The method proposed have significant performances when considering short time simulations. As it can be seen in Figure 2.3, the rank of the tensor approximation is constantly increasing. Constructing better approximations is the object of the ADAPT project presented in the next section.

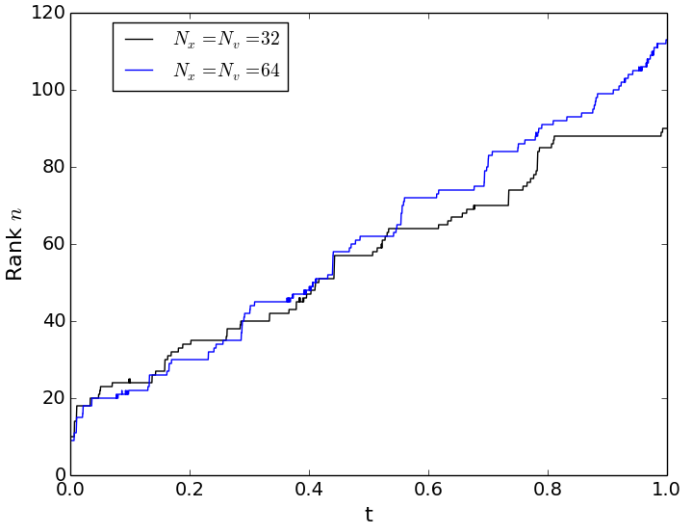


Figure 2.3: Evolution of the rank of the approximate solution for the 3D-3D Landau damping test case as a function of time.

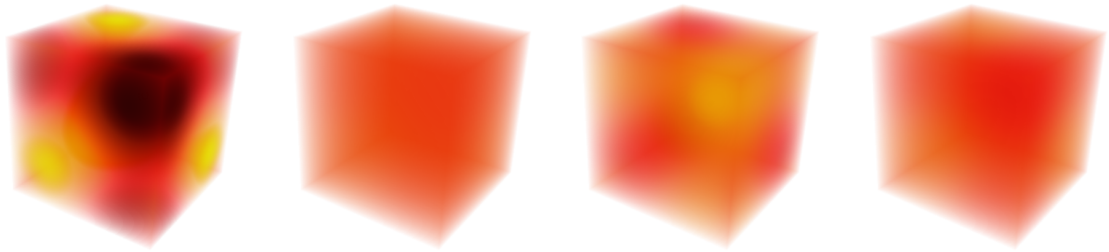


Figure 2.4: Fluctuations of density $1 - \rho(t, x)$ for the 3D-3D Landau damping test case at times $t = 0, 0.33, 0.67, 1.0$ from left to right.

2.4.2 The ADAPT project

The project ADAPT (Adaptive Dynamical Approximations via Parallel Tensor methods) is a ANR JCJC ongoing project, started in december 2018, investigating tensor methods to build parsimonious versatile discretisations for high-dimensional PDEs. Tensor formats have been intensively developed, leading to the introduction of many ways to exploit separation of variables and, in general, multilinear algebra, to provide a representation of high dimensional data.

To this end and to make tensor methods efficient to solve realistic scale problems by exploiting at best the computational power available, two main points are addressed in ADAPT:

1. The tensor construction and the multi-linear algebra operations involved when solving high-dimensional problems are still sequential in most of the cases. Indeed, they rely on successive applications of Singular Value Decomposition (SVD) which is intrinsically sequential. This can be a bottleneck in view of large realistic applications, since the parallelisation of the SVD steps (as well as the other operations) may not provide a sufficient gain and result in a poor scalability. The objective is to design efficient parallel methods for tensor construction and computations.

2. When solving high-dimensional problems, the tensor is not assigned; instead, it is specified through a set of equations and tensor data. As consequence it is often impossible to know *a priori* what is the most efficient format and, given a format, it is difficult to assess *a priori* the rank (the number of terms in the tensor decomposition). What it is known is a prescribed accuracy on the solution representation. Therefore, our goal is to devise numerical methods able to (dynamically) adapt the rank and the discretisation (possibly even the tensor format) to respect the chosen error criterion. This could, in turn, improve the efficiency and reduce the computational burden.

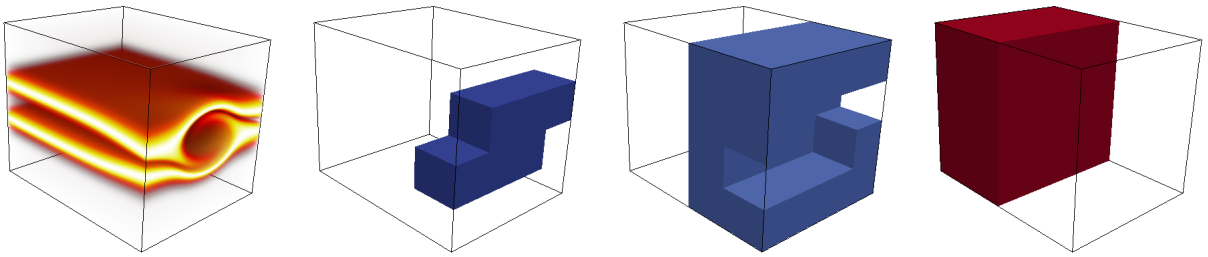
These sought improvements could make the definition of parsimonious discretisations for Kinetic Theory and Uncertainty Quantification problems more efficient and suitable for a High Performance Computing paradigm.

The first work proposed is presented in [12], and it consists in the introduction of an adaptive piece-wise tensor approximation. Neither the subtensor partitioning nor the rank of the approximation of each subtensor are fixed *a priori*. Instead, they are determined in order to fulfill a prescribed error criterion and optimise, to some extent, the storage. Two contributions are proposed:

1. Given a subtensor partitioning, a greedy algorithm is proposed to construct a piece-wise approximation, based on parsimony, such that the approximation error on the tensor is respected.

2. A method is proposed to optimise the partitioning tree by fusing together subtensors for which a low-rank approximation on the merged subtensor allow to save some memory while guaranteeing the same precision.

The observation which is at the genesis of the method is the following: given a multivariate function to be approximated, such as, for instance, the solution of the Vlasov-Poisson system



(a) (b) (c) (d)

Figure 2.5: Vlasov-Poisson solution, section 2.4.2: (a) the tensor entries, in red the largest entries; (b) the small size subtensors, (c) and (d) the mid size and the larger size subtensors. The largest subtensors are in the complement of the cube.

presented above, it is not true, in general, that it can be well approximated by a global low rank tensor. Instead, it can be true that there exists a domain partitioning such that a low rank approximation can be defined on each subdomain. To make this heuristic precise, a sufficient condition on the regularity of the function and on the norms used to measure the error is investigated. The result is summarised in the following:

5. *Proposition.* *Let* $\Omega_1 = \cdots = \Omega_d = (0,1)$ *so that* $\mathbf{\Omega} := (0,1)^d$. *For all* $M \in \mathbb{N}^*$, *let us consider* $P^M := \{(\frac{m-1}{M}, \frac{m}{M})\}_{1\leq m\leq M}$ *be a collection of subsets of* $(0,1)$ *and let* $\boldsymbol{P}^M$ *be the tensorized domain partition of* $\mathbf{\Omega}$ *associated to the collection of domain partitions* $(P_j)_{1\leq j\leq d}$ *where* $P_j = P^M$ *for all* $1 \leq j \leq d$. *Let* $k \in \mathbb{N}^*$, $1 \leq p \leq q \leq \infty$ *and* $\varepsilon > 0$. *We denote* $\lambda := \frac{k}{d} - \frac{1}{p} + \frac{1}{q} > 0$. *Let* $\mathcal{F} \in W^{k,p}(\mathbf{\Omega})$ *such that* $\|\mathcal{F}\|_{W^{k,p}(\mathbf{\Omega})} \leq 1$. *Then, there exists a constant* $C > 0$ *which depends only on* k, p, d, q *such that for all* $M \in \mathbb{N}^*$ *such that* $\ln M \geq -\frac{1}{d\lambda}\ln\left(\frac{\varepsilon}{C}\right)$, *there exists a tensor* $\mathcal{F}^{CPF}$ *of Canonical Partitioning Format with domain partition* $\boldsymbol{P}^M$ *and rank* $R \leq \frac{(k-1+d)!}{(k-1)!d!}$ *such that*

$$\|\mathcal{F} - \mathcal{F}^{CPF}\|_{L^q(\mathbf{\Omega})} \leq \varepsilon. \tag{2.34}$$

The details of the proof are proposed in [12]. The result states simply that if the embedding $W^{k,p}(\Omega) \hookrightarrow L^q(\Omega)$ is compact (Rellich-Kondrakov theorem), than the heuristic is true.

All the details about the method and the algorithms proposed can be found in [12]. Hereafter, we detail the two main algorithms (Algorithms 3-4). In these, the tensor approximation constructed in each of the subdomains is in HOSVD format. The modifications to adapt these algorithms to other tensor formats are straightforward.

Several numerical experiments are proposed therein to assess the performances of the method. We report here the case of the adaptive compression of a solution of a Vlasov-Poisson system.

In Figure 2.4.2.a) the solution of a Vlasov-Poisson equation is shown. The test is a 1D-1D benchmark, called double stream instability. For this test is can be seen that, for short

Algorithm 3 PF-Greedy-HOSVD

Input:
$\mathcal{A} \in \mathbb{R}^{\boldsymbol{I}} \leftarrow$ a tensor of order d
$\mathcal{P} \leftarrow$ an admissible partition of $\boldsymbol{I}$
$\varepsilon > 0 \leftarrow$ error tolerance criterion

Output:
Set of Ranks $(\boldsymbol{R}^{\boldsymbol{J}})_{\boldsymbol{J}\in\mathcal{P}} \subset \mathbb{N}^d$
Set of local errors $(\varepsilon^{\boldsymbol{J}})_{\boldsymbol{J}\in\mathcal{P}}$ satisfying $\sum_{\boldsymbol{J}\in\mathcal{P}} |\varepsilon^{\boldsymbol{J}}|^2 < |\varepsilon|^2$.

Begin:
Set $\boldsymbol{R}^{\boldsymbol{J}} := (0, \cdots, 0)$ for all $\boldsymbol{J} \in \mathcal{P}$
while $\displaystyle\sum_{\boldsymbol{J}\in\mathcal{P}} \sum_{1\le j\le d} \sum_{R_j^{\boldsymbol{J}}+1\le q_j\le p_j(\mathcal{A}^{\boldsymbol{J}})} \left|\sigma_j^{q_j}(\mathcal{A}^{\boldsymbol{J}})\right|^2 \ge \varepsilon^2$ **do**
    Select $1 \le j_0 \le d$ and $\boldsymbol{J}_0 \in \mathcal{P}$ such that

$$(j_0, \boldsymbol{J}_0) = \underset{1\le j\le d,\ \boldsymbol{J}\in\mathcal{P}}{\operatorname{argmax}} \sigma_j^{R_j^{\boldsymbol{J}}+1}(\mathcal{A}^{\boldsymbol{J}}).$$

    if $\boldsymbol{R}^{\boldsymbol{J}_0} = (0, \cdots, 0)$ **then**
        Set $\boldsymbol{R}^{\boldsymbol{J}_0} := (1, \cdots, 1)$
    else
        Assume that $\boldsymbol{R}^{\boldsymbol{J}_0} = (R_1^{\boldsymbol{J}_0}, \cdots, R_d^{\boldsymbol{J}_0})$.
        $R_{j_0}^{\boldsymbol{J}_0} \leftarrow R_{j_0}^{\boldsymbol{J}_0} + 1$.
Define $\varepsilon^{\boldsymbol{J}} := \sqrt{\sum_{1\le j\le d} \sum_{R_j^{\boldsymbol{J}}+1\le q_j\le p_j(\mathcal{A}^{\boldsymbol{J}})} \left|\sigma_j^{q_j}(\mathcal{A}^{\boldsymbol{J}})\right|^2}$ for all $\boldsymbol{J} \in \mathcal{P}$.
return $\left(\boldsymbol{R}^{\boldsymbol{J}}\right)_{\boldsymbol{J}\in\mathcal{P}}$ and $\left(\varepsilon^{\boldsymbol{J}}\right)_{\boldsymbol{J}\in\mathcal{P}}$

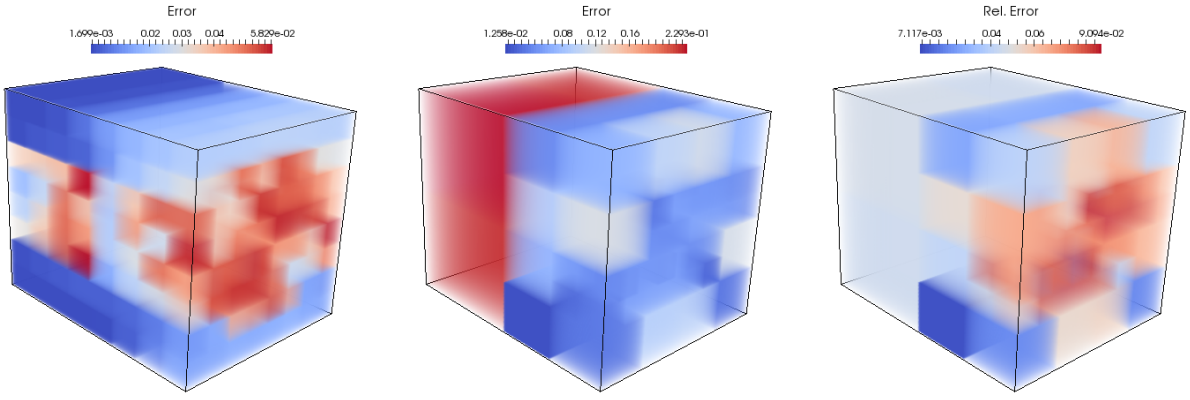
time, the solution is well tensorised in $x - v$, hence being well approximated by a low rank CP. On the contrary, for larger time, it develops a vortex-like structure in space-momentum, that makes it intrinsically high-rank. The method, as it can be seen in Figures 2.4.2.b-d) is able to automatically partition the domain and detect the regions in which the solution is low-rank. In Figure 2.4.2.a-b) the error is shown after the first phase of the method (a greedy algorithm applied to the maximally partitioned domain) and the error after the optimisation of the partitioning tree. In Figure 2.4.2.c) the error renormalised with the subdomain size is shown. The method tends to uniformly distribute the error with respect to the subtensor size, which seems to be beneficial in optimising the storage.

Algorithm 4 PF-MERGE

Input:
$\mathcal{A} \in \mathbb{R}^{\boldsymbol{I}} \leftarrow$ a tensor of order d
$T_{\boldsymbol{I}}^{\text{init}}$ an initial partition tree of $\boldsymbol{I}$
$\varepsilon > 0 \leftarrow$ error tolerance criterion

Output:
$T_{\boldsymbol{I}}$ a final partition tree of $\boldsymbol{I}$
A set of leaf errors $(\varepsilon^{\boldsymbol{J}})_{\boldsymbol{J} \in \mathcal{L}(T_{\boldsymbol{I}})}$
A set of leaf ranks $\left(\boldsymbol{R}^{\boldsymbol{J}}\right)_{\boldsymbol{J} \in \mathcal{L}(T_{\boldsymbol{I}})} \subset \mathbb{N}^d$

Begin:
Set $T_{\boldsymbol{I}} = T_{\boldsymbol{I}}^{\text{init}}$.
Compute $((\boldsymbol{R}^{\boldsymbol{J}})_{\boldsymbol{J} \in \mathcal{L}(T_{\boldsymbol{I}})}, (\bar{\varepsilon}^{\boldsymbol{J}})_{\boldsymbol{J} \in \mathcal{L}(T_{\boldsymbol{I}})}) = \text{PF-Greedy-HOSVD}(\mathcal{A}, \mathcal{L}(T_{\boldsymbol{I}}), \varepsilon)$
Compute $\eta^2 := \varepsilon^2 - \sum_{\boldsymbol{J} \in \mathcal{L}(T_{\boldsymbol{I}})} |\bar{\varepsilon}^{\boldsymbol{J}}|^2$.
For all $\boldsymbol{J} \in \mathcal{L}(T_{\boldsymbol{I}})$, define $\varepsilon^{\boldsymbol{J}} := \sqrt{|\bar{\varepsilon}^{\boldsymbol{J}}|^2 + \frac{|\boldsymbol{J}|}{|\boldsymbol{I}|}\eta^2}$.
Set $\mathcal{N}_{\text{totest}} = \mathcal{L}^p(T_{\boldsymbol{I}})$ and $\mathcal{N}_{\text{nomerge}} = \emptyset$.
while $\mathcal{N}_{\text{totest}} \neq \emptyset$ **do**
    Choose $\boldsymbol{J}_0 \in \mathcal{N}_{\text{totest}}$.
    $(T_{\boldsymbol{I}}^{\text{fin}}, merge, (\varepsilon^{\text{fin},\boldsymbol{J}})_{\boldsymbol{J} \in \mathcal{L}(T_{\boldsymbol{I}}^{\text{fin}})}, (\boldsymbol{R}^{\text{fin},\boldsymbol{J}})_{\boldsymbol{J} \in \mathcal{L}(T_{\boldsymbol{I}}^{\text{fin}})}) = \text{MERGE}(\mathcal{A}, T_{\boldsymbol{I}}, (\varepsilon^{\boldsymbol{J}})_{\boldsymbol{J} \in \mathcal{L}(T_{\boldsymbol{I}})}, (\boldsymbol{R}^{\boldsymbol{J}})_{\boldsymbol{J} \in \mathcal{L}(T_{\boldsymbol{I}})}, \boldsymbol{J}_0)$
    if $merge = \text{true}$ **then**
        $T_{\boldsymbol{I}} = T_{\boldsymbol{I}}^{\text{fin}}$
        $\mathcal{N}_{\text{totest}} = \mathcal{L}^p(T_{\boldsymbol{I}}^{\text{fin}}) \setminus \mathcal{N}_{\text{nomerge}}$.
        $(\varepsilon^{\boldsymbol{J}})_{\boldsymbol{J} \in \mathcal{L}(T_{\boldsymbol{I}})} = (\varepsilon^{\text{fin},\boldsymbol{J}})_{\boldsymbol{J} \in \mathcal{L}(T_{\boldsymbol{I}}^{\text{fin}})}$, $(\boldsymbol{R}^{\boldsymbol{J}})_{\boldsymbol{J} \in \mathcal{L}(T_{\boldsymbol{I}})} = \left(\boldsymbol{R}^{\text{fin},\boldsymbol{J}}\right)_{\boldsymbol{J} \in \mathcal{L}(T_{\boldsymbol{I}}^{\text{fin}})}$
    else
        $\mathcal{N}_{\text{nomerge}} = \mathcal{N}_{\text{nomerge}} \cup \{\boldsymbol{J}_0\}$; $\mathcal{N}_{\text{totest}} = \mathcal{N}_{\text{totest}} \setminus \{\boldsymbol{J}_0\}$.
    return $T_{\boldsymbol{I}}$, $(\varepsilon^{\boldsymbol{J}})_{\boldsymbol{J} \in \mathcal{L}(T_{\boldsymbol{I}})}$, $(\boldsymbol{R}^{\boldsymbol{J}})_{\boldsymbol{J} \in \mathcal{L}(T_{\boldsymbol{I}})}$



(a) (b) (c)

Figure 2.6: Vlasov-Poisson case, section 2.4.2:

Chapter 3

Dealing with uncertainty.

Context

When considering biophysical systems, variability in the measured quantities is an undeniable fact. The first cause is a natural variability, which is an intrinsic component of living systems. As a paradigmatic example, consider the action potential of one single cell, or the cardiac heart rate of the human body. Beside the fact that the system is behaving in a non-deterministic way, there are other sources of variability, which are related to the ability of knowing the system behaviour and describe it. When defining a model of a system, there is a inevitable model error. Moreover, measurements are always partial and corrupted by noise.

The first works in the literature considered the problem of the Forward Uncertainty Quantification. This consists in assuming that parameter and data have a certain probability density distribution and the model outcome are estimated. A generic presentation of Uncertainty Quantification (UQ) can be found in [GHO17, Smi13, LPS14]. One of the first works proposing a Galerkin approximation is [XK02]. The need of developing efficient methods was investigated in numerous works, among which [Xiu07, XH05, BNTT11, NLM09, BG04] . An analysis of polynomial chaos expansions is proposed in [EMSU12]. More recently, a Multi-level approach have been extensively studied to deal with scenarios in which a hierarchy of models can be constructed, [Gil08, CHAN$^+$15, TSGU13]. The obtained solutions can be exploited, for instance, to perform a sensitivity analysis [CLMM09], which is a step to move from forward uncertainty quantification to another kind of analysis, related to data assimilation. An example of application in biomedical engineering is proposed in [SKM$^+$15].

In this manuscript, three main contributions are described, which are not related to Forward Uncertainty Quantification. The first contribution concerns identifiability: given a model, which is supposed to describe the system behaviour sufficiently well, can we estimate the QoIs by using the available observations? This is a property of the model used and the available observations. It is particularly pertinent in biomedical applications, since, usually, we do not have one model, but a hierarchy of possible models.

3.1 Information theoretic quantities and practical identifiability

When considering a dynamical system, the ability to retrieve the value of hidden parameters or, more generally a set of QoI, by exploiting available measurements, is referred to as identifiability.

The objective of this contribution is threefold:

1. Use Information Theory to quantify practical identifiability in noisy scenarios, potentially when the model is a system of parametric PDEs.
2. Introduce a notion of virtual measurement tool, making sense of the identifiability estimation for practical purposes.
3. Modify the classical estimator to better adapt to the model based applications.

3.1.1 Entropy Equivalent Variance.

Structural identifiability was first introduced by Bellman in [BÅ70] and it consists in assessing, in a noiseless setting, whether the observable to QoI map is injective. This can be done, based on differential algebra, when systems of parametric Ordinary Differential Equations are in polynomial or rational form ([Něm10, MEDI09]). For more general systems of ODEs or for systems described by parametric Partial Differential Equations, structural identifiability is still an open problem. Moreover, when considering realistic applications, observations are always affected by some noise. Practical identifiability analyses were proposed in several works, among which we cite [BRK01, RKM$^+$09], in which local sensitivity or profile likelihood are used to measure a degree of identifiability.

In what follows, we focus on this problem: the QoI to be estimated are hidden parameters of the dynamical system of interest. Henceforth, they influence the system evolution and potentially the observables. Otherwise stated, observables convey a certain amount of information about the parameters that determine the system state evolution. This vague sentence needs to be made precise. In particular, the goal is to provide a rigorous characterisation about this information, and propose a way to quantify it.

The starting point for this study is to consider a hidden parameter ϑ as a random variable, with prior density distribution $p(\vartheta)$. When an observation $y \in \mathbb{R}^m$ is available, the posterior density distribution $p(\vartheta|y)$ can be computed by means of the Bayes theorem. The notion of differential entropy is introduced to quantify the amount of uncertainty and express the gain in information brought by the observation as the reduction in the amount of uncertainty:

$$H(p(x)) = \int_\Omega p(x) \log\left(\frac{1}{p(x)}\right)\, dx. \tag{3.1}$$

In [Jay57a, KL51, Hob69, HC73] several notions of differential entropy are analysed and compared. The remarkable fact is that, when the gain in information between prior and posterior is computed, the different possible definitions of differential entropy lead to the same expression:

$$\Delta H = H(p(\vartheta)) - H(p(\vartheta|y)). \tag{3.2}$$

Identifiability analysis is a form of *a priori* analysis, and it describes a property of the system and the kind of observation performed on it. In this sense, the key remark proposed in [27] is that, when identifiability is analysed, the gain in information have to be computed prior of receiving any specific observation. Henceforth, we propose to assess the expected gain in information, *i.e.* the gain in information averaged among all the possible observations coming from the system (whose evolution depends upon ϑ). This leads precisely to the mutual information between the parameters and the observables. Several combinations of parameters and conditional mutual information between them can be estimated to provide a quantification of practical identifiability in complex scenarios, when noisy observations are available, and analyse compensation effects between parameters.

An important point is the ability to make sense of an expected gain of information and to interpret it. Having gained a certain amount of *bits* or *nats* does not provide a direct intuition on the fact that a parameter can be identified or not. To this end the notion of Entropy Equivalent Variance (EEV) is introduced. Performing a Bayesian parameter estimation problem is like building a virtual measurement device that performs a direct observation of the parameters. Let us make the hypothesis that this device provides noisy measurements, with an additive independent Gaussian noise. What would the variance of these measurements be if the device measured the parameters by getting the same expected gain in information as the Bayesian estimation? This variance is called EEV.

In [27], the quantification of the mutual information is performed by using the KSG estimator (presented in [KSG04]). The method is based on nearest neighbors and it is an extension of the work proposed in [KL87]. Several numerical experiments are described in [27], both on systems of ODEs and PDEs. In particular, it can be observed that different physical regimes are characterised by different degrees of identifiability, and that the method provides estimations in accordance with intuition. In the tests, it emerges that the estimators proposed in the literature (the KL estimator for the entropy, and the KSG for the mutual information) are not working in an ideal setting. In particular, in all the cases we are interested in a parametric model describing the system induces a joint parameter-observable density $p(\vartheta, y)$ which is concentrated around a small dimensional manifold embedded in the Euclidean space containing the domains of parameters and observables. This is what motivates the methodological investigation on entropy and mutual information estimation based on samples. A first work in this sense is presented in [25] and the main contribution is detailed hereafter.

3.1.2 A modified nearest-neighbours entropy estimator.

Let $i \in \mathbb{N}^*,\ i = 1, \dots, N_s$ and let the samples of a random vector be denoted by $x^{(i)} \in \mathbb{R}^n$. The probaility density distribution is denoted by $p(x)$. The goal is to estimate the differential entropy based on a Monte-Carlo method that reads:

$$\hat{H} = \sum_{i=1}^{N_s} \log\left(\frac{1}{\log(p(x^{(i)}))}\right).$$

To this end, it is key to have an estimation of the logarithm of the probability density evaluated at the sample $x^{(i)}$. The KL and KSG estimators provides that based on the KNN algorthm and the mathematical relationships that describe the samples distribution as function of the density.

The ball of radius ε_i is constructed around the sample $x^{(i)}$, such that it contains $k \in \mathbb{N}^*$ neighbors:

$$\mathcal{B}^q(x^{(i)}, \varepsilon_i) = \left\{ x \in \mathbb{R}^n \text{ such that : } \|x - x^{(i)}\|_{\ell^{q,n}} \leq \varepsilon_i \right\}.$$
$$\sharp \left\{ x^{(j)} \in \mathcal{B}^q(x^{(i)}, \varepsilon_i) \right\} = k, \ \forall i.$$

The probability mass of the ball $\mathcal{B}^q(x^{(i)}, \varepsilon_i)$ is defined as:

$$P_i = \int_{\mathcal{B}^q(x^{(i)}, \varepsilon_i)} p(x) dx.$$

The clever mathematical idea proved in [KL87] is that the expected value of the logarithm of the probability mass is related to the Euler digamma function:

$$\mathbb{E}(\log(P)) = \psi(N_s) - \psi(k). \tag{3.3}$$

To relate the probability mass to the probability density the authors of the KL (and KSG) estimator introduce an approximation hypothesis: the density is approximately constant in each ball, so that $P_i = p(x^{(i)})\mu\left(\mathcal{B}^q(x^{(i)}, \varepsilon_i)\right) = p(x^{(i)})\mu_i^{(q)}$, where μ denote the measure of the set (the volume of the ball). Putting the estimates together entails:

$$\hat{H} = \psi(N_s) - \psi(k) + \frac{1}{N_s} \sum_{i=1}^{N_s} \log\left(\mu_i^{(q)}\right).$$

In the case of $\ell^{\infty,n}-$balls, which is the one considered in the present work, the measure of the set reduces to $\mu_i^{(\infty)} = \mu\left(\mathcal{B}^\infty(x^{(i)}, \varepsilon_i)\right) = (2\varepsilon_i)^n$.

The main idea of the contribution presented in [25] is to substitute the constant in the ball approximation by a local Gaussian approximation. In each ball $\mathcal{B}^q(x^{(i)}, \varepsilon_i)$, the density is approximated by a gaussian of the form:

$$g_i = (x - \overline{x}^{(i)})^T S_i^{-1} (x - \overline{x}^{(i)}),$$

where $\overline{x}^{(i)}$ and S_i are the mean and the covariance respectively. These are estimated empirically by considering the first $l \in \mathbb{N}^*$ neighbors (and usually $l > k$). The approximation of the probability density in the $i-$th ball reads:

$$p(x) \approx p(x^{(i)}) \frac{g_i(x)}{g_i(x^{(i)})}.$$

The expression of the resulting estimator is:

$$G_i = \int_{\mathcal{B}^q(x^{(i)}, \varepsilon_i)} g_i(x) \ dx, \ \forall i,$$
$$\hat{H} = \psi(N_s) - \psi(k) - \sum_{i=1}^{N_s} \log\left(g_i(x^{(i)})\right) + \frac{1}{N_s} \sum_{i=1}^{N_s} \log(G_i).$$

Several tests are proposed in [25]. Concerning the choice of the parameter k, its effect is perfectly analogous to the one observed in the classical KL estimator. The choice of the parameter l reflects the characteristic length of the changes in the density $p(x)$, and it can improve in a substantial way the entropy estimation, especially in cases in which the probability density is concentrated around a small dimensional manifold embedded in $\mathbb{R}^n$. A principle of theoretical analysis is reported in [25]. The result reads as follows:

6. *Proposition.* *Let P_i^{KL} denote the probability mass in the $i-$the ball as estimated by the classical KL estimator. Let the probability density be $p \in \mathcal{C}^2$ in each ball. Let the Hessian of the residual of the approximation in each ball be $\nabla^2 R$ such that its maximum and minimum eigenvalues are $\zeta_i^{\max,\min}$. The statistical error due to MC approximation is denoted by e_s. It holds:*

$$|H - \hat{H}| \geq \left| e_s + \frac{d2^{d-1}}{3N} \sum_{i=1}^{N} \frac{\zeta_i^{\min} \varepsilon_i^{d+2}}{P_i^{KL} + \frac{|\zeta_i^{\max}| d2^{d-1}}{3}} \varepsilon_i^{d+2} \right|,$$

$$|H - \hat{H}| \leq e_s + \frac{d2^{d-1}}{3N} \sum_{i=1}^{N} \frac{|\zeta_i^{\max}|}{P_i^{KL}} \varepsilon_i^{d+2}.$$

The tests show that the estimator is robust even in less classical and singular cases, such as the multivariate Beta distribution. The extension of such a method to the estimation of the mutual information is currently under investigation.

3.2 Backward Uncertainty Quantification: moment matching

The forward problem in UQ consists in computing the model output variability given the variability in the parameters and the data of the system. In this section, the backward problem is discussed, namely, given a collection of experimental measurements, the variability in the input parameter and data is computed. The variability in the data can be caused for instance by an heterogeneity in the physical settings [CDE06, SZ06], or fluctuations in the experimental condition, inherent variability in a population.

The aim of the contribution is twofold.

1. A non-parametric and non-intrusive method to estimate the uncertain parameters probability density function (PDF) by exploiting the observable variability is proposed.

2. We investigates a method to make this estimation "parsimonious", *i.e.* requiring as few model evaluations as possible and as few observables (or DOFs) as possible.

To tackle the first problem, two different strategies may be envisioned. First, one could estimate the model parameters associated with each experimental sample using classical inverse problem tools such as Bayesian approaches [KS06, BT11, SBN$^+$13, WZ04, Kou09] or genetic algorithms [HNGK09]. These strategies would yield a collection of parameters values from which the PDF would be computed by using histograms or more sophisticated PDF estimation techniques [AA13]. As straightforward as this approach is, it becomes computationally intensive

as the number of experimental samples grows larger. Second, one may see the experimental data set as a whole, which has the advantage of being both computationally cheaper and more robust to noise and low-quality measurements. We focus on the second strategy and present an adaptation of the well-known problem of moments [ST43]. The problem of moments matching has been used as an inverse problem tool with success in various contexts [GNSG11, SZ06, PS09]. A reasonable regularization of the problem of moments is the maximum entropy principle, (also referred to as Jaynes principle) which is rooted in information theory [Jay57b, MP84]. In most cases however, parameters of a model are not directly observable. Therefore, one needs a technique that takes into account the observable variability.

In this context, we introduce an "observable moment matching" method which consists in maximizing the PDF entropy under the constraints of matching the moments of the observable itself (not of the parameters). This is a two-step method. First, the model is evaluated for a fixed number of parameters samples and the corresponding outputs, i.e. the simulated observables, are stored. Second, the PDF is found by an iterative process that maximizes its entropy under the constraints of matching the moments of the experimental and simulated observables.

To address the second problem, we propose an algorithm that selects the DOFs in the physical domain where the moments are to be matched in order to alleviate the cost of the inverse problem – which is crucial for complex models such as PDEs – and to improve its conditioning. This algorithm exploits the sensitivity information provided by the pre-computed model evaluations. The sensitivity Gram matrix, computed for every DOF, reveals active subspaces [Con15, CELI15] of the parameter space. The DOFs are selected by clustering the active subspaces and choosing their best representatives. This strategy allows for a reduction of the number of DOFs by several orders of magnitude and therefore proves to drastically reduce the computational cost of the inverse problem without requiring any additional evaluation of the model.

We present, hereafter, the main steps of the method. Let us consider a data set that exhibits variability and a physical model assumed to accurately account for the observations. Let $\mathcal{D} \subseteq \mathbb{R}^d$ be an open subset, the physical domain (space, time or space-time), in which the governing equations are written.

Let $(\Theta, \mathcal{A}, \mathcal{P})$ be a complete probability space, Θ being the set of outcomes, $\mathcal{A}$ a σ–algebra and $\mathcal{P}$ a probability measure. The model can be written in a compact notation as:

$$\mathcal{L}(u(x, \vartheta)) = 0, \tag{3.4}$$

where $\mathcal{L}$ denotes a generic nonlinear differential operator.

The vector $\vartheta = \left(\theta_1, \ldots, \theta_{n_p}\right) \in \Theta$ denotes the uncertain parameters of the model and Θ is a bounded subset of $\mathbb{R}^{n_p}$, sometimes referred to as the stochastic domain [ZG08].

A set of measurements $\{y_1, \ldots, y_N\}$ is available. Each measurement y_i is assumed to take the following form:

$$y = g\left(u(x, \vartheta)\right) + \epsilon, \tag{3.5}$$

where g is a function describing the measurement process and ϵ is the noise, assumed to be additive and independent. For practical reasons, g is normalized to take values in $[0, 1]$. Let $\mathbb{E}$ be the expectation operator. We make the hypothesis that the random fields associated with the observables are p-integrable, that is: $\int_{\mathcal{D}} |\mathbb{E}(y^p)| \, dx < M$, where the exponent p is the highest

available moment. The variability in the observations is due to two main contributions: the variability in the parameters and the noise in the measurement process. In a classical forward Uncertainty Quantification (UQ) context, given the probability density function (PDF) of the parameters ρ, the moments of the observables are computed. In the present work, an inverse problem is solved which consists in finding the PDF of the parameters that generates the observed variability in a set of available data. Let us introduce the m^{th} order empirical moment of the measurements:

$$\mu_m(x) = \frac{1}{N}\sum_{i=1}^{N} y_i(x)^m \approx \mathbb{E}((g+\epsilon)^m), \tag{3.6}$$

and the m^{th} order moment of the simulations:

$$\mu_m^\rho(x) = \int_{\vartheta\in\Theta} (y_{sim}(\vartheta))^m \, \rho(\vartheta)\mathrm{d}\vartheta = \mathbb{E}(y_{sim}^m), \tag{3.7}$$

where y_{sim} are the observations of the simulated system. Under the assumption that the noise is additive, independent and with a known structure, it is straightforward to account for its influence on the measurements moments. Using the linearity of the expectation operator and the independence of the noise, it follows from definition (3.5) that:

$$\mathbb{E}\left[y^m\right] = \sum_{k=0}^{m} \binom{m}{k} \mathbb{E}\left[g^m\right] \mathbb{E}\left[\epsilon^{m-k}\right].$$

As an example, consider the case where the noise follows a zero-mean normal distribution with a known variance τ^2: $\epsilon \sim \mathcal{N}(0,\tau^2)$. Then, the following corrections may be applied to the first three empirical moments defined in Eq. (3.6):

$$\begin{aligned}
\tilde{\mu}_1(x) &= \mu_1(x),\\
\tilde{\mu}_2(x) &= \mu_2(x) - \tau^2,\\
\tilde{\mu}_3(x) &= \mu_3(x) - 3\tau^2\mu_1(x).
\end{aligned}$$

In the numerical experiments, the noise is assumed to be gaussian and its level is defined as the ratio $4\tau/A$ where A is the signal amplitude. In the Supplementary Material, the effect of τ^2 on the PDF estimation is investigated.

Only Gaussian noises are considered here. However, the same procedure may be applied to any noise whose power moments are known. If the noise structure is completely unknown, a strategy can be set up to estimate it but it is not investigated in the present work.

The overall algorithm aims at estimating the PDF ρ of the uncertain parameters $\vartheta \in \Theta \subseteq \mathbb{R}^p$, given some empirical moments of the observables.

The Jaynes principle of maximum entropy is applied (see [Jay57b]): the PDF is sought so that it has the maximum entropy under the constraints that the experimental and simulated moments be equal. Two additional constraints correspond to the positivity and the PDF normalization. This leads to the following optimization problem:

$$\begin{cases} \textbf{Minimize:} & \int_\Theta \rho \log(\rho) \\ \textbf{Subject to:} & \tilde{\mu}_m(x) - \mu_m^\rho(x) = 0, \quad \forall x \in \mathcal{D}, 1 \leq m \leq N_m, \\ & \rho(\vartheta) \geq 0, \qquad\qquad \forall \vartheta \in \Theta, \\ & \int_\Theta \rho = 1. \end{cases} \tag{3.8}$$

In what follows, this is referred to as the Observable Moment Matching (OMM) problem.

The classical problem of moments consists in finding a PDF ρ of the parameters θ_k from the knowledge of a finite number N_m of its power moments $\mu_{m,k}$, $m = 1, \ldots, N_m$, $k = 1, \ldots, n_p$:

$$\mathbb{E}_\rho \left[\theta_k^m\right] = \mu_{m,k}, \qquad m = 1, \ldots, N_m, \ k = 1, \ldots, n_p,$$

where $\mathbb{E}_\rho(\cdot)$ denotes the expectation operator given a density function ρ. This problem has been extensively discussed in the literature and has been addressed by adopting a wide range of strategies. When only a finite number of moments are known, which is often the case in practice, the problem becomes under-determined. Indeed, there exists an infinite number of densities that have the same N_m moments. Therefore, one needs to introduce a regularization in order to obtain a unique distribution function among all the feasible solutions.

This problem has been successfully used in situations where the moments of the model parameters are directly measurable, for instance in the context of microstructure reconstruction [GNSG11, SZ06, PS09]. In general however, the moments of the model parameters are not observable. Therefore, we propose to apply the moment matching constraints not on the parameters but on the observable itself.

To regularize the problem, the maximum entropy principle is used: find the PDF that maximizes the entropy under the constraint of matching the first N_m moments, where the differential entropy reads: $S(\rho) = -\int_\Theta \rho \log(\rho)$. There are three main reasons why this choice of regularization is well suited to the present case. First, from an information theory point of view, the maximum entropy PDF is considered the best choice when a limited amount of information is available (here, only a finite number of moments are known). This principle was first introduced by Jaynes [Jay57b] and was successfully applied to numerous practical cases [MP84, SZ06]. Second, $-S(\rho)$ is a convex cost function which enables the use of efficient optimization tools. Last, ρ can be written as an exponential term (see below), which dispenses the addition of an inequality constraint ensuring its positivity.

A set of constraint functions is introduced, expressing the mismatch between the moments of the measured observable and the moment of the simulated observable. They read:

$$c_m(x) = \mu_m^\rho(x) - \tilde{\mu}_m(x) = \int_\Theta g^m(x, \vartheta)\rho(\vartheta)\ d\vartheta - \tilde{\mu}_m(x), \quad m = 1, \ldots, N_m. \tag{3.9}$$

Introducing the Lagrange multipliers $\lambda(x) = (\lambda_m(x))_{m=1\ldots N_m}$, λ_0 and $\nu(\vartheta)$, the initial optimization problem (3.8) is recast in the following saddle-point problem:

$$\inf_\rho \sup_{\lambda, \lambda_0, \nu \geq 0} \mathcal{L}\left(\rho, \lambda, \lambda_0, \nu\right), \tag{3.10a}$$

with

$$\mathcal{L}\left(\rho, \lambda, \nu\right) = \int_\Theta \rho \log(\rho) - \sum_{m=1}^{N_m} \int_\mathcal{D} \lambda_m(x) c_m(x)\ dx - \lambda_0 \left(\int_\Theta \rho - 1 \right) - \int_\Theta \rho\nu. \tag{3.10b}$$

The necessary conditions for optimality are derived by cancelling out $\partial\mathcal{L}/\partial\rho$ and $\partial\mathcal{L}/\partial\lambda$:

$$\rho = \exp(\nu + \lambda_0 - 1) \exp\left(\sum_{m=1}^{N_m} \int_{\mathcal{D}} g^m \lambda_m \, dx\right), \tag{3.11}$$

$$\int_{\Theta} g^m \exp(\lambda_0 - 1) \exp\left(\sum_{h=1}^{N_m} \int_{\mathcal{D}} g^h \lambda_h \, dx\right) \, d\vartheta - \tilde{\mu}_m = 0, \; m = 1, \ldots, N_m. \tag{3.12}$$

In the present case, by virtue of the entropy regularization, the primal variable ρ can be expressed in an analytic form as a function of the dual variable and the positivity constraint is automatically satisfied (Eq.(3.11)). Hence, the solution of the system can be reduced to the solution of a nonlinear problem for the dual variable (Eq.(3.12)).

In the following, we suppose the problem is well-posed in the sense that the solution exists and is unique. It is shown in Appendix B of [GS13] that it is the case when the constraints are algebraically independent. The algebraic independence of the constraints, as formulated in [GS13], is equivalent to saying that there exists a non-zero measure subset $\tilde{\Theta}$ of Θ such that, for any nonzero vector $v^{(m)}$, $m = 1, \ldots, N_m$ in $L^2(\mathcal{D})$, one has:

$$\mathcal{A}_{\tilde{\Theta}} := \int_{\tilde{\Theta}} \left(\sum_{m=1}^{N_m} \left\langle v^{(m)}, g^m \right\rangle_{L^2(\mathcal{D})} \right)^2 d\vartheta > 0. \tag{3.13}$$

The optimality conditions reduce henceforth to a non-linear system of equations to be solved for the dual variables. The problem is discretised in Θ by means of a Quasi Montecarlo method and with standard discretisation methods in $\mathcal{D}$. The size of the optimisation problem to be solved is the number of degrees of freedom in the physical space (space-time) times the number of moments available: it can be, therefore, prohibitive. Moreover, it is often ill-conditioned. To overcome these drawbacks a reduction method is proposed, that aims at identifying the degrees of freedom in the physical space which are more pertinent in view of solving the problem. An algorithm is proposed, that selects a subset $\mathcal{S}$ of the full set of DOFs $\mathcal{D}$, based on a gradient algorithm which is rooted in the global sensitivity analysis of the model.

For each dof x_j, we consider a numerical approximation of the exact Sensitivity Gram Matrix (SGM):

$$C^j = \int_{\Theta} \left[\nabla_{\vartheta} g(x_j, \vartheta)\right] \left[\nabla_{\vartheta} g(x_j, \vartheta)\right]^T \rho \, d\vartheta,$$

This matrix encodes the local sensitivity covariance. For a given x_j the eigenvalues are denoted by $\eta_1^j, \ldots, \eta_{n_p}^j$, in descending order. The corresponding eigenvectors, denoted by $\mathbf{e}_1^j, \ldots, \mathbf{e}_{n_p}^j$, form an orthonormal basis of the parameter space. The vector $\mathbf{e}_1^j$ corresponds to the direction (in the parameter space) of maximum variation, on average, of g at x_j. Its associated eigenvalue η_1^j corresponds to the mean-squared directional derivative of the observable along the direction $\mathbf{e}_1^j$ [Con15][Lemma 3.1].

Each x_j is therefore associated with a dominant direction in the parameter space $\mathbf{e}_1^j$ and its corresponding eigenvalue η_1^j.

We are now able to address the initial problem: on the one hand, the DOFs where the variation of the observable is not significant are characterized by a low first eigenvalue. A threshold on η_1^j may be applied to remove the DOFs where the observable variation amplitude

is lower than the noise level. On the other hand, the DOFs that are redundant from the observable point of view are characterized by "similar" dominant directions. The similarity function between two (unit-norm) vectors is defined as the absolute value of the cosine of the angle between them.

Knowing this, we propose to divide the set of N_x dominant directions into N_k clusters using an agglomerative hierarchical clustering algorithm. In the present work, we used the Scikit-learn library [PVG+11] which provides a Python implementation of an agglomerative hierarchical algorithm that accepts user-defined similarity functions.

Once the N_x DOFs of the full physical set are divided into N_k clusters, the ones with maximum trace of $\hat{\mathbf{C}}^j$ are chosen as their cluster representatives. The output of the CS algorithm is a nested sequence of subsets $\mathcal{S}^{(1)} \subset \ldots \subset \mathcal{S}^{(N_x)}$ and we denote by N_k the cardinality of a given subset $\mathcal{S}$. The agglomerative clustering guarantees that the sequence of selected subsets is *nested*. This means that if $\mathcal{S}^{(n)}$ and $\mathcal{S}^{(n+1)}$ respectively count n and $n+1$ elements, then they have n elements in common. From a practical viewpoint, the full sequence of clusters can be computed once so that there is no additional cost linked to the clustering when N_k increases. Furthermore, in our simulations, we noticed that the residual had a smoother behavior as N_k increases compared to other clustering techniques. The detailed algorithm and some numerical experiments with systems of parametric PDEs are proposed in [19]. We report in Algorithm 3.2 the steps of the overall numerical strategy.

Algorithm 5 OMM algorithm

Initialise:

- Corrected experimental moments: $\tilde{\mu}_{j,m}$, $j = 1, \ldots, N_{\mathbf{x}}$, $m = 1, \ldots, N_m$
- Number of stochastic samples: N_c
- Tolerance parameters : α, ϵ_{Newton}

Step 1:

- Build the simulation set $\{g_{i,j}\}$;

Initial guess $\rho^{(0,0)}$: uniform distribution over Θ $\quad j = 1$, $N_k = 0$;
while $\|\mathbf{R}^{(j-1,N_k)}\|_2$ not converged **do**
    Apply CS on $\rho^{(j-1,0)}$ (**Step 2**) $\rightarrow$ nested sequence $\mathcal{S}^{(j,1)} \subset \ldots \subset \mathcal{S}^{(j,N_{\mathbf{x}})}$;
    $n = 1$;
    while $\|\mathbf{R}^{(j-1,n)}\|_2$ not converged **do**
        Apply OMM with $\mathcal{S}^{(j,n)}$ (**Step 3**) $\rightarrow \rho^{(j,n+1)}$;
        $n \leftarrow n + 1$;
    $j \leftarrow j + 1$;
    $N_k \leftarrow n$

3.3 Composite biomarkers: towards learning-simulation interaction

In numerous applications in biophysics, physiology and medicine, the system of interest is studied by monitoring a number of quantities, called biomarkers. For example in electrophysiology, an action potential can be viewed through simple quantities like the action potential duration, the amplitude, the rate of depolarization, *etc.* In hemodynamics, the systolic and diastolic pressures, or the pulse wave velocity, are typical biomarkers extracted from pressure measurements.

The biomarkers are obtained by applying a nonlinear map to the signal measured during experiments or clinical observations. They convey some information on hidden quantities, that are not directly measured. For example in hemodynamics, the pulse wave velocity can be linked to the arterial stiffness. When performing parameter estimation for a biophysical model, it is often much more convenient to work with biomarkers than with the whole signals. A natural question is therefore: which biomarker should be chosen to estimate a given parameter?

Biomarkers are usually proposed by the community, based on physical intuition and experimental observations. They are often relevant in qualitatively describing the hidden quantities. However, in most practical applications, although the biomarkers exhibit a good correlation with respect to the hidden quantity they are designed to monitor, they have a non-negligible correlation with respect to others, making them less robust or of difficult interpretation.

In this contribution, a strategy to automatically design biomarkers is proposed. The basic ideas of the approach are:

1. To design composite biomarkers that are maximally correlated with the hidden quantities they have to reveal, and minimally correlated with respect to all the others.

2. To provide a set of quantities making the parameter estimation better conditioned.

3. A greedy method has been recently developed for classification problems.

3.3.1 Numerical design of composite markers.

The biomarker design problem may be interpreted as a feature selection problem. Most of the literature considers the problem of selecting features in the input space in order to predict a given output (that may be the output of a computational model). Even though the aim of this work is reverse we will momentarily consider, for the sake of comparison, the biomarkers to be inputs and the parameters of interests to be the outputs. A common strategy to select a subset of the available features is by ranking or eliminating them according to a given criterion or score. This score may be based on a sensitivity analysis (e.g. first-order sensitivity indices [GB15]), based on information theory (e.g. Fisher information matrix [CABCL09]) or on the input covariance matrix [OMO14]. For other feature selection techniques and an overview of the field, the interested reader is referred to [GE03]. Other methods consist in selecting directions in the parameter space: instead of selecting a subset of features, linear combinations of the features are sought. Principle Component Analysis is a classical exemple in this respect. The same principle holds for functional-PCA [GLMG06], its counterpart applied to the case where the input space is a function space. Neither of these approaches take into account the relationship between

inputs and outputs. In Active Subspaces [Con15], directions in the input space are sought so that the gradient of the output with respect to these directions is maximum. In Partial Least Squares [WRWD84], directions are sought in the parameter space so that their covariance with the output is maximum. In this regard, this approach bears some similarities with the present work. Indeed, we look for biomarkers that are maximally correlated with their respective parameters. The main difference is that we also add the constraint that they are minimally correlated with all remaining parameters. Another way to reduce the input space is to perform a sparse linear regression using the Lasso algorithm [YGPD13]. The use of the ℓ^1 norm penalization makes this approach similar to the present work even though the cost function to be minimized is different.

Another aspect which makes the present approach different from previous works is that the feature, or biomarker, selection is performed in order to simplify future inverse problems. This issue has been addressed in [DDM17] but is rarely the focus of feature selection studies. For a comprehensive review of inverse problem techniques, the reader is referred to [KS06].

Our method is based on a semi-empirical approach. A mathematical model of the system of interest is considered and a database of simulations is built, by taking meaningful scenarios into account. Then, a dictionary of linear and nonlinear forms of the observable is considered. Such a dictionary is user-defined and therefore the efficiency of the proposed method relies on a relevant choice of its entries. Typically, such a dictionary is built by considering first common or "classical" biomarkers defined in the literature and then by incorporating agnostic features to enrich the space spanned by the dictionary entries. The composite biomarker is defined as a linear combination of the elements of the dictionary. The linear combination is sought such that the resulting biomarker is maximally correlated to the hidden quantity it refers to, and minimally correlated with respect to all the others.

From a practical point of view, at the computational cost of one single *offline* database computation (done once for all), the expansion coefficients of the biomarkers on a dictionary of observable forms are computed, for a given experimental setup or physical system. Then this result can be exploited for an unlimited number of experiments. As a by-product, when doing parameter identification, the ℓ^2-distance in the space of the biomarkers defines a metric which is the ℓ^2-distance in the space of the hidden quantities, up to a controlled perturbation. This makes the inverse problem less ill-conditioned and, in general, easier to solve.

The method is detailed.

Let $(\Theta, \mathcal{A}, \mathcal{P})$ be a complete probability space, Θ being the set of outcomes, $\mathcal{A}$ a σ–algebra and $\mathcal{P}$ a probability measure. Let $\mathbb{E}[\cdot]$ denote the expectation operator. In the following, random variables are denoted by boldface letters.

We assume we have an ODE or PDE model, depending on $p \in \mathbb{N}^*$ parameters $\boldsymbol{\vartheta}_1, \dots, \boldsymbol{\vartheta}_p$, which, for the sake of simplicity, are assumed to be zero-mean unit-variance and mutually uncorrelated (This is an hypothesis on the prior of the parameters). The observable of the model is defined as a function v from $\mathbb{R}^p$ to $\mathbb{R}^M$ which corresponds to the output of the parameterized equations. The observable can be viewed as a random vector $\boldsymbol{v}$ of the following form:

$$\boldsymbol{v} : \left\{ \begin{array}{ccc} (\Theta, \mathcal{A}, \mathcal{P}) & \longrightarrow & \mathbb{R}^M \\ (\boldsymbol{\vartheta}_1, \dots, \boldsymbol{\vartheta}_p) & \longmapsto & v\left(\boldsymbol{\vartheta}_1, \dots, \boldsymbol{\vartheta}_p\right). \end{array} \right. \tag{3.14}$$

A biomarker g is a function from the observable, which is an element of $\mathbb{R}^M$ to $\mathbb{R}$, and it can be

viewed in the present context as a random variable $\boldsymbol{g}$ defined as follows:

$$\boldsymbol{g}: \left\{ \begin{array}{ccc} (\Theta, \mathcal{A}, \mathcal{P}) & \longrightarrow & \mathbb{R} \\ (\boldsymbol{\vartheta}_1, \ldots, \boldsymbol{\vartheta}_p) & \longmapsto & g\left(v\left(\boldsymbol{\vartheta}_1, \ldots, \boldsymbol{\vartheta}_p\right)\right). \end{array} \right. \tag{3.15}$$

A set of $n_g \in \mathbb{N}^*$ functions g can be actually computed given a number of $n_s \in \mathbb{N}^*$ instances of the observables, for different parameters samples. This is a dictionary G of instances. The marker meant to reveal the parameter ϑ_j is denoted by $y^{(j)}$ and, by hypothesis, it is sought as a linear combination of the dictionary entries:

$$y^{(j)} = \sum_{i=1}^{n_g} w_i^{(j)} g^{(j)}.$$

The goal is to compute the weights $w_i^{(j)} \in \mathbb{R}$ in such a way that $y^{(j)}$ is maximally correlated to ϑ_j and minimally correlated to all the $\vartheta_i,\ i \neq j$.

Let the dictionary entries be samples for the centered random variable $\tilde{\mathbf{g}} = \mathbf{g} - \mathbb{E}[\mathbf{g}]$. The following quantities are introduces, which are useful to define the problem:

$$C_{kl} := \mathbb{E}[\tilde{\mathbf{g}}_k \boldsymbol{\vartheta}_l], \quad 1 \leq k \leq n_g,\ 1 \leq l \leq p, \tag{3.16}$$

$$V_{kl} := \mathbb{E}[\tilde{\mathbf{g}}_k \tilde{\mathbf{g}}_l], \quad 1 \leq k \leq n_g, 1 \leq l \leq n_g, \tag{3.17}$$

$$\nu(w^j) = (w^{(j)})^T V w^{(j)}, \quad 1 \leq j \leq p, \tag{3.18}$$

$$e_l^{(j)} := \delta_{lj}, \quad 1 \leq j \leq p,\ 1 \leq l \leq p. \tag{3.19}$$

Let $\lambda^{(j)} \in \mathbb{R}^+$ be a penalisation parameter. The optimisation problem to be solved reads:

$$\boxed{w_*^{(j)} = \arg \min_{w^{(j)} \in \mathbb{R}^{n_g}} \mathcal{J}_\lambda(w^{(j)})}, \tag{3.20}$$

$$\mathcal{J}_\lambda(w^{(j)}) = \mathcal{J}(w^{(j)}) + \frac{\lambda^{(j)}}{n_g} \|w^{(j)}\|_{\ell^1(\mathbb{R}^{n_g})}, \tag{3.21}$$

$$\mathcal{J}(w^{(j)}) = \frac{1}{2} \|C^T w^{(j)} - e^{(j)}\|^2_{\ell^2(\mathbb{R}^p)} + \frac{\xi}{2} \left(\nu\left(w^{(j)}\right) - 1\right)^2. \tag{3.22}$$

A proposition is reported, which is detailed in [18], and that helps in understanding the properties of the solutions of the problem.

7. *Proposition.* *Let $\lambda^{(j)} = 0$. The composite marker $\mathbf{y}^{(j)}$ can be decomposed as:*

$$\mathbf{y}^{(j)} = \boldsymbol{\vartheta}^T (e^{(j)} + \gamma) + \mathbf{q}, \tag{3.23}$$

where $\mathbf{q} : (\Theta, \mathcal{A}, \mathcal{P}) \to \mathbb{R}$. Moreover, it exists a constant $K > 0$ such that:

$$\|\boldsymbol{\vartheta}^T \gamma\|_{L^2(\Theta)} \leq \sqrt{2} (\mathcal{J}_{\lambda=0})^{1/2}, \tag{3.24}$$

$$\|\mathbf{q}\|_{L^2(\Theta)} \leq K (\mathcal{J}_{\lambda=0})^{1/4}. \tag{3.25}$$

The result of the above Proposition is meaningful for two reasons:

1. It states that the value of the functional can be used as an *a posteriori* indicator of the quality of the dictionary chosen.

2. If the solution $w_*^{(j)}$ is such that the value of the functional $\mathcal{J}$ is small, observing $y^{(j)}$ is equivalent, up to a small perturbation, to observing $\vartheta^{(j)}$. The composite marker can be used as a non-linear preconditioner for assimilation problems.

We report in the Algorithm 3.3.1 the main steps of the method, and refer to [18] for all the details of the method and the discretisation.

Algorithm 6 Composite Biomarkers

Initialise:

- Sparsity promoting parameter: $\lambda^{(j)}$
- Initial guess: $w^{(j),0}$
- Stopping criteria: absTol and relTol
- Constraint tolerance: ϵ_c
- Constraint penalization parameter (initial guess): ξ
- Constraint penalization parameter update factor: $\alpha > 1$

$i \leftarrow 0$
Compute residual: $r^{(0)} := J_\lambda(w^{(j),0})$
while $r^{(i)} <$ absTol **or** $\left|r^{(i)} - r^{(i-1)}\right| <$ relTol $\times\, r^{(i)}$ **do**
  if $\left|\nu(w^{(j),i}) - 1\right| > \epsilon_c$ **then**
    Increase penalisation: $\xi \leftarrow \alpha\xi$
  Compute gradient: $\partial_{w_\pm^{(j)}} J_\lambda(w^{(j),i})$
  Nesterov descent: $w^{(j),i+1} := f(w^{(j),i}, \partial_{w^{(j)}} J_\lambda(w^{(j),i})$
  Update residual: $r^{(i+1)} := J_\lambda(w^{(j),i+1})$
  $i \leftarrow i + 1$

The ℓ^1 penalty constraint is dealt with by means of a classical approach, consisting in doubling the number of the unknowns. The penalisation of the variance ν is treated with an increasing penalty (continuation), and the parameter λ is fixed by using an $L-2$ curve criterion. A Nesterov accelerated projected gradient is used to perform the computation. Several numerical experiments are described in [18, 30].

3.3.2 A greedy algorithm for classification problems.

This contribution is motivated by the problems in saphety pharmacology, but the result of the investigation is a rather general, and it can be applied to a broad spectrum of problems in engineering. The starting point is the same as for the construnction of composit markers, namely, a dictionary of quantities extracted from the observables.

Let G be an ensemble of signals, provided from experimental measurements, numerical simulations (or both). Let $n_s \in \mathbb{N}^*$ be the number of samples that will be used to train the classifier:

for each $G^{(i)},\ i = 1, \dots, n_s$ a set of $n_g \in \mathbb{N}^*$ quantities are extracted from the signal. These can be either informed linear or non-linear forms identified by experimental insight or more agnostic features, such as point values of the signal, local average, Fourier or Wavelets coefficients. We refer to the set of these quantities for all the available signals as the dictionary entries $G^{(i)}_j \in \mathbb{R},\ i = 1, \dots, n_s,\ j = 1, \dots, n_g$. The present work deals with classification problems: given an observable signal coming from a physical system, we want to determine to which class in a set of possible classes the system belongs to. For sake of simplicity, the method is derived in the case of binary classification: its extension to multiple classes is straightforward. As it frequently happens science and engineering, the size n_g of quantities that can be extracted from the signal can be extremely large. Moreover, the number of available samples n_s, due to experimental constraints and to the complexity of the systems at hand, can be small if compared to n_g. This regime, called *high dimensional/low sample size* in the learning community, is particularly critic when performing classification and regression tasks. The mathematical reason is that we wish to identify a function whose domain dimension n_g is large, and hence we are possibly exposed to the phenomenon of the *curse of dimensionality*, introduced for the first time by Bellman in [Bel15] and related to learning theory in [SZ03]. In [CDD$^+$12, FSV12, MUV15] a theoretical analysis is proposed that describes the ability of approximating a high-dimensional ridge function by point queries. These analyses shows how the problem in some regimes admits approximations that are not exposed to the curse of dimensionality. From a probabilistic viewpoint, for a given sample size, when the dictionary size becomes too large, the classification error increases: this is referred to as the Hughes phenomenon [Hug68, Tru79].

The method is described hereafter. Let $X_{n_g} \in \mathbb{R}^{n_g}$ be a random vector of the probability space $(\Omega, \mathcal{A}, \mathbb{P})$. We assume that the probability density function (pdf) of X_{n_g} is a mixture of the form:

$$\rho(g) = \rho_0(g)\pi_0 + \rho_1(g)\pi_1, \tag{3.26}$$

where $\rho_i(g) = \rho_i(g|y_* = i)$, the conditional probability density of g given that its label is $y_* = i$. The scalars π_i are the weights of the mixture and they can be seen as the *a priori* probability mass of being in the class i. It holds $\pi_0 + \pi_1 = 1$.

A classifier is defined in Definition 1 with $n = n_g$.

1. *Definition.* *Let $g \in \mathbb{R}^n$ be an observation coupled with a label y. A binary classifier is a function $\mathcal{C}_n$ such that the following holds:*

$$\mathcal{C}_n : \mathbb{R}^n \to \{0, 1\}, \tag{3.27}$$

$$g \mapsto y. \tag{3.28}$$

Some geometrical notations are introduced. Let $k \leq n_g$. The Grassmann manifold Gr_{k,n_g} is the set of $k-$dimensional linear subspace of $\mathbb{R}^{n_g}$. The method proposed in the present work can be seen as an optimisation on the compact Stiefel manifold, denoted by $\mathcal{M}_{k,n_g}$, whose definition is recalled in Definition 2. An element of the Stiefel manifold will be denoted by M.

2. *Definition.* *A Stiefel manifold $\mathcal{M}_{k,n_g}$ is a set of all the k-frames in $\mathbb{R}^{n_g}$:*

$$\mathcal{M}_{k,n_g} \triangleq \left\{ Y = (Y_1, \dots, Y_k), Y_i \in \mathbb{R}^{n_g} | Y_i^T Y_j = \delta_{ij}, \forall 1 \leq i, j \leq k \right\}, \tag{3.29}$$

so that the elements of the compact Stiefel manifold are the $k \times n_g$ matrices with orthonormal columns. The Stiefel manifold $\mathcal{M}_{n_g,n_g} = O(n_g)$ is the orthogonal group. An element $R \in O(n_g)$ satisfies $R^T R = RR^T = I_{n_g}$. It can be seen, roughly speaking, as the concatenation of an element of the Stiefel manifold, and its orthogonal complement: $R = [M, M^\perp]$. Let us consider the endomorphism induced by R, and how the probability density ρ is transformed accordingly. A change of coordinates is applied to the expression in Equation 3.26, leading to:

$$\rho(\xi) = \rho_0(R\xi)\pi_0 + \rho_1(R\xi)\pi_1, \tag{3.30}$$

that holds since $\det(R) = 1$.

An element M of a Stiefel manifold is used to reduce the input dimension: $x \in \mathbb{R}^k$ (the dimension k is, also, an outcome of the proposed method).

Let $\mathcal{C}_k$ be a classifier in the projected space of dimension k (Definition 1 with $n = k$). It is defined as:

3. *Definition.* *The classifier $\mathcal{C}_k$ in the subspace of dimension $k \ll n_g$ is defined as follow:*

$$\begin{array}{c} \mathcal{C}_k : \mathbb{R}^k \longrightarrow \{0,1\} \\ x = M^T g \mapsto y \end{array}, \tag{3.31}$$

where $g \in \mathbb{R}^{n_g}$ is an observation, $M \in \mathcal{M}_{k,n_g}$ and y is the label in the projected space.

The objective is to find $M \in \mathcal{M}_{k,n_g}$ which maximizes the success rate of the classifier $\mathcal{C}_k$. This has to be made more precise. In particular, an objective function is introduced, related to the classification success rate. This function could, in general, depend upon the classifier (defined by the function $\mathcal{C}$); in what follows we will propose an objective function that is intrinsically related to the ability of distinguishing between two classes, and that can be applied to all types of classifiers. By the properties of orthogonality of the Stiefel manifold and its orthogonal complement, the pdf p in the projected space of dimension $k < n_g$ corresponds to the marginals of ρ (see Equation 3.32). Indeed, let $M \in \mathcal{M}_{k,n_g}$ and $R = [M, M^\perp]$. Let an input $x = M^T g$; we denote by $\xi \in \mathbb{R}^{n_g}$ the vector $\xi = R^T g$. It follows that $x = [\xi_1; \ldots; \xi_k]$. Since R is an element of the orthogonal group, it holds:

$$p(x) = \int_{\mathbb{R}^{n_g - k}} \rho(\xi) d\xi_{k+1} \ldots d\xi_{n_g}, \tag{3.32}$$

and hence:

$$p(x) = p_0(x)\pi_0 + p_1(x)\pi_1. \tag{3.33}$$

An important consequence is that p is a mixture of the same form as ρ, and, moreover, $p_i(x)$ is the conditional probability:

$$p_i(x) = p(x|y_* = i), \tag{3.34}$$

for $i = 0$ or 1. The input space is subdivided into three distinct regions, in relation to what the classifier $\mathcal{C}_k$ (see 3) would provide, based on a probability argument. We denote by $S_0 \subseteq \mathbb{R}^k$, $S_1 \subseteq \mathbb{R}^k$ and $S_2 \subseteq \mathbb{R}^k$:

4. *Definition.*

$$\begin{cases} S_0 \triangleq \{x = M^T g \in \mathbb{R}^k | \pi_0 p_0(x) > \pi_1 p_1(x)\} \\ S_1 \triangleq \{x = M^T g \in \mathbb{R}^k | \pi_1 p_1(x) < \pi_0 p_0(x)\} \\ S_2 \triangleq \{x = M^T g \in \mathbb{R}^k | \pi_0 p_0(x) = \pi_1 p_1(x)\} \end{cases} . \tag{3.35}$$

It follows that:

- $S_i \cap S_j = \emptyset, \forall i \neq j$.
- $\cup_{i=0}^{2} S_i = S \subseteq \mathbb{R}^k$.

Let (g, y_*) be a couple such that $g \in \mathbb{R}^{n_g}$ is an observation and $y_* \in \{0, 1\}$ the corresponding label (the true label). Let A_S be the ensemble of the success events, *i.e.* when the classifier $\mathcal{C}_k$ provides as result $y = y_*$. The set of success events can be defined as:

5. *Definition.*

$$\begin{cases} A_{S_0} \triangleq \{y_* = 0 \wedge \pi_0 p_0 > \pi_1 p_1\} \\ A_{S_1} \triangleq \{y_* = 1 \wedge \pi_1 p_1 > \pi_0 p_0\} \\ A_{S_2} \triangleq \{y_* = 0, 1 \wedge \pi_0 p_0 = \pi_1 p_1\} \end{cases} , \tag{3.36}$$

And,

$$A_S \triangleq \cup_{i=0}^{2} A_{S_i}. \tag{3.37}$$

From the success events defined in Definition 5, the measure of the success events $\mu(A_S)$ can be obtained by quantifying the measure of the set:

$$\mu(A_S) = \int_{S_0} \pi_0 p_0(x) dx + \int_{S_1} \pi_1 p_1(x) dx + \frac{1}{2} \int_{S_2} p(x) dx, \tag{3.38}$$

The $\frac{1}{2}$ factor is justified by the fact that we expect to have half of the realizations to be well classified on S_2. This score is analogous to the excess risk measure proposed in [BCD$^+$14]. In [26], this score is related to the most commonly used divergences and metrics. The results are summarised in the following Propositions. The equivalence of the score with the total variation δ is proved in:

8. *Proposition.* *Let $p(x)$ be defined as in Eq.*(3.33)-(3.34)*, and the quantity $\mu(A_S)$ be defined as in Eq.*(3.38)*. It holds:*

$$\frac{1}{2} + \min(\pi_0, \pi_1) \delta_{TV}(p_0, p_1) \leq \mu(A_S) \leq \frac{1}{2} + \max(\pi_0, \pi_1) \delta_{TV}(p_0, p_1). \tag{3.39}$$

The Pinsker inequality makes it possible to prove a similar equivalence for the Hellinger distance d_H:

9. *Proposition.* *Let P_0 and P_1 be two probability distributions on S and p_0 and p_1 the corresponding pdf. Then,*

$$\frac{1}{2} + \min(\pi_0, \pi_1) d_H^2(P_0, P_1) \leq \mu(A_S) \leq \frac{1}{2} + \max(\pi_0, \pi_1) \sqrt{2} d_H(P_0, P_1). \tag{3.40}$$

Particularly relevant is the symmetrised Kullbach-Leibler divergence D_{KL}, for which we can prove:

10. ***Proposition.*** *Let P_0 and P_1 be two probability distributions on S (see Definition of the set S in 4) with pdf p_0 and p_1.*

If $\log\left(\frac{p_0}{p_1}\right) \in L^\infty(S)$ and, moreover, $D_{KL}(p_i||p_j) < +\inf$ for $i \neq j, i, j = 0$ or 1 then, $\exists c > 0$ such that:

$$2\left(\frac{\mu(A_S) - \frac{1}{2}}{\max(\pi_0, \pi_1)}\right)^2 \leq D_{SKL}(P_0, P_1) \leq c\frac{\mu(A_S) - \frac{1}{2}}{\min(\pi_0, \pi_1)}. \tag{3.41}$$

The goal is to maximise the score $\mu(A_S)$. Optimising over all the possible elements of the Stiefel manifolds (of multiple and unknown dimension k) would be prohibitive. To circumvent this, a double greedy approach is proposed. The heuristics are the following: the smaller the dimension of the input, the better it is in terms of palliating the curse of dimensionality; aiming at reducing possible overfitting phenomena, the sparser the orthonormal vectors of M, the better it is. Henceforth, the strategy which is investigated is the following: we start with $k = 1$ and look for a vector of unitary norm such that at each step of a greedy method, we maximise $\mu(A_S)$. When the error on a validation set stagnates and start increasing (early stopping criterion [Pre98]), we start considering $k = 2$. The first column vector of M is the result of the previous step of the method, and by a greedy approach we construct a second unitary norm column vector, orthogonal to the first one. This can be iterated until the error on a validation set starts increasing as soon as we start building the $(k+1)-$th vector.

Let $n_s, n_v \in \mathbb{N}^*$ be the number of the samples used in the training and the validation phases respectively. A training and a validation datasets $(g^{(i)}, y_*^{(i)})_{i=1}^{n_s}$, $(g^{(i)}, y_*^{(i)})_{i=1}^{n_v}$ are given, that consist of couples of dictionary entries and corresponding labels.

Let $\widehat{M}_{k,n_g} \in \mathcal{M}_{k,n_g}$ be the element of the Stiefel Manifold selected at the $k-$th outer iteration of the method. The goal is to find a vector $\omega_* \in \mathbb{R}^{n_g}$, orthogonal to all the columns of the matrix $\hat{M}_{k,n_g}$, such that:

$$\begin{aligned} \widehat{M}_{k+1,n_g} &= [\widehat{M}_{k,n_g}, \omega_*], \\ x \in \mathbb{R}^{k+1},\ x &= \widehat{M}_{k+1,n_g}^T g, \\ \omega_* &= \arg\inf_{\omega \in \mathbb{R}^{n_g}} \mu(A_S). \end{aligned} \tag{3.42}$$

When n_g is large, this optimisation can be costly. Furthermore, when the vector ω is sparse the classification tends to be less prone to overfitting phenomena. For these reasons, ω is constructed in a greedy way. At first $\|\omega\|_{\ell^{0,n_g}} = 1$, so that only one dictionary entry is chosen, by computing the value of the score (on the training dataset) for all possible choices and keeping the best. The numerical approximation of the cost $\mu(A_s)$ is detailed in [26]. The double greedy algorithm is detailed hereafter.

At the beginning of the $l-$th inner iteration, $\|\omega\|_{\ell^{0,n_g}} = l - 1$, $l - 1$ dictionary entries have been chosen and we have to choose the $l-$th one. Let the chosen indices be in the set $c^{(k+1)} = \{i_1, \ldots, i_{l-1}\}$. The $l-$th non zero entry has to be chosen among the indices $i \in c_c^{(k+1)}$,

the complementary set of $c^{(k+1)}$. Moreover, the best values of the selected entries of ω are sought, such that the result of the classification is the best possible (in the sense of the score introduced). Once one candidate to be the $l-$th non zero component is proposed, an optimisation task on the entries of ω is performed by using the CMAES method, detailed in [KMH$^+$04]. This does not guarantee automatically that ω is orthogonal to the subspace spanned by the column of $\widehat{M}_{k,n_g}$. Otherwise stated, $[\widehat{M}_{k,n_g}, \omega] \in Gr_{k+1,n_g}$. The projection onto the Stiefel manifold is obtained by QR decomposition. Let $Q_m \in \mathbb{R}^{n_g \times k+1}$, $R_m \in \mathbb{R}^{k+1 \times k+1}$, it holds:

$$Q_m R_m = [\widehat{M}_{k,n_g}, \omega], \tag{3.43}$$

$$\widehat{M}_{k+1,n_g} = Q_m. \tag{3.44}$$

Among all the possible optimised choices for the $l-$th component, the one that maximises the score is chosen. As said, the stopping criterion for these iterations is the early stopping strategy [Pre98]: the score is computed on the validation set. A stagnation of the score ends the inner iteration. As soon as increasing the dimension of the Stiefel manifold does not produce an improvement on the score computed on the validation, the outer iterations end. Once the algorithm terminates, the element of the Stiefel manifiold is obtained.

The algorithm is detailed in [26]. A principle of theoretical analysis is proposed, in which it is shown that the algorithm, when it performs an update, always improves the score.

11. *Proposition.* *Let $\widehat{M}_{k,n_g} \in \mathcal{M}_{k,n_g}$ and $\widehat{M}_{k,n_g} = [\widehat{M}_{k-1,n_g}, \omega]$; let $1 \leq m < n_g$, and $\|\omega\|_{\ell^{0,n_g}} = m$. The set of non-zero entries of ω is denoted by c, whose cardinality is $\#c = m$. Let $\tilde{\omega} \in \mathbb{R}^{n_g}$. The set of non-zero entries of $\tilde{\omega}$ is $\tilde{c}$, $\#c = m+1$. It holds $c \subset \tilde{c}$. Then,*

$$\max_{\tilde{\omega}} \mu(A_S) \geq \max_{\omega} \mu(A_S). \tag{3.45}$$

12. *Proposition.* *Let $M_{k,n_g} \in \mathcal{M}_{k,n_g}$ and the associated score be $\mu(A_S^{(k)})$. Let $M_{k+1,n_g} \in \mathcal{M}_{k+1,n_g}$ such that:*

$$M_{k+1,n_g} = [M_{k,n_g}, \omega],$$

where $\omega \in \mathbb{R}^{n_g}$ and the associated score be $\mu(A_S^{(k+1)})$. Then:

$$\mu(A_S^{(k+1)}) \geq \mu(A_S^{(k)}). \tag{3.46}$$

Several comparisons to other dimension reduction methods applied to classification are proposed therein.

3.4 Conclusions.

In this section of the manuscript, several contributions were proposed to address some of the methodological issues encountered in solving realistic data assimilation problems.

The first fundamental problem is related to identifiability, whose estimation is of the utmost importance but remains out of reach for complex systems. The set up of a numerical method to estimate mutual information as well as tensor methods to approximate the solution of large many-parameters systems of PDEs would mitigate the difficulties of this estimation.

Second, the problem of estimating the probability density distributions of a model parameter set by exploiting data collected in a population (or a set of experiments) has been investigated. The main challenges related to this problem are the computational cost and the need for a regularisation, that, for the present case, was chosen as the differential entropy.

The last problem presented, the introduction of composite markers, made it possible to address several realistic applications. Moreover, it highlights an eventual interaction between the learning strategies (currently intensively developed) and the mathematical modeling. The preliminary investigations as well as the numerical experiments on realistic data sets are, for sure, insufficient to have a neat definite picture of the problem. However, the results suggest that for most of the applications (at least the ones in biomedical engineering), the regime of the data is such that integrating the *a posteriori* knowledge coming from available datasets with *a priori* knowledge coming from mathematical modeling is unavoidable to get good predictions. This leads to the necessity of describing the problem from a mathematical point of view and defining systematically the possible methods to couple these different knowledge sources. In particular, given an available experimental dataset and several possible models, how an *in silico* models population can be constructed to maximise the reliability of the sough predictions? Moreover, what is the best way to define the method that actually computes the prediction?

Conclusions

In this document, a brief synthetic review of my research activity has been proposed. Some general conclusions are presented hereafter as well as some perspectives.

Numerous applications in biomedical engineering could potentially take advantage of numerical methods. These are a promising systematic way to improve the currently used techniques. Data assimilation is one of the main thematic areas of investigation in this respect. Several cases have been presented, in different systems, with different kinds of data and wished predictions. In tumour growth the data are images and the goal is to produce a forecast in time. In safety pharmacology, data are electrograms and the prediction sought might be a set of binary *yes/no* answers. Variational and sequential approaches have been tested. Their pertinence and efficiency largely depend upon the problem nature to be treated. In each of the presented experiments, encouraging results were obtained, showing that the applications targeted are non unfeasible and could be helpful in gaining insight. Despite the heterogeneity in the nature of the systems, of the data and of the sought predictions, all the problems share a common structure. In particular, they highlight several issues preventing the current models and methods to be directly deployed in realistic applications. These issues can be roughly divided into two classes: the often prohibitive computational cost needed to perform the estimations and the ability to account for the uncertainty.

The perspective in data assimilation consists in integrating the results of the methodological investigations. In particular the focus will be on two different applications:

1. The sequential assimilation for the 3D reconstruction of blood flow from ultrasound and MRI images.

2. The inverse problems in safety pharmacology, to detect dangerous molecules from multiple sources of information.

Concerning high-dimensional problems and reduced order models, four different topics were presented in this manuscript. They refer to different challenges encountered when solving the problems presented in the Data Assimilation chapter of the present manuscript.

The first set of works is about the reduction of fluid-structure interaction and multi-physics systems. In particular, we focused on cases in which the physics is such that the equations can be simplified in a way that helps in reducing the computational cost. The reduction of a thin structure to a Robin generalised boundary condition for the fluid problem was considered. The second work was about the reduction by means of an online adapted reduced basis

of the Poincaré-Steklov operator for multi-physics multi-compartment models. The second contribution is on optimal transport. In this, some effort was done on the numerical methods to approximate the solution of the optimal transport in Benamou-Brenier and Kantorovich formulations. A possible way of dealing with unbalanced optimal transport was also considered. A third thematic is the reduction of advection-dominated problems. These are a challenge to the classical projection based ROMs. Two possible approaches were considered: a dynamical basis method derived as a numerical analogue of the Lax-Pairs and reduction with respect to different metrics. The last set of works deals with tensor methods. Some ways to construct dynamical adaptive parsimonious discretisations for high-dimensional PDEs were presented.

The perspectives in model reduction are the following:

1. Investigate the reduction methods for advection-dominated problems. In particular, try to understand whether some metric spaces are better suited to perform fast and accurate approximations of solutions featured by traveling and progressive waves.

2. Study and develop tensor methods for the efficient solution of high-dimensional PDEs. This could be of particular relevance to data-assimilation and uncertainty quantification.

In uncertainty quantification, briefly exposed in the last chapter, three main contributions were presented. In the first one, a pragmatical characterisation of the identifiability has been proposed, based on information theoretic quantities. This fits naturally into a Bayesian framework. In particular, it makes it possible to spot situations in which some of the parameters are poorly identifiable or compensate with others. A modification of the KL entropy estimator has been proposed, that better performs in the cases of interest. The second contribution deals with a population stochastic inverse problem: the probability density distribution of the model parameters is estimated by means of a non-parametric approach in order for the model outputs statistics to match the observed ones, at best. This have numerous applications and it can be used to improve the priors to be used in Bayesian inverse problems. The last work presented consists in the definition of a semi-empirical method to design composite markers. From the point of view of the application, these can be seen as a corrected or improved version of the biomarkers currently used in the scientific community to relate measurements to some QoIs. From a numerical point of view, composite markers can be interpreted as non-linear preconditioners for assimilation problems. A greedy algorithm have been proposed to deal with classification problems. This is an example of learning-simulation interaction.

The perspectives of these works are the following:

1. Define a method that improves KSG mutual information estimator for the cases in which the joint density distribution is concentrated around a small-dimensional manifold embedded into the joint space.

2. Investigate the learning-simulation interaction; in particular, define a systematic way to exploit, at best, a combination of real data and mathematical modeling.

The division in three distinct parts of these topics is clearly more for the sake of presentation than for a real thematic separation. Indeed, several connections exist between the different

points presented, not only between data assimilation and the methodological contributions, but also between the different methodological contributions presented. For instance, setting up parsimonious discretisations is a technical need to foster the development of the methods proposed in the last chapter of this manuscript. The main perspective of the works described in this manuscript consists therefore in analysing and exploring all these connections.

References.

[AA13] Aravind Alwan and Narayan R Aluru. Improved statistical models for limited datasets in uncertainty quantification using stochastic collocation. *Journal of Computational Physics*, 255:521–539, 2013.

[AAC16] R. Abgrall, D. Amsallem, and R. Crisovan. Robust model reduction by L1-norm minimization and approximation via dictionaries: application to nonlinear hyperbolic problems. *Advanced Modeling and Simulation in Engineering Sciences*, 3(1):1, 2016.

[ACF14] Luigi Ambrosio, Maria Colombo, and Alessio Figalli. On the lagrangian structure of transport equations: the Vlasov-Poisson system. *arXiv preprint arXiv:1412.3608*, 2014.

[AG13] Luigi Ambrosio and Nicola Gigli. A user's guide to optimal transport. In *Modelling and optimisation of flows on networks*, pages 1–155. Springer, 2013.

[AGPT09] M. Astorino, J-F. Gerbeau, O. Pantz, and K. Traoré. Fluid-structure interaction and multi-body contact. Application to aortic valves. *Comp. Meth. Appl. Mech. Engng.*, 198:3603–3612, 2009.

[AH16] D. Amsallem and B. Haasdonk. Pebl-rom: Projection-error based local reduced-order models. *Advanced Modeling and Simulation in Engineering Sciences*, 3(1):6, 2016.

[AH17] Babak Maboudi Afkham and Jan S Hesthaven. Structure preserving model reduction of parametric hamiltonian systems. *SIAM Journal on Scientific Computing*, 39(6):A2616–A2644, 2017.

[AL85] V.I. Agoshkov and V.I. Lebedev. Poincaré-steklov's operators and the methods of partition of the domain in variational problems. *Computational Processes and Systems*, 2:173–227, 1985.

[Ars75] Aleksei Alekseevich Arsenev. Existence in the large of a weak solution to the Vlasov system of equations. *Zhurnal Vychislitelnoi Matematiki i Matematicheskoi Fiziki*, 15:136–147, 1975.

[AZF12] D. Amsallem, M. J. Zahr, and C. Farhat. Nonlinear model order reduction based on local reduced-order bases. *International Journal for Numerical Methods in Engineering*, 92(10):891–916, 2012.

[AZW15] David Amsallem, Matthew J Zahr, and Kyle Washabaugh. Fast local reduced basis updates for the efficient reduction of nonlinear systems with hyper-reduction. *Advances in Computational Mathematics*, 41(5):1187–1230, 2015.

[BÅ70] R. Bellman and K.J. Åström. On structural identifiability. *Mathematical biosciences*, 7(3-4):329–339, 1970.

[BB00] Jean-David Benamou and Yann Brenier. A computational fluid mechanics solution to the monge-kantorovich mass transfer problem. *Numerische Mathematik*, 84(3):375–393, 2000.

[BBOVA+13] Oliver J Britton, Alfonso Bueno-Orovio, Karel Van Ammel, Hua Rong Lu, Rob Towart, David J Gallacher, and Blanca Rodriguez. Experimentally calibrated population of models predicts and explains intersubject variability in cardiac cellular electrophysiology. *Proceedings of the National Academy of Sciences*, 110(23):E2098–E2105, 2013.

[BCD+14] Peter Binev, Albert Cohen, Wolfgang Dahmen, Ronald DeVore, et al. Classification algorithms using adaptive partitioning. *The Annals of Statistics*, 42(6):2141–2163, 2014.

[BČG+14] M Bukač, S Čanić, Roland Glowinski, Boris Muha, and Annalisa Quaini. A modular, operator-splitting scheme for fluid–structure interaction problems with thick structures. *International journal for numerical methods in fluids*, 74(8):577–604, 2014.

[BCOW17] Peter Benner, Albert Cohen, Mario Ohlberger, and Karen Willcox. *Model reduction and approximation: theory and algorithms*, volume 15. SIAM, 2017.

[BCZH06] Y Bazilevs, VM Calo, Y Zhang, and Thomas JR Hughes. Isogeometric fluid–structure interaction analysis with applications to arterial blood flow. *Computational Mechanics*, 38(4-5):310–322, 2006.

[BD85] Claude Bardos and Pierre Degond. Global existence for the Vlasov-Poisson equation in 3 space variables with small initial data. In *Annales de l'IHP Analyse non linéaire*, volume 2, pages 101–118, 1985.

[BEKM16] D. Bigoni, A. Engsig-Karup, and Y. Marzouk. Spectral tensor-train decomposition. *SJSC*, 38(4):A2405–A2439, 2016.

[Bel15] Richard E Bellman. *Adaptive control processes: a guided tour*, volume 2045. Princeton university press, 2015.

[Ben03] Jean-David Benamou. Numerical resolution of an "unbalanced" mass transport problem. *ESAIM: Mathematical Modelling and Numerical Analysis*, 37(5):851–868, 2003.

[BF14] Erik Burman and Miguel A Fernández. Explicit strategies for incompressible fluid-structure interaction problems: Nitsche type mortaring versus robin–robin coupling. *International Journal for Numerical Methods in Engineering*, 97(10):739–758, 2014.

[BG04] Hans-Joachim Bungartz and Michael Griebel. Sparse grids. *Acta numerica*, 13:147–269, 2004.

[BG15] Jonas Ballani and Lars Grasedyck. Hierarchical tensor approximation of output quantities of parameter-dependent PDEs. *SIAM/ASA Journal on Uncertainty Quantification*, 3(1):852–872, 2015.

[BHL93] Gal Berkooz, Philip Holmes, and John L Lumley. The proper orthogonal decomposition in the analysis of turbulent flows. *Annual review of fluid mechanics*, 25(1):539–575, 1993.

[BMS05] Peter Benner, Volker Mehrmann, and Danny C Sorensen. *Dimension reduction of large-scale systems*, volume 45. Springer, 2005.

[BNTT11] Joakim Bäck, Fabio Nobile, Lorenzo Tamellini, and Raul Tempone. Stochastic spectral galerkin and collocation methods for pdes with random coefficients: a numerical comparison. In *Spectral and high order methods for partial differential equations*, pages 43–62. Springer, 2011.

[Bra16] JU Brackbill. On energy and momentum conservation in particle-in-cell plasma simulation. *Journal of Computational Physics*, 317:405–427, 2016.

[BRD15] Rosalie Bélanger-Rioux and Laurent Demanet. Compressed absorbing boundary conditions via matrix probing. *SIAM Journal on Numerical Analysis*, 53(5):2441–2471, 2015.

[Bre87] Yann Brenier. Décomposition polaire et réarrangement monotone des champs de vecteurs. *CR Acad. Sci. Paris Sér. I Math.*, 305:805–808, 1987.

[Bre91] Yann Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Communications on pure and applied mathematics*, 44(4):375–417, 1991.

[BRK01] Roland Brun, Peter Reichert, and Hans R Künsch. Practical identifiability analysis of large environmental simulation models. *Water Resources Research*, 37(4):1015–1030, 2001.

[BSU16] Markus Bachmayr, Reinhold Schneider, and André Uschmajew. Tensor networks and hierarchical tensors for the solution of high-dimensional partial differential equations. *Foundations of Computational Mathematics*, 16(6):1423–1472, 2016.

[BT11] George EP Box and George C Tiao. *Bayesian inference in statistical analysis*, volume 40. John Wiley & Sons, 2011.

[CABCL09] Ariel Cintrón-Arias, HT Banks, Alex Capaldi, and Alun L Lloyd. A sensitivity matrix based methodology for inverse problem formulation. *Journal of Inverse and Ill-Posed Problems*, 17(6):545–564, 2009.

[Caf00] Luis A Caffarelli. Monotonicity properties of optimal transportation and the fkg and related inequalities. *Communications in Mathematical Physics*, 214(3):547–563, 2000.

[Car15] K. Carlberg. Adaptive h-refinement for reduced-order models. *International Journal for Numerical Methods in Engineering*, 102(5):1192–1210, 2015.

[CCMA17] N. Cagniart, R. Crisovan, Y. Maday, and R. Abgrall. Model order reduction for hyperbolic problems: a new framework. 2017.

[CD15] Albert Cohen and Ronald DeVore. Kolmogorov widths under holomorphic mappings. *IMA Journal of Numerical Analysis*, 36(1):1–12, 2015.

[CDD+12] Albert Cohen, Ingrid Daubechies, Ronald DeVore, Gerard Kerkyacharian, and Dominique Picard. Capturing ridge functions in high dimensions from point queries. *Constructive Approximation*, 35(2):225–243, 2012.

[CDE06] Mike Christie, Vasily Demyanov, and Demet Erbas. Uncertainty quantification for porous media flows. *Journal of Computational Physics*, 217(1):143–158, 2006.

[CDM13] Frédérique Charles, Bruno Després, and Michel Mehrenberger. Enhanced convergence estimates for semi-lagrangian schemes application to the Vlasov–Poisson equation. *SIAM Journal on Numerical Analysis*, 51(2):840–863, 2013.

[CDQ14] Claudia Maria Colciago, Simone Deparis, and Alfio Quarteroni. Comparisons between reduced order models and full 3D models for fluid–structure interaction problems in haemodynamics. *Journal of Computational and Applied Mathematics*, 265:120–138, 2014.

[CELI15] Paul G Constantine, Michael Emory, Johan Larsson, and Gianluca Iaccarino. Exploiting active subspaces to quantify uncertainty in the numerical simulation of the hyshot ii scramjet. *Journal of Computational Physics*, 302:1–20, 2015.

[CH14] Paul Cazeaux and Jan S Hesthaven. Multiscale time-integration for particle-in-cell methods. Technical report, Elsevier, 2014.

[CHAN+15] Nathan Collier, Abdul-Lateef Haji-Ali, Fabio Nobile, Erik Von Schwerin, and Raúl Tempone. A continuation multilevel monte carlo algorithm. *BIT Numerical Mathematics*, 55(2):399–432, 2015.

[Cia00] Philippe G. Ciarlet. *Mathematical elasticity. Vol. III*, volume 29 of *Studies in Mathematics and its Applications*. North-Holland Publishing Co., Amsterdam, 2000. Theory of shells.

[CLMM09] Thierry Crestaux, Olivier Le Maıtre, and Jean-Marc Martinez. Polynomial chaos expansion for sensitivity analysis. *Reliability Engineering & System Safety*, 94(7):1161–1172, 2009.

[CLS09] Nicolas Crouseilles, Guillaume Latu, and Eric Sonnendrücker. A parallel Vlasov solver based on local cubic spline interpolation on patches. *Journal of Computational Physics*, 228(5):1429–1446, 2009.

[CMS10] Nicolas Crouseilles, Michel Mehrenberger, and Eric Sonnendrücker. Conservative semi-lagrangian schemes for Vlasov equations. *Journal of Computational Physics*, 229(6):1927–1953, 2010.

[CMS19] N. Cagniart, Y. Maday, and B. Stamm. Model order reduction for problems with large convection effects. In *Contributions to Partial Differential Equations and Applications*, pages 131–150. Springer, 2019.

[Con15] Paul G Constantine. *Active Subspaces: Emerging Ideas for Dimension Reduction in Parameter Studies*. SIAM, 2015.

[CPSV18] Lenaic Chizat, Gabriel Peyré, Bernhard Schmitzer, and François-Xavier Vialard. An interpolating distance between optimal transport and fisher–rao metrics. *Foundations of Computational Mathematics*, 18(1):1–44, 2018.

[CPW14] Paul G Constantine, Eric T Phipps, and Timothy M Wildey. Efficient uncertainty propagation for network multiphysics systems. *International Journal for Numerical Methods in Engineering*, 99(3):183–202, 2014.

[CRD+11] Paolo Crosetto, Philippe Reymond, Simone Deparis, Dimitrios Kontaxakis, Nikolaos Stergiopulos, and Alfio Quarteroni. Fluid–structure interaction simulation of aortic blood flow. *Computers & Fluids*, 43(1):46–57, 2011.

[CVJS16] William J Crumb, Jose Vicente, Lars Johannesen, and David G Strauss. An evaluation of 30 clinical drugs against the comprehensive in vitro proarrhythmia assay (cipa) proposed ion channel panel. *Journal of pharmacological and toxicological methods*, 81:251–262, 2016.

[CVK16] H Cho, Daniele Venturi, and George E Karniadakis. Numerical methods for high-dimensional probability density function equations. *Journal of Computational Physics*, 305:817–837, 2016.

[CZAL13] Liqun Cao, Lei Zhang, Walter Allegretto, and Yanping Lin. Multiscale asymptotic method for Steklov eigenvalue equations in composite media. *SIAM Journal on Numerical Analysis*, 51(1):273–296, 2013.

[DCP+16] C. C. Drovandi, N. Cusimano, S. Psaltis, B. A. J. Lawson, A. N. Pettitt, P. Burrage, and K. Burrage. Sampling methods for exploring between-subject variability in cardiac electrophysiology experiments. *Journal of The Royal Society Interface*, 13(121), 2016.

[DD91] Laurent Desvillettes and Jean Dolbeault. On long time asymptotics of the Vlasov—Poisson—Boltzmann equation. *Communications in partial differential equations*, 16(2-3):451–489, 1991.

[DDFQ06] Simone Deparis, Marco Discacciati, Gilles Fourestey, and Alfio Quarteroni. Fluid–structure algorithms based on Steklov–Poincaré operators. *Computer Methods in Applied Mechanics and Engineering*, 195(41):5797–5812, 2006.

[DDM17] Shaun M Davidson, Paul D Docherty, and Rua Murray. The dimensional reduction method for identification of parameters that trade-off due to similar model roles. *Mathematical Biosciences*, 285:119–127, 2017.

[DKLM14] Sergey Dolgov, Boris N Khoromskij, Alexander Litvinenko, and Hermann G Matthies. Computation of the response surface in the tensor train data format. *arXiv preprint arXiv:1406.2816*, 2014.

[DL04] Socrates Dokos and Nigel H Lovell. Parameter estimation in cardiac ionic models. *Progress in biophysics and molecular biology*, 85(2):407–431, 2004.

[DSL08] Vin De Silva and Lek-Heng Lim. Tensor rank and the ill-posedness of the best low-rank approximation problem. *SIAM Journal on Matrix Analysis and Applications*, 30(3):1084–1127, 2008.

[EMSU12] Oliver G Ernst, Antje Mugler, Hans-Jörg Starkloff, and Elisabeth Ullmann. On the convergence of generalized polynomial chaos expansions. *ESAIM: Mathematical Modelling and Numerical Analysis*, 46(2):317–339, 2012.

[EP14] Jens L Eftang and Anthony T Patera. A port-reduced static condensation reduced basis element method for large component-synthesized structures: approximation and a posteriori error estimation. *Advanced Modeling and Simulation in Engineering Sciences*, 1(1):1, 2014.

[Fig07] Alessio Figalli. Existence, uniqueness, and regularity of optimal transport maps. *SIAM Journal on Mathematical Analysis*, 39(1):126–137, 2007.

[FL18] F. Feppon and P. F. J. Lermusiaux. A geometric approach to dynamical model order reduction. *SIAM Journal on Matrix Analysis and Applications*, 39(1):510–538, 2018.

[FLV15] Miguel A Fernández, Mikel Landajuela, and Marina Vidrascu. Fully decoupled time-marching schemes for incompressible fluid/thin-walled structure interaction. *Journal of Computational Physics*, 297:156–181, 2015.

[FN14] Zheng Fang and David P Nicholls. An operator expansions method for computing Dirichlet–Neumann operators in linear elastodynamics. *Journal of Computational Physics*, 272:266–278, 2014.

[FQV10] Luca Formaggia, Alfio Quarteroni, and Allesandro Veneziani. *Cardiovascular Mathematics: Modeling and simulation of the circulatory system*, volume 1. Springer Science & Business Media, 2010.

[FS03] Francis Filbet and Eric Sonnendrücker. Comparison of eulerian vlasov solvers. *Computer Physics Communications*, 150(3):247–266, 2003.

[FSV12] Massimo Fornasier, Karin Schnass, and Jan Vybiral. Learning functions of few arbitrary linear parameters in high dimensions. *Foundations of Computational Mathematics*, 12(2):229–262, 2012.

[FVCJ+06] C Alberto Figueroa, Irene E Vignon-Clementel, Kenneth E Jansen, Thomas JR Hughes, and Charles A Taylor. A coupled momentum method for modeling blood flow in three-dimensional deformable arteries. *Computer methods in applied mechanics and engineering*, 195(41):5685–5706, 2006.

[GAG11] Lionel Gendre, Olivier Allix, and Pierre Gosselet. A two-scale approximation of the Schur complement and its use for non-intrusive coupling. *International Journal for Numerical Methods in Engineering*, 87(9):889–905, 2011.

[GB15] R Gul and S Bernhard. Parametric uncertainty and global sensitivity analysis in a model of the carotid bifurcation: Identification and ranking of most sensitive model parameters. *Mathematical Biosciences*, 269:104–116, 2015.

[GB18] F. J Gonzalez and M. Balajewicz. Learning low-dimensional feature dynamics using deep convolutional recurrent autoencoders. *arXiv preprint arXiv:1808.01346*, 2018.

[GBRQ14] Philip Gemmell, Kevin Burrage, Blanca Rodriguez, and T Alexander Quinn. Population of computational rabbit-specific ventricular action potential models for investigating sources of variability in cellular repolarisation. *PLoS One*, 9(2):e90112, 2014.

[GE03] Isabelle Guyon and André Elisseeff. An introduction to variable and feature selection. *Journal of machine learning research*, 3(Mar):1157–1182, 2003.

[GFA+16] Kai Germaschewski, William Fox, Stephen Abbott, Narges Ahmadi, Kristofor Maynard, Liang Wang, Hartmut Ruhl, and Amitava Bhattacharjee. The plasma simulation code: A modern particle-in-cell code with patch-based load-balancing. *Journal of Computational Physics*, 318:305–326, 2016.

[GHO17] Roger Ghanem, David Higdon, and Houman Owhadi. *Handbook of uncertainty quantification*, volume 6. Springer, 2017.

[Gil08] Michael B Giles. Multilevel monte carlo path simulation. *Operations Research*, 56(3):607–617, 2008.

[GKT13] Lars Grasedyck, Daniel Kressner, and Christine Tobler. A literature survey of low-rank tensor approximation techniques. *GAMM-Mitteilungen*, 36(1):53–78, 2013.

[Gla96] Robert T Glassey. *The Cauchy problem in kinetic theory*. SIAM, 1996.

[GLMG06] P Gokulakrishnan, AD Lawrence, PJ McLellan, and EW Grandmaison. A functional-pca approach for analyzing and reducing complex chemical mechanisms. *Computers & chemical engineering*, 30(6):1093–1101, 2006.

[GLV14] Emmanuel Grenier, Violaine Louvet, and Paul Vigneaux. Parameter estimation in non-linear mixed effects models with saem algorithm: extension from ode to pde. *ESAIM: Mathematical Modelling and Numerical Analysis*, 48(5):1303–1329, 2014.

[GM96] Wilfrid Gangbo and Robert J McCann. The geometry of optimal transportation. *Acta Mathematica*, 177(2):113–161, 1996.

[GNSG11] J Guilleminot, A Noshadravan, C Soize, and RG Ghanem. A probabilistic model for bounded elasticity tensor random fields with application to polycrystalline microstructures. *Computer Methods in Applied Mechanics and Engineering*, 200(17):1637–1648, 2011.

[GS13] Johann Guilleminot and Christian Soize. On the statistical dependence for the components of random elasticity tensors exhibiting material symmetry properties. *Journal of elasticity*, 111(2):109–130, 2013.

[GW92] J. P. Gazeau and P. Winternitz. Symmetries of variable coefficient korteweg–de vries equations. *Journal of mathematical physics*, 33(12):4087–4102, 1992.

[Hac12] Wolfgang Hackbusch. *Tensor spaces and numerical tensor calculus*, volume 42. Springer Science & Business Media, 2012.

[HC73] Arthur Hobson and Bin-Kang Cheng. A comparison of the shannon and kullback information measures. *Journal of Statistical Physics*, 7(4):301–310, 1973.

[HK07] Wolfgang Hackbusch and Boris N Khoromskij. Tensor-product approximation to operators and functions in high dimensions. *Journal of Complexity*, 23(4-6):697–714, 2007.

[HKP13] Dinh Bao Phuong Huynh, David J Knezevic, and Anthony T Patera. A static condensation reduced basis element method: approximation and a posteriori error estimation. *ESAIM: Mathematical Modelling and Numerical Analysis*, 47(1):213–251, 2013.

[HLW06] Ernst Hairer, Christian Lubich, and Gerhard Wanner. *Geometric numerical integration: structure-preserving algorithms for ordinary differential equations*, volume 31. Springer Science & Business Media, 2006.

[HNGK09] Nikolaus Hansen, André SP Niederberger, Lino Guzzella, and Petros Koumoutsakos. A method for handling uncertainty in evolutionary optimization with an application to feedback control of combustion. *Evolutionary Computation, IEEE Transactions on*, 13(1):180–197, 2009.

[Hob69] Arthur Hobson. A new theorem of information theory. *Journal of Statistical Physics*, 1(3):383–391, 1969.

[HP18] J. S. Hesthaven and C. Pagliantini. Structure-preserving reduced basis methods for hamiltonian systems with a nonlinear poisson structure. Technical report, 2018.

[Hug68] Gordon Hughes. On the mean accuracy of statistical pattern recognizers. *IEEE transactions on information theory*, 14(1):55–63, 1968.

[Hwa04] Hyung Ju Hwang. Regularity for the Vlasov–Poisson system in a convex domain. *SIAM journal on mathematical analysis*, 36(1):121–171, 2004.

[IQR16] Laura Iapichino, Alfio Quarteroni, and Gianluigi Rozza. Reduced basis method and domain decomposition for elliptic problems in networks and complex parametrized geometries. *Computers & Mathematics with Applications*, 71(1):408–430, 2016.

[Jay57a] Edwin T Jaynes. Information theory and statistical mechanics. *Physical review*, 106(4):620, 1957.

[Jay57b] Edwin T Jaynes. Information theory and statistical mechanics. *Physical review*, 106(4):620, 1957.

[JCB+15] Ross H Johnstone, Eugene TY Chang, Rémi Bardenet, Teun P De Boer, David J Gavaghan, Pras Pathmanathan, Richard H Clayton, and Gary R Mirams. Uncertainty and variability in models of the cardiac action potential: Can we build trustworthy models? *Journal of molecular and cellular cardiology*, 2015.

[Kan42] Leonid Vitalevich Kantorovich. On the transfer of masses. *Dokl. Akad. Nauk SSSR*, 37, 1942.

[KB09] Tamara G Kolda and Brett W Bader. Tensor decompositions and applications. *SIAM review*, 51(3):455–500, 2009.

[Kho07] Boris Khoromskij. Structured data-sparse approximation to high order tensors arising from the deterministic boltzmann equation. *Mathematics of computation*, 76(259):1291–1315, 2007.

[Kho11] Boris N Khoromskij. O (dlog n)-quantics approximation of n-d tensors in high-dimensional numerical modeling. *Constructive Approximation*, 34(2):257–280, 2011.

[Kho12] Boris N Khoromskij. Tensors-structured numerical methods in scientific computing: Survey on recent advances. *Chemometrics and Intelligent Laboratory Systems*, 110(1):1–19, 2012.

[KKNT15] D. Kressner, R. Kumar, F. Nobile, and C. Tobler. Low-rank tensor approximation for high-order correlation functions of gaussian random fields. *SIAM/ASA JUQ*, 3(1):393–416, 2015.

[KL51] Solomon Kullback and Richard A Leibler. On information and sufficiency. *The annals of mathematical statistics*, 22(1):79–86, 1951.

[KL87] LF Kozachenko and Nikolai N Leonenko. Sample estimate of the entropy of a random vector. *Problemy Peredachi Informatsii*, 23(2):9–16, 1987.

[KL10] O. Koch and C. Lubich. Dynamical tensor approximation. *SIAM Journal on Matrix Analysis and Applications*, 31(5):2360–2375, 2010.

[KMH+04] Stefan Kern, Sibylle D Müller, Nikolaus Hansen, Dirk Büche, Jiri Ocenasek, and Petros Koumoutsakos. Learning probability distributions in continuous evolutionary algorithms–a comparative review. *Natural Computing*, 3(1):77–112, 2004.

[Kor15a] Katharina Kormann. A semi-lagrangian vlasov solver in tensor train format. *SIAM Journal on Scientific Computing*, 37(4):B613–B632, 2015.

[Kor15b] Katharina Kormann. A semi-lagrangian Vlasov solver in tensor train format. *SIAM Journal on Scientific Computing*, 37(4):B613–B632, 2015.

[Kou09] Phaedon-Stelios Koutsourelakis. A multi-resolution, non-parametric, Bayesian framework for identification of spatially-varying model parameters. *Journal of computational physics*, 228(17):6184–6211, 2009.

[KS06] Jari Kaipio and Erkki Somersalo. *Statistical and computational inverse problems*, volume 160. Springer Science & Business Media, 2006.

[KS08] Tamara G Kolda and Jimeng Sun. Scalable tensor decompositions for multi-aspect data mining. In *2008 Eighth IEEE international conference on data mining*, pages 363–372. IEEE, 2008.

[KS11] Boris N Khoromskij and Christoph Schwab. Tensor-structured galerkin approximation of parametric and stochastic elliptic PDEs. *SIAM Journal on Scientific Computing*, 33(1):364–385, 2011.

[KSG04] Alexander Kraskov, Harald Stögbauer, and Peter Grassberger. Estimating mutual information. *Physical review E*, 69(6):066138, 2004.

[LBLM09] Claude Le Bris, Tony Lelievre, and Yvon Maday. Results and questions on a nonlinear approximation approach for solving high-dimensional partial differential equations. *Constructive Approximation*, 30(3):621, 2009.

[LC18] K. Lee and K. Carlberg. Model reduction of dynamical systems on nonlinear manifolds using deep convolutional autoencoders. *arXiv preprint arXiv:1812.08373*, 2018.

[LFNR16] Daniel M Lombardo, Flavio H Fenton, Sanjiv M Narayan, and Wouter-Jan Rappel. Comparison of detailed and simplified models of human atrial myocytes to recapitulate patient specific properties. *PLOS Comput Biol*, 12(8):e1005060, 2016.

[LMS18] Matthias Liero, Alexander Mielke, and Giuseppe Savaré. Optimal entropy-transport problems and a new hellinger–kantorovich distance between positive measures. *Inventiones mathematicae*, 211(3):969–1117, 2018.

[LOV15] Christian Lubich, Ivan V Oseledets, and Bart Vandereycken. Time integration of tensor trains. *SIAM Journal on Numerical Analysis*, 53(2):917–941, 2015.

[LP91] Pierre-Louis Lions and Benoit Perthame. Propagation of moments and regularity for the 3-dimensional Vlasov-Poisson system. *Inventiones mathematicae*, 105(1):415–430, 1991.

[LP10] Haw-Ling Liew and Peter M Pinsky. Matrix-Padé via Lanczos solutions for vibrations of fluid–structure interaction. *International Journal for Numerical Methods in Engineering*, 84(10):1183–1204, 2010.

[LPS14] Gabriel J Lord, Catherine E Powell, and Tony Shardlow. *An introduction to computational stochastic PDEs*, volume 50. Cambridge University Press, 2014.

[LRSV13] C. Lubich, T. Rohwedder, R. Schneider, and B. Vandereycken. Dynamical approximation by hierarchical tucker and tensor-train tensors. *SJMAA*, 34(2):470–494, 2013.

[LS13] Trung Bao Le and Fotis Sotiropoulos. Fluid–structure interaction of an aortic heart valve prosthesis driven by an animated anatomic left ventricle. *Journal of Computational Physics*, 244(0):41 – 62, 2013. Multi-scale Modeling and Simulation of Biological Systems.

[MDS+18] Daniel Millard, Qianyu Dang, Hong Shi, Xiaou Zhang, Chris Strock, Udo Kraushaar, Haoyu Zeng, Paul Levesque, Hua-Rong Lu, Jean-Michel Guillon, et al. Cross-site reliability of human induced pluripotent stem-cell derived cardiomyocyte based safety assays using microelectrode arrays: Results from a blinded cipa pilot study. *Toxicological Sciences*, page kfy110, 2018.

[MEDI09] Nicolette Meshkat, Marisa Eisenberg, and Joseph J DiStefano III. An algorithm for finding globally identifiable parameter combinations of nonlinear ode models using gröbner bases. *Mathematical biosciences*, 222(2):61–72, 2009.

[MMQ16] Y. Maday, A. Manzoni, and A. Quarteroni. An online intrinsic stabilization strategy for the reduced basis approximation of parametrized advection-dominated problems. *Comptes Rendus Mathematique*, 354(12):1188–1194, 2016.

[MNZ15] Eleonora Musharbash, Fabio Nobile, and Tao Zhou. Error analysis of the dynamically orthogonal approximation of time dependent random pdes. *SIAM Journal on Scientific Computing*, 37(2):A776–A810, 2015.

[Mon81] Gaspard Monge. Mémoire sur la théorie des déblais et des remblais. *Histoire de l'Académie Royale des Sciences de Paris*, 1781.

[Mor05] PJ Morrison. Hamiltonian and action principle formulations of plasma physicsa). *Physics of Plasmas (1994-present)*, 12(5):058102, 2005.

[MP84] L. R. Mead and N. Papanicolaou. Maximum entropy in the problem of moments. *Journal of Mathematical Physics*, 25(8):2404–2417, 1984.

[MRH15] Immanuel Martini, Gianluigi Rozza, and Bernard Haasdonk. Reduced basis approximation and a-posteriori error estimation for the coupled Stokes-Darcy system. *Advances in Computational Mathematics*, 41(5):1131–1157, 2015.

[MRS14] Éric Madaule, Marco Restelli, and Eric Sonnendrücker. Energy conserving discontinuous Galerkin spectral element method for the Vlasov–Poisson system. *Journal of Computational Physics*, 279:261–288, 2014.

[MS18] S. Mowlavi and T. P. Sapsis. Model order reduction for stochastic dynamical systems with continuous symmetries. *SIAM Journal on Scientific Computing*, 40(3):A1669–A1695, 2018.

[MT11] Eve Marder and Adam L Taylor. Multiple models to capture the variability in biological neurons and networks. *Nature neuroscience*, 14(2):133–138, 2011.

[MUV15] S Mayer, T Ullrich, and J Vybıral. Entropy and sampling numbers of classes of ridge functions, vol. 42, 2015.

[MVL11] M. Muskulus and S. Verduyn-Lunel. Wasserstein distances in the analysis of time series and dynamical systems. *Physica D: Nonlinear Phenomena*, 240(1):45–58, 2011.

[MW82] Jerrold E. Marsden and Alan Weinstein. The hamiltonian structure of the Maxwell-Vlasov equations. *Physica D: Nonlinear Phenomena*, 4(3):394 – 406, 1982.

[MXA+12] P. Moireau, N. Xiao, M. Astorino, C. A. Figueroa, D. Chapelle, C. A. Taylor, and J-F. Gerbeau. External tissue support and fluid-structure simulation in blood flows. *Biomechanics and Modeling in Mechanobiology*, 11(1):1–18, 2012.

[Nat95] Ramesh Natarajan. Domain decomposition using spectral expansions of Steklov–Poincaré operators. *SIAM Journal on Scientific Computing*, 16(2):470–495, 1995.

[Nat97] Ramesh Natarajan. Domain decomposition using spectral expansions of Steklov–Poincaré operators ii: A matrix formulation. *SIAM Journal on Scientific Computing*, 18(4):1187–1199, 1997.

[NB18] N. J. Nair and M. Balajewicz. Transported snapshot model order reduction approach for parametric, steady-state fluid flows containing parameter-dependent shocks. *International Journal for Numerical Methods in Engineering*, 2018.

[Něm10] Jana Němcová. Structural identifiability of polynomial and rational systems. *Mathematical biosciences*, 223(2):83–96, 2010.

[NLM09] Anthony Nouy and Olivier P Le Maıˆtre. Generalized spectral decomposition for stochastic nonlinear problems. *Journal of computational physics*, 228(1):202–235, 2009.

[Nou10] Anthony Nouy. A priori model reduction through proper generalized decomposition for solving time-dependent partial differential equations. *CMAME*, 199(23):1603–1626, 2010.

[NV08] F. Nobile and C. Vergara. An effective fluid-structure interaction formulation for vascular dynamics by generalized robin conditions. *SIAM Journal on Scientific Computing*, 30(2):731–763, 2008.

[OMO14] Johnny T Ottesen, Jesper Mehlsen, and Mette S Olufsen. Structural correlation method for model reduction and practical estimation of patient specific parameters illustrated on heart rate regulation. *Mathematical biosciences*, 257:50–59, 2014.

[OR13] M. Ohlberger and S. Rave. Nonlinear reduced basis approximation of parameterized evolution equations via the method of freezing. *Comptes Rendus Mathematique*, 351(23-24):901–906, 2013.

[Ose11] Ivan V Oseledets. Tensor-train decomposition. *SIAM Journal on Scientific Computing*, 33(5):2295–2317, 2011.

[PBWB14] B. Peherstorfer, D. Butnaru, K. Willcox, and H.-J. Bungartz. Localized discrete empirical interpolation method. *SIAM Journal on Scientific Computing*, 36(1):A168–A192, 2014.

[PDB+16] E Pueyo, CE Dangerfield, OJ Britton, L Virág, K Kistamás, N Szentandrássy, N Jost, A Varró, PP Nánási, K Burrage, et al. Experimentally-based computational investigation into beat-to-beat variability in ventricular repolarization and its response to ionic current inhibition. *PloS one*, 11(3):e0151461, 2016.

[Pin12] Allan Pinkus. *N-widths in Approximation Theory*, volume 7. Springer Science & Business Media, 2012.

[Pir14] Olivier Pironneau. Simplified Fluid-Structure Interactions for Hemodynamics. In Sergio R. Idelsohn, editor, *Numerical Simulations of Coupled Problems in Engineering*, volume 33 of *Computational Methods in Applied Sciences*, pages 57–70. Springer International Publishing, 2014.

[PM08] Martin Campos Pinto and Michel Mehrenberger. Convergence of an adaptive semi-lagrangian scheme for the Vlasov-Poisson system. *Numerische Mathematik*, 108(3):407–444, 2008.

[PR14] Benedetto Piccoli and Francesco Rossi. Generalized wasserstein distance and its application to transport equations with source. *Archive for Rational Mechanics and Analysis*, 211(1):335–358, 2014.

[Pre98] Lutz Prechelt. Automatic early stopping using cross validation: quantifying the criteria. *Neural Networks*, 11(4):761–767, 1998.

[PS05] Rodrigo R Paz and Mario A Storti. An interface strip preconditioner for domain decomposition methods: application to hydrology. *International journal for numerical methods in engineering*, 62(13):1873–1894, 2005.

[PS09] Edoardo Patelli and Gerhart Schuëller. On optimization techniques to reconstruct microstructures of random heterogeneous media. *Computational Materials Science*, 45(2):536–549, 2009.

[PVG+11] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[PW15] Benjamin Peherstorfer and Karen Willcox. Online adaptive model reduction for nonlinear systems via low-rank updates. *SIAM Journal on Scientific Computing*, 37(4):A2123–A2150, 2015.

[QMN15] Alfio Quarteroni, Andrea Manzoni, and Federico Negri. *Reduced basis methods for partial differential equations: an introduction*, volume 92. Springer, 2015.

[QR14] Alfio Quarteroni and Gianluigi Rozza. *Reduced Order Methods for Modeling and Computational Reduction*, volume 9. Springer, 2014.

[QV99] A. Quarteroni and A. Valli. *Domain Decomposition methods for partial differential equations.* Oxford University Press, 1999.

[RBZ+17] Fabien Raphel, Muriel Boulakia, Nejib Zemzemi, Yves Coudière, Jean-Michel Guillon, Philippe Zitoun, and Jean-Frédéric Gerbeau. Identification of ion currents components generating field potential recorded in mea from hipsc-cm. *IEEE Transactions on Biomedical Engineering*, 65(6):1311–1319, 2017.

[RGMP14] T Chacon Rebollo, V Girault, F Murat, and O Pironneau. Analysis of a simplified coupled fluid-structure model for computational hemodynamics. *Journal de l'Ecole Polytechnique*, 2014.

[RKM+09] Andreas Raue, Clemens Kreutz, Thomas Maiwald, Julie Bachmann, Marcel Schilling, Ursula Klingmüller, and Jens Timmer. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics*, 25(15):1923–1929, 2009.

[RRG+86] Fane Robinson, Charles E Riva, Juan E Grunwald, Benno L Petrig, and Stephen H Sinclair. Retinal blood flow autoregulation in response to an acute increase in blood pressure. *Investigative ophthalmology & visual science*, 27(5):722–726, 1986.

[RSSK14] N. Ravindran, N. Sidiropoulos, S. Smith, and G. Karypis. Memory-efficient parallel computation of tensor and matrix products for big tensor decomposition. In *2014 48th Asilomar Conference on Signals, Systems and Computers*, pages 581–585. IEEE, 2014.

[San15] Filippo Santambrogio. Optimal transport for applied mathematicians. *Birkäuser, NY*, 55:58–63, 2015.

[SBN+13] Mikael Sunnåker, Alberto Giovanni Busetto, Elina Numminen, Jukka Corander, Matthieu Foll, and Christophe Dessimoz. Approximate bayesian computation. *PLoS computational biology*, 9(1):e1002803, 2013.

[SBOW+14] Carlos Sánchez, Alfonso Bueno-Orovio, Erich Wettwer, Simone Loose, Jana Simon, Ursula Ravens, Esther Pueyo, and Blanca Rodriguez. Inter-subject variability in human atrial action potential in sinus rhythm versus chronic atrial fibrillation. *PloS one*, 9(8):e105897, 2014.

[SG11] Christoph Schwab and Claude Jeffrey Gittelson. Sparse tensor discretizations of high-dimensional parametric and stochastic PDEs. *Acta Numerica*, 20:291–467, 2011.

[SKM+15] Daniele E Schiavazzi, Ethan O Kung, Alison L Marsden, Catriona Baker, Giancarlo Pennati, Tain-Yen Hsia, Anthony Hlavacek, Adam L Dorfman, Modeling of Congenital Hearts Alliance (MOCHA) Investigators, et al. Hemodynamic effects of left pulmonary artery stenosis after superior cavopulmonary connection: a patient-specific multiscale modeling study. *The Journal of thoracic and cardiovascular surgery*, 149(3):689–696, 2015.

[Smi13] Ralph C Smith. *Uncertainty quantification: theory, implementation, and applications*, volume 12. Siam, 2013.

[ST43] James Alexander Shohat and Jacob David Tamarkin. *The problem of moments*. American Mathematical Society, 1943.

[SVNL05] Z Syed, E Vigmond, S Nattel, and LJ Leon. Atrial cell action potential parameter fitting using genetic algorithms. *Medical and Biological Engineering and Computing*, 43(5):561–571, 2005.

[SZ03] Steve Smale and Ding-Xuan Zhou. Estimating the approximation error in learning theory. *Analysis and Applications*, 1(01):17–41, 2003.

[SZ06] S. Sankaran and N. Zabaras. A maximum entropy approach for property prediction of random microstructures. *Acta Materialia*, 54(8):2265–2276, 2006.

[TBR18] D. Torlo, F. Ballarin, and G. Rozza. Stabilized weighted reduced basis methods for parametrized advection dominated problems with random inputs. *SIAM/ASA Journal on Uncertainty Quantification*, 6(4):1475–1502, 2018.

[Tru79] Gerard V Trunk. A problem of dimensionality: A simple example. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (3):306–307, 1979.

[TSGU13] Aretha L Teckentrup, Robert Scheichl, Michael B Giles, and Elisabeth Ullmann. Further analysis of multilevel monte carlo methods for elliptic pdes with random coefficients. *Numerische Mathematik*, 125(3):569–600, 2013.

[vAdB04] R. van Loon, P.D. Anderson, J. de Hart, and F. Baaijens. A combined fictitious domain/adaptative meshing method for fluid-structure interaction in heart valves. *Int. J. Num. Meth. Fluids*, 46:533–544, 2004.

[vAv06] R. van Loon, P.D. Anderson, and F.N. van de Vosse. A fluid-structure interaction method with solid-rigid contact for heart valve dynamics. *J. Comp. Phys.*, 217:806–823, 2006.

[Vil03] Cédric Villani. *Topics in optimal transportation*. Number 58. American Mathematical Soc., 2003.

[Vil08] Cédric Villani. *Optimal transport: old and new*, volume 338. Springer Science & Business Media, 2008.

[Wel15] G. Welper. Transformed snapshot interpolation. *arXiv preprint arXiv:1505.01227*, 2015.

[Wel17] G. Welper. *h* and *hp*-adaptive interpolation by transformed snapshots for parametric and stochastic hyperbolic pdes. *arXiv preprint arXiv:1710.11481*, 2017.

[WHC+04] Erich Wettwer, Ottó Hála, Torsten Christ, Jürgen F Heubach, Dobromir Dobrev, Michael Knaut, András Varró, and Ursula Ravens. Role of ikur in controlling action potential shape and contractility in the human atrium influence of chronic atrial fibrillation. *Circulation*, 110(16):2299–2306, 2004.

[WRWD84] Svante Wold, Arnold Ruhe, Herman Wold, and WJ Dunn, III. The collinearity problem in linear regression. the partial least squares (pls) approach to generalized

inverses. *SIAM Journal on Scientific and Statistical Computing*, 5(3):735–743, 1984.

[WSJY16] Xingyu Wang, Roman Samulyak, Xiangmin Jiao, and Kwangmin Yu. AP-Cloud: Adaptive Particle-in-Cloud method for optimal solutions to Vlasov–Poisson equation. *Journal of Computational Physics*, 316:682–699, 2016.

[WZ04] Jingbo Wang and Nicholas Zabaras. A Bayesian inference approach to the inverse heat conduction problem. *International Journal of Heat and Mass Transfer*, 47(17):3927–3941, 2004.

[XH05] Dongbin Xiu and Jan S Hesthaven. High-order collocation methods for differential equations with random inputs. *SIAM Journal on Scientific Computing*, 27(3):1118–1139, 2005.

[Xiu07] Dongbin Xiu. Efficient collocational approach for parametric uncertainty analysis. *Communications in computational physics*, 2(2):293–309, 2007.

[XK02] Dongbin Xiu and George Em Karniadakis. The wiener–askey polynomial chaos for stochastic differential equations. *SIAM journal on scientific computing*, 24(2):619–644, 2002.

[XOMN10] Jin Xu, Peter N Ostroumov, Brahim Mustapha, and Jerry Nolen. Scalable direct vlasov solver with discontinuous galerkin method on unstructured mesh. *SIAM Journal on Scientific Computing*, 32(6):3476–3494, 2010.

[YGPD13] Sylvia Young, Michael E Goddard, Jennie E Pryce, and Guang Deng. Kernel methods and haplotypes used in selection of sparse dna markers for protein yield in dairy cattle. *Mathematical biosciences*, 243(1):57–66, 2013.

[YKI+18] Daiju Yamazaki, Takashi Kitaguchi, Masakazu Ishimura, Tomohiko Taniguchi, Atsuhiro Yamanishi, Daisuke Saji, Etsushi Takahashi, Masao Oguchi, Yuta Moriyama, Sanae Maeda, et al. Proarrhythmia risk prediction using human induced pluripotent stem cell-derived cardiomyocytes. *Journal of pharmacological sciences*, 136(4):249–256, 2018.

[ZG08] N. Zabaras and B. Ganapathysubramanian. A scalable framework for the solution of stochastic inverse problems using a sparse grid collocation approach. *Journal of Computational Physics*, 227(9):4697–4735, 2008.

[ZYO+15] Zheng Zhang, Xiu Yang, Ivan V Oseledets, George E Karniadakis, and Luca Daniel. Enabling high-dimensional hierarchical uncertainty quantification by anova and tensor-train decomposition. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 34(1):63–76, 2015.

Author's works.

[1] Matteo Aletti, Jean-Frédéric Gerbeau, and Damiano Lombardi. Modeling autoregulation in three-dimensional simulations of retinal hemodynamics. *Journal for Modeling in Ophthalmology*, 1(1):88–115, 2016.

[2] Matteo Aletti, Jean-Frédéric Gerbeau, and Damiano Lombardi. A simplified fluid–structure model for arterial flow. application to retinal hemodynamics. *Computer Methods in Applied Mechanics and Engineering*, 306:77–94, 2016.

[3] Matteo Aletti and Damiano Lombardi. A reduced-order representation of the poincaré–steklov operator: an application to coupled multi-physics problems. *International Journal for Numerical Methods in Engineering*, 111(6):581–600, 2017.

[4] Athmane Bakhta and Damiano Lombardi. An a posteriori error estimator based on shifts for positive hermitian eigenvalue problems. *Comptes Rendus Mathematique*, 356(6):696–705, 2018.

[5] Michel Bergmann, Thierry Colin, Angelo Iollo, Damiano Lombardi, Olivier Saut, and Haysam Telib. Reduced order models at work in aeronautics and medicine. In *Reduced Order Methods for Modeling and Computational Reduction*, pages 305–332. Springer, 2014.

[6] Eric Cances, Virginie Ehrlacher, David Gontier, Antoine Levitt, and Damiano Lombardi. Numerical quadrature in the brillouin zone for periodic schrodinger operators. *arXiv preprint arXiv:1805.07144*, 2018.

[7] Thierry Colin, Angelo Iollo, Damiano Lombardi, and Olivier Saut. Prediction of the evolution of thyroidal lung nodules using a mathematical model. *ERCIM News*, 2010(82):37–38, 2010.

[8] Thierry Colin, Angelo Iollo, Damiano Lombardi, and Olivier Saut. System identification in tumor growth modeling using semi-empirical eigenfunctions. *Mathematical Models and Methods in Applied Sciences*, 22(06):1250003, 2012.

[9] Thierry Colin, Angelo Iollo, Damiano Lombardi, Olivier Saut, Françoise Bonichon, and Jean Palussière. Some models for the prediction of tumor growth: General framework and applications to metastases in the lung. In *Computational Surgery and Dual Training*, pages 289–314. Springer, 2014.

[10] F Cornelis, O Saut, P Cumsille, D Lombardi, A Iollo, J Palussiere, and T Colin. In vivo mathematical modeling of tumor growth from imaging data: Soon to come in the future? *Diagnostic and Interventional Imaging*, 94(6):593–600, 2013.

[11] Rafael Coyaud, Virginie Ehrlacher, Damiano Lombardi, et al. Approximation of optimal transport problems with marginal moments constraints. Technical report, 2019.

[12] Virginie Ehrlacher, Laura Grigori, Damiano Lombardi, and Hao Song. Adaptive hierarchical subtensor partitioning for tensor compression. *hal preprint:https://hal.inria.fr/hal-02284456*, 2019.

[13] Virginie Ehrlacher and Damiano Lombardi. A dynamical adaptive tensor method for the vlasov–poisson system. *Journal of Computational Physics*, 339:285–306, 2017.

[14] Virginie Ehrlacher, Damiano Lombardi, O Mula, and F-X Vialard. Nonlinear model reduction on metric spaces. application to one-dimensional conservative pdes in wasserstein spaces. *arXiv preprint arXiv:1909.06626*, 2019.

[15] Felipe Galarce, Jean-Frédéric Gerbeau, Damiano Lombardi, and Olga Mula. State estimation with nonlinear reduced models. application to the reconstruction of blood flows with doppler ultrasound images. *arXiv preprint arXiv:1904.13367*, 2019.

[16] Jean-Frédéric Gerbeau and Damiano Lombardi. Approximated lax pairs for the reduced order integration of nonlinear evolution equations. *Journal of Computational Physics*, 265:246–269, 2014.

[17] Jean-Frédéric Gerbeau, Damiano Lombardi, and Elisa Schenone. Reduced order model in cardiac electrophysiology with approximated lax pairs. *Advances in Computational Mathematics*, 41(5):1103–1130, 2015.

[18] Jean-Frédéric Gerbeau, Damiano Lombardi, and Eliott Tixier. How to choose biomarkers in view of parameter estimation. *Mathematical biosciences*, 303:62–74, 2018.

[19] Jean-Frédéric Gerbeau, Damiano Lombardi, and Eliott Tixier. A moment-matching method to study the variability of phenomena described by partial differential equations. *SIAM Journal on Scientific Computing*, 40(3):B743–B765, 2018.

[20] Angelo Iollo and Damiano Lombardi. A lagrangian scheme for the solution of the optimal mass transfer problem. *Journal of Computational Physics*, 230(9):3430 – 3442, 2011.

[21] Angelo Iollo and Damiano Lombardi. Advection modes by optimal mass transfer. *Physical Review E*, 89(2):022923, 2014.

[22] Damiano Lombardi. *Inverse problems in tumor growth modelling*. PhD thesis, Ph. D. thesis, Institut de Mathématiques de Bordeaux (September 2011), 2011.

[23] Damiano Lombardi. Inverse problems in 1d hemodynamics on systemic networks: A sequential approach. *International journal for numerical methods in biomedical engineering*, 30(2):160–179, 2014.

[24] Damiano Lombardi and Emmanuel Maitre. Eulerian models and algorithms for unbalanced optimal transport. *ESAIM: Mathematical Modelling and Numerical Analysis*, 49(6):1717–1744, 2015.

[25] Damiano Lombardi and Sanjay Pant. Nonparametric k-nearest-neighbor entropy estimator. *Phys. Rev. E*, 93:013310, Jan 2016.

[26] Damiano Lombardi and Fabien Raphel. A greedy dimension reduction method for classification problems. *https://hal.inria.fr/hal-02280502*, 2019.

[27] Sanjay Pant and Damiano Lombardi. An information-theoretic approach to assess practical identifiability of parametric dynamical systems. *Mathematical biosciences*, 268:66–79, 2015.

[28] Fabien Raphel, Tessa de Korte, Damiano Lombardi, Stefan Braam, and Jean-Frédéric Gerbeau. A greedy classifier optimisation strategy to assess ion channel blocking activity and pro-arrhythmia in hipsc-cardiomyocytes. *HAL preprint https://hal.inria.fr/hal-02276945*, 2019.

[29] Eliott Tixier, Damiano Lombardi, Blanca Rodriguez, and Jean-Frédéric Gerbeau. Modelling variability in cardiac electrophysiology: A moment-matching approach. *Journal of the Royal Society Interface*, 14(133):20170238, 2017.

[30] Eliott Tixier, Fabien Raphel, Damiano Lombardi, and Jean-Frédéric Gerbeau. Composite biomarkers derived from micro-electrode array measurements and computer simulations improve the classification of drug-induced channel block. *Frontiers in physiology*, 8:1096, 2018.