

Campagne 2020 Contrats Doctoraux Instituts/Initiatives

Proposition de Projet de Recherche Doctoral (PRD)

Appel à projet ISCD-Institut des Sciences du calcul des Données 2020

Intitulé du Projet de Recherche Doctoral : Nonlinear reduced models and machine learning in forward modeling and inverse problems

Directeur de Thèse porteur du projet (titulaire d'une HDR) :

NOM : **COHEN** Prénom : **Albert**
Titre : Professeur des Universités ou

e-mail : cohen@ann.jussieu.fr
Adresse professionnelle : 4, Place Jussieu, 75005 Paris
(site, adresse, bât., bureau)

Unité de Recherche :
Intitulé : Laboratoire Jacques-Louis Lions
Code (ex. UMR xxxx) : UMR 7598

ED386-Sciences Mathématiques Paris Ce

Ecole Doctorale de rattachement de l'équipe & d'inscription du doctorant :

Doctorants actuellement encadrés par le directeur de thèse (préciser le nombre de doctorants, leur année de 1^{ère} inscription et la quotité d'encadrement) : Matthieu Dolbeault, normalien en pré-thèse depuis 2019.

Co-encadrant :

NOM : **Mula** Prénom : **Olga**
Titre : Maître de Conférences des Universités HDR
OU
e-mail : mula@ceremade.dauphine.fr

Unité de Recherche :
Intitulé : Ceremade
Code (ex. UMR xxxx) : UMR 7534

Choisissez un élément :

Ecole Doctorale de rattachement : Ou si ED non Alliance SU : **ED de Paris Dauphine**

Doctorants actuellement encadrés par le co-directeur de thèse (préciser le nombre de doctorants, leur année de 1^{ère} inscription et la quotité d'encadrement) : Felipe Galarce, 1^{ère} inscription en Nov. 2017, co-encadrement à 30%, soutenance espérée en Novembre 2020.

Cotutelle internationale : Non Oui, précisez Pays et Université :

Description du projet de recherche doctoral (en français ou en anglais)

3 pages maximum – interligne simple – Ce texte sera diffusé en ligne

Détailler le contexte, l'objectif scientifique, la justification de l'approche scientifique ainsi que l'adéquation à l'initiative/l'Institut.

Le cas échéant, préciser le rôle de chaque encadrant ainsi que les compétences scientifiques apportées. Indiquer les publications/productions des encadrants en lien avec le

projet.

Préciser le profil d'étudiant(e) recherché.

**Merci de nommer votre fichier pdf :
«ACRONYME de l'institut/initiative_2_NOM Porteur Projet_2020 »**

**à envoyer simultanément par e-mail à l'ED de rattachement et au programme :
cd_instituts_et_initiatives@listes.upmc.fr avant le 30 mars.**

PhD project: Nonlinear reduced models and machine learning in forward modeling and inverse problems

1. General context

Complex physical systems described by partial differential equations (PDEs) often involve extra parameters in addition to the usual time and space variables, that can be described by a vector $y = (y_1, \dots, y_d) \in Y \subset \mathbb{R}^d$. These parameters may represent physical properties such as the thermal or electric conductivity of a material, or geometrical properties such as the boundary of the domain. They may either be of deterministic or stochastic nature, depending on the considered application. In this general context, denoting by $u = u(y)$ the solution to the PDE which is assumed to be well-defined in a suitable function space V for each $y \in Y$, two principal classes of numerical problems are typically considered:

Forward numerical simulation: the solution to the PDE is typically computed by a numerical (finite difference or finite element) solver. In applications where the solution needs to be computed for many different parameter values, repeated use of this solver becomes computationally intensive. One is thus led to search for a quickly computable and accurate approximation to the parameter-to-solution map $y \mapsto u(y)$.

Inverse problems: the exact value of the parameter y is unknown as well as the solution $u = u(y)$, and one instead observes measurements in terms of a vector $z = (z_1, \dots, z_m) \in \mathbb{R}^m$ where z_i is the image of u by a linear functional ℓ_i which models a sensing device, possibly affected by noise. Typical inverse problems are either to recover an approximation of u (state estimation) or of y (parameter estimation) from z .

2. Linear reduced modeling

The last decade has witnessed an important development of *linear reduced modeling* strategies for efficiently tackling the above problems. These strategies aim at building linear spaces V_n of hopefully low dimension n that are specifically tailored to the approximation of all solutions to the parametric PDE gathered in the *solution manifold* $\mathcal{M} := \{u(y) : y \in Y\}$.

Examples of linear reduced models include proper orthogonal decompositions, reduced basis, or sparse polynomial expansions [6]. They contrast with standard approximation tools such as polynomials, finite elements, Fourier series or wavelets, which are designed to approximate well all sufficiently smooth functions, and are not specifically adapted to \mathcal{M} . In contrast, it was shown in [1] that reduced basis spaces generated by snapshots $u_i = u(y^i)$, where $y^1, \dots, y^n \in Y$ are properly selected, have optimal approximation performance, comparable to the Kolmogorov n -width $d_n(\mathcal{M})$ which is the minimum of $\text{dist}(\mathcal{M}, E)_V$ taken over all spaces E of dimension n .

These approaches have proved to be successful for a large class of problems including linear and nonlinear elliptic and parabolic PDEs, with parameters entering the coefficients and domain geometry in various ways. However, they fall short when solutions exhibit parameter dependent discontinuities or sharp transitions, which are typical in hyperbolic conservation laws or transport dominated problems. In fact, for these problems, the Kolmogorov n -width decays slowly with n , and therefore linear methods are doomed to perform poorly. Relevant examples where this difficulty arises can be found in the field of fluid dynamics. In the thesis, we will focus on two applications: one from hemodynamics related to forward and inverse problems in blood flows in carotid arteries. The second application is related to the rapid reconstruction of air pollution maps in the city of Paris. For both problems we will have access to real data from ongoing collaborations with medical doctors and a company called Numtech working in the field of pollution cartography.

3. Objective of the thesis

The objective of the PhD thesis is the development, analysis and practical implementation of *fast* and *nonlinear* model reduction strategies for the afore mentioned classes of problems. One important ingredient for this will be to combine *physical modeling* exploiting the PDE structure with *data driven strategies* which have been highly developed in a machine learning context.

In forward numerical simulation, one first approach to be tested will be *non-intrusive*: a finite number of snapshots u_i is generated by a high resolution solver, and the approximation to the solution map is computed by interpolating these snapshots in a nonlinear way. This may be thought as a high-dimensional regression problems where the data are generated by the user, and cannot be too many since each sample costs one run of the solver. Nonlinear strategies from machine learning such as deep neural networks or random forest, will be adapted to this setting and compared to other nonlinear approximation strategies such as rational fractions. One main difficulty to be addressed is how to adapt these strategies so that the resulting numerical solutions preserve the physical properties of the model, such as maximum principle or mass conservativity. This will lead to also consider *intrusive* approaches where reduced order numerical schemes will be developed via a data-driven strategy (for example in the computation of the numerical fluxes).

In inverse problems, the starting point will be the Parametrized Background Data Weak (PBDW) state estimation method introduced in [9] which builds up on linear reduced modeling: defines the estimate \tilde{u} by minimizing the distance from V_n among all elements $v \in V$ which agree with the measurement vector z . This approach has the advantage of being easy to implement and coming with certified error bounds [2]. It can be viewed as a deterministic alternative to Bayesian inversion [7] which searches for a plausible solution based on a probabilistic prior. PBDW will serve as a first building block towards nonlinear state estimation strategies. One typical avenue will be to consider an offline selected collection of reduced basis spaces, through a proper splitting of parameter space, each of them leads to a different estimator through the PBDW method. Then, statistical techniques of classification, model selection or agregation [8, 10] will be employed for the derivation of a final estimator. The envisaged strategies thus combine ideas from statistical estimation and machine learning with reduced modeling techniques. It will also be interesting to compare them with “learning only” strategies that instead ignore the model and only operate from large data sets. As already brought up, we will apply our methodology to applications from hemodynamics and air pollution.

1 PhD supervisors

The PhD will be supervised by Prof. Albert Cohen from Sorbonne Univesity, Laboratoire Jacques Louis Lions (LJLL), and Assistant Prof. Olga Mula from Paris Dauphine University, Centre de Recherche en Mathematiques de la Decision (Ceremade).

Prof. Cohen is a specialist in nonlinear approximation theory and has contributions in different topics like wavelet and harmonic analysis. His current research is focused on high-dimensional problems whose efficient numerical treatment is challenged by the so-called curse of dimensionality. One relevant application in this area are direct and inverse problems involving parametric PDEs and the development of reduced modeling techniques to overcome the high-dimensionality. Ass. Prof. Mula is a numerical analyst specializing in model reduction of parametrized PDEs and their application to real-life inverse problems. She has also contributions on the numerical solution of transport dominated problems. Relevant works of the supervisors related to the present project are [1, 2, 3, 4, 5].

2 Candidate's profile

The PhD student should have a training background in physical and numerical modeling with PDEs, as well as in data science.

References

- [1] P. Binev, A. Cohen, W. Dahmen, R. DeVore, G. Petrova, and P. Wojtaszczyk, *Convergence Rates for Greedy Algorithms in Reduced Basis Methods*, SIAM Journal of Mathematical Analysis **43**, 1457-1472, 2011.
- [2] P. Binev, A. Cohen, W. Dahmen, R. DeVore, G. Petrova, and P. Wojtaszczyk, *Data assimilation in reduced modeling*, SIAM Journal on Uncertainty Quantification, **5**, 1–29, 2017.
- [3] P. Binev, A. Cohen, O. Mula and J. Nichols, *Greedy algorithms for optimal measurements selection in state estimation using reduced models*, SIAM Journal on Uncertainty Quantification **43**, 1101-1126, 2018.
- [4] A. Cohen, W. Dahmen, R. DeVore, J. Fadili, O. Mula and J. Nichols, *Optimal reduced model algorithms for data-based state estimation*, Submitted to SIAM Journal on Numerical Analysis, 2019.
- [5] F. Galarce, J.F. Gerbeau, D. Lombardi, O. Mula, *Fast Reconstruction of Flows from Ultrasound Measurements*, Submitted to CMAME, 2020.
- [6] A. Cohen and R. DeVore, *Approximation of high dimensional parametric pdes*, Acta Numerica 24, 1-159, 2015.
- [7] M. Dashti and A.M. Stuart, *The Bayesian Approach to Inverse Problems*, Handbook of Uncertainty Quantification, Editors R. Ghanem, D. Higdon and H. Owhadi, Springer, 2015.
- [8] A. Juditsky and A. Nemirovski, *Functional aggregation for nonparametric regression*, The Annals of Statistics, 2000.
- [9] Y. Maday, A.T. Patera, J.D. Penn and M. Yano, *A parametrized-background data-weak approach to variational data assimilation: Formulation, analysis, and application to acoustics*, Int. J. Numer. Meth. Eng., submitted, 2014.
- [10] A. Tsybakov, *Optimal rates of aggregation*, in Learning Theory and Kernel Machines, Springer, 2003.